

Matrix Reordering for Noisy Disordered Matrices: Optimality and Computationally Efficient Algorithms

T. Tony Cai and Rong Ma

Abstract

Motivated by applications in single-cell biology and metagenomics, we investigate the problem of matrix reordering based on a noisy disordered monotone Toeplitz matrix model. We establish the fundamental statistical limit for this problem in a decision-theoretic framework and demonstrate that a constrained least squares estimator achieves the optimal rate. However, due to its computational complexity, we analyze a popular polynomial-time algorithm, spectral seriation, and show that it is suboptimal. To address this, we propose a novel polynomial-time adaptive sorting algorithm with guaranteed performance improvement. Simulations and analyses of two real single-cell RNA sequencing datasets demonstrate the superiority of our algorithm over existing methods.

I. INTRODUCTION

Consider the following noisy disordered matrix model

$$\mathbf{Y} = \mathbf{\Pi}\mathbf{\Theta}\mathbf{\Pi}^\top + \mathbf{Z}, \quad (\text{I.1})$$

where $\mathbf{Y}, \mathbf{\Theta}, \mathbf{Z} \in \mathbb{R}^{n \times n}$ are symmetric matrices, \mathbf{Y} is observed, \mathbf{Z} is the noise matrix with independent (up to symmetry) sub-Gaussian entries with mean zero and variance σ^2 , $\mathbf{\Theta}$ is a deterministic signal matrix with certain structural patterns of interest, and $\mathbf{\Pi} \in \mathbb{R}^{n \times n}$ is an unknown permutation matrix that simultaneously permutes the columns and rows of the signal matrix $\mathbf{\Theta}$. This paper investigates the noisy matrix reordering problem, where the aim is to recover the underlying permutation $\mathbf{\Pi}$ (Figure 1) based on the observed noisy disordered matrix \mathbf{Y} .

This matrix reordering problem, also known as the matrix seriation problem, has a long history in data analysis and data mining [1, 2]. It is often encountered when there is structural information contained in the true signal matrix $\mathbf{\Theta}$, while only the corresponding noisy and disordered matrix $\mathbf{Y} = \mathbf{\Pi}\mathbf{\Theta}\mathbf{\Pi}^\top + \mathbf{Z}$ is available. The importance of such a statistical problem lies in many applications (Section I-A), where important features about the object under investigation are embedded in the structural pattern of $\mathbf{\Theta}$, but are largely concealed from its noisy disordered

T. Tony Cai is with the Department of Statistics and Data Science at the University of Pennsylvania, Philadelphia, PA 19104 USA (E-mail: tcgai@wharton.upenn.edu).

Rong Ma is with the Department of Statistics at Stanford University, Stanford, CA 94305 USA (email: rongm@stanford.edu).

The research of Tony Cai was supported in part by NSF Grant DMS-2015259 and NIH grants R01-GM129781 and R01-GM123056. The research of Rong Ma was supported by Professor David Donoho from Stanford University.

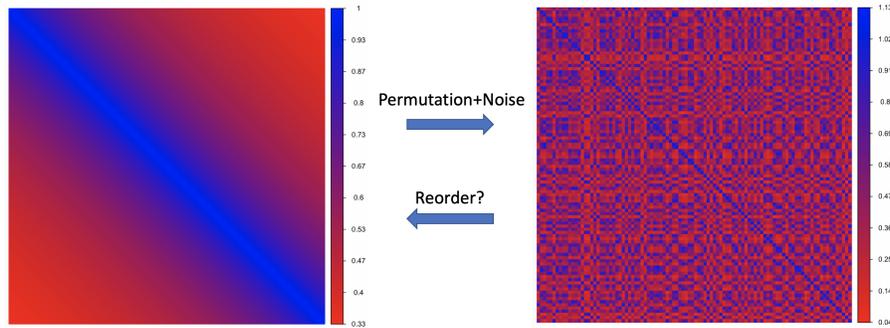


Fig. 1. A graphical illustration of the matrix reordering problem. Left: a symmetric monotone Toeplitz matrix of dimension 100×100 under Setting 6 of Section VI. Right: the observed noisy disordered matrix with an arbitrary permutation and the Gaussian noise.

observations. For instance, when Θ is a symmetric monotone Toeplitz matrix (Figure 1 left), after an arbitrary simultaneous permutation of its rows and columns, such a structural pattern is not easily discernible or recovered from a noisy observation \mathbf{Y} of the disordered matrix (Figure 1 right).

A. Reordering Monotone Toeplitz Matrices and Applications

The matrix reordering problem arises naturally in many applications [1, 2, 3]. The goal is to recover latent regularity and structural patterns contained in the noisy disordered data. When viewed through the lens of model (I.1), many of the applications involve reordering a noisy disordered matrix with some underlying monotone and Toeplitz structure, that is,

$$\Theta = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 & \dots & \theta_{n-1} \\ \theta_1 & \theta_0 & \theta_1 & \dots & \theta_{n-2} \\ \theta_2 & \theta_1 & \theta_0 & \dots & \theta_{n-3} \\ \vdots & & & \ddots & \\ \theta_{n-1} & \theta_{n-2} & \theta_{n-3} & \dots & \theta_0 \end{bmatrix}, \quad (\text{I.2})$$

where $\theta_0 \geq \theta_1 \geq \theta_2 \geq \dots \geq \theta_{n-1}$. The following are two specific examples.

Example 1 (Pseudotemporal ordering of single cells): Single-cell analysis promises to revolutionize the treatment of common and rare diseases and provides insights into some of the most fundamental processes in biology. As an important problem in single-cell biology, pseudotemporal cell ordering aims to determine the pattern of a dynamic process experienced by cells and then arrange cells according to their progression through the process, based on single-cell RNA sequencing data collected at multiple time points. This problem can be formulated as a noisy matrix reordering problem (I.1), where the entries of Θ represent true cell-to-cell similarities of n cells in their transcriptomic profiles. For studies involving cells undergoing a dynamic process such as differentiation, monotonic patterns are often observed in the gene expression of the cells along the progression path [4, 5, 6]. In such cases, the true similarity matrix Θ for the well-ordered cells can be modelled by a symmetric monotone Toeplitz matrix

as in (I.2), and the goal is to recover the total order of the cells based on the noisy and disordered measurements \mathbf{Y} of cell-to-cell similarities. For example, [7] considered a power decaying Toeplitz matrix with parameters $\theta_i = \alpha^i$ for some $\alpha \in (0, 1)$, $i = 1, 2, \dots, n - 1$, to infer the underlying pseudotemporal ordering. See also Section VII for more detailed discussion and the analyses of two real datasets.

Example 2 (Genome assembly): In metagenomics and bioinformatics, genome assembly refers to the process of taking a large number of short DNA sequences and putting them back together in correct order to reconstruct the original sequence. In particular, the task of assembling a draft genome from shotgun metagenomic sequencing data can be treated as a noisy matrix reordering problem (I.1), where each entry of Θ characterizes the true genome distance between a pair of contigs, or fragments of DNA sequence in the target genome. Specifically, for a set of n ordered contigs on one arm of the circular chromosome, their true pairwise genome distance matrix may be well approximated by a symmetric monotone Toeplitz matrix as in (I.2), and our goal is to recover the original genome order of these contigs based on the noisy and disordered measurements \mathbf{Y} of their pairwise genome distances. Among existing works, [8] considered a graph-based model, which assumed a Hamiltonian path structure, or tridiagonal Toeplitz distance matrix (defined below) for the n contigs on one arm; [9, 10] considered a linear monotone model for the contigs on one arm, which implies a linear decaying Toeplitz matrix (defined below) for the pairwise distance among these contigs.

The matrix reordering problem also has important applications in combinatorial exploratory data analysis and data visualization [3, 2]. Specifically, for a given data matrix of interest, a proper reordering of its columns and rows may bring forth a more informative representation with structural patterns directly accessible or even visible to the analysts, thus providing critical guidance for downstream analysis [11, 12]. The importance of such a task in data analysis may be partially reflected by the current availability of a large variety of matrix visualization tools with an automatic matrix reordering option in standard statistical softwares, such as R. Among them, heatmap [13] and corrplot [14] are probably the most commonly used tools.

Motivated by these interesting applications, this paper takes model (I.1) as a prototype underlying various matrix reordering problems and focuses on reordering noisy symmetric matrices with latent monotone and Toeplitz structures. In particular, we consider the simple noise structure with independent sub-Gaussian entries up to symmetry. Throughout, for any matrix Θ admitting the expression (I.2), we call the entries corresponding to the value θ_0 the main diagonals of Θ , and the entries corresponding to the value θ_i for $i \in \{1, \dots, n - 1\}$ the i -th principal diagonals of Θ .

In this paper, we consider the class of *ridged monotone Toeplitz matrices* defined by

$$\mathcal{T}_n = \left\{ \Theta \in \mathbb{R}^{n \times n} : \begin{array}{l} \Theta \text{ admits the expression (I.2)} \\ \theta_1 \geq \theta_2 \geq \dots \geq \theta_{n-1} \geq 0, \quad \theta_1 - \theta_{\lceil n/2 \rceil} \geq \theta_{\lceil n/2 \rceil} - \theta_{n-1} \end{array} \right\}. \quad (\text{I.3})$$

Note that the above definition puts no restriction on the main diagonal entries θ_0 – this is because for any $\Theta \in \mathcal{T}_n$, the main diagonal entries of $\Pi\Theta\Pi^\top$ are invariant to the permutation Π , and therefore does not contain any information that helps for matrix reordering. The nonnegativity condition $\theta_i \geq 0$ and the direction of monotonicity are not essential here (Section VIII). The "ridge" condition $\theta_1 - \theta_{\lceil n/2 \rceil} \geq \theta_{\lceil n/2 \rceil} - \theta_{n-1}$ ensures that the total amount of variations in the first $\lceil n/2 \rceil$ principal diagonals is no less than the variations in the rest of the diagonals. Such a characterization is required for technical reasons (Remark ??) but is in conformity with a wide range of applications. In particular, the class \mathcal{T}_n includes as special cases many interesting matrices that arise commonly in practice and have been discussed in different contexts. As a few examples, we note that \mathcal{T}_n includes,

- the tridiagonal Toeplitz matrices [15, 16, 17] where $\theta_1 > 0$ and $\theta_j = 0$ for all $j \geq 2$;
- the band monotone Toeplitz matrices [18] where $\theta_1 \geq \theta_2 \geq \dots \geq \theta_k = \dots = \theta_{n-1} = 0$ where $k < \lfloor n/2 \rfloor$;
- the linear decaying Toeplitz matrices [19] where $\theta_j = \alpha + \beta(n-j)$ for all $1 \leq j \leq n-1$ for some constants $\alpha, \beta > 0$; and
- the polynomial decaying Toeplitz matrices [20, 21] where $\theta_j = Mj^{-\beta}$ for all $1 \leq j \leq n-1$ for some constants $\beta, M \in (0, \infty)$.

B. Exact Matrix Reordering

Throughout, we identify a permutation matrix $\Pi \in \mathbb{R}^{n \times n}$ with its corresponding permutation π , as an element in the symmetric group \mathcal{S}_n . Suppose $\Pi \in \mathcal{S}_n$ is the underlying true permutation in $\Pi\Theta\Pi^\top$ and let $\Pi' \in \mathcal{S}_n$ be any given permutation. We quantify the distance between Π' and Π by the following 0-1 loss function

$$\tau_{\Theta}(\Pi, \Pi') = 1\{\Pi\Theta\Pi^\top \neq \Pi'\Theta\Pi'^\top\}. \quad (\text{I.4})$$

Note that $\Pi' = \Pi$ is only a special case of $\tau_{\Theta}(\Pi, \Pi') = 0$. The loss $\tau_{\Theta}(\Pi, \Pi') = 0$ if and only if the two permutations produce the same disordered matrix, allowing for $\Pi \neq \Pi'$. The loss function $\tau_{\Theta}(\Pi, \Pi')$ takes into account the possible equivalence classes among the permutations, caused by the specific structures of the signal matrix Θ . That is, we consider the exact permutation recovery modulo any inherent ambiguity caused by structure of the signal matrix. For example, when Θ is Toeplitz with distinct diagonal elements, then the loss function will identify any two permutations up to a complete reversal.

Let $\hat{\Pi}$ be any estimator of Π based on the observed matrix \mathbf{Y} . We define the estimation risk associated with the true parameters Θ and Π as

$$\mathbb{E}_{\Theta, \Pi}[\tau_{\Theta}(\hat{\Pi}, \Pi)] = P_{\Theta, \Pi}(\Pi\Theta\Pi^\top \neq \hat{\Pi}\Theta\hat{\Pi}^\top), \quad (\text{I.5})$$

where the expectation on the left-hand side and the probability measure on the right-hand side are both with respect to the random observation \mathbf{Y} for given (Θ, Π) .

To evaluate the performance of an estimator, we consider the probability of exact matrix reordering over a parameter space $\mathcal{T}'_n \times \mathcal{S}'_n = \{(\Theta, \Pi) : \Theta \in \mathcal{T}'_n, \Pi \in \mathcal{S}'_n\}$ for some subsets $\mathcal{T}'_n \subseteq \mathcal{T}_n$ and $\mathcal{S}'_n \subseteq \mathcal{S}_n$. In particular, we will identify sufficient and/or necessary conditions for parameter spaces of the form $\mathcal{T}'_n \times \mathcal{S}'_n$ such that, as $n \rightarrow \infty$, a given estimator could achieve exact matrix reordering uniformly over $\mathcal{T}'_n \times \mathcal{S}'_n$ with high probability. In this way, rigorous comparisons between various estimators can be made by comparing their respective conditions for exact matrix reordering. Specifically, for any $\mathcal{T}'_n \times \mathcal{S}'_n$, we define a hyper-parameter $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n)$ by

$$\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) = \inf_{\Theta \in \mathcal{T}'_n} \inf_{\substack{\Pi_1, \Pi_2 \in \mathcal{S}'_n \\ \Pi_1 \Theta \Pi_1^\top \neq \Pi_2 \Theta \Pi_2^\top}} \|\Pi_1 \Theta \Pi_1^\top - \Pi_2 \Theta \Pi_2^\top\|_F. \quad (\text{I.6})$$

Intuitively, the parameter $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n)$ quantifies how distinguishable two permuted versions of a matrix are within $\mathcal{T}'_n \times \mathcal{S}'_n$ – a larger value of $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n)$ indicates a potentially bigger contrast between any two permuted matrices and vice versa. As will be shown shortly, this hyper-parameter reflects the overall signal strength, and therefore the fundamental difficulty of reordering matrices over a given parameter space.

C. Main Results and Contributions

The main results of this paper can be summarized as follows.

- 1) (Rate-optimal estimator) For any subsets $\mathcal{T}'_n \subseteq \mathcal{T}_n$ and $\mathcal{S}'_n \subseteq \mathcal{S}_n$ such that $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \gtrsim \sigma\sqrt{n \log n}$, there exists a constrained least square estimator (LSE) that achieves exact matrix reordering for any $(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n$ with high probability.
- 2) (Fundamental information threshold) There exist some subsets $\mathcal{T}'_n \subset \mathcal{T}_n$ and $\mathcal{S}'_n \subset \mathcal{S}_n$ satisfying $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \asymp \sigma\sqrt{n \log n}$, such that no permutation estimator could achieve exact matrix reordering for all $(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n$ with high probability.
- 3) (Suboptimality of the spectral seriation algorithm) There exist some subsets $\mathcal{T}'_n \subset \mathcal{T}_n$ and $\mathcal{S}'_n \subset \mathcal{S}_n$ satisfying $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \asymp \sigma n^3$, such that the commonly used spectral seriation algorithm cannot achieve exact matrix reordering for any $(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n$ with high probability.
- 4) (An improved polynomial-time algorithm) For any subsets $\mathcal{T}'_n \subseteq \mathcal{T}_n$ and $\mathcal{S}'_n \subseteq \mathcal{S}_n$ such that $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \gtrsim \sigma n^2$, there exists a polynomial-time algorithm, proposed in Section IV, that achieves exact matrix reordering for any $(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n$ with high probability.

The results are illustrated in Figure 2. Parts 1) and 2) of the main results together show that a phase transition occurs at order $\sigma\sqrt{n \log n}$: no method can achieve the exact order recovery with high probability if the signal strength (as measured by ρ^*) is at or below this level and the constrained LSE recovers the order exactly with high probability whenever the signal strength is above this level.

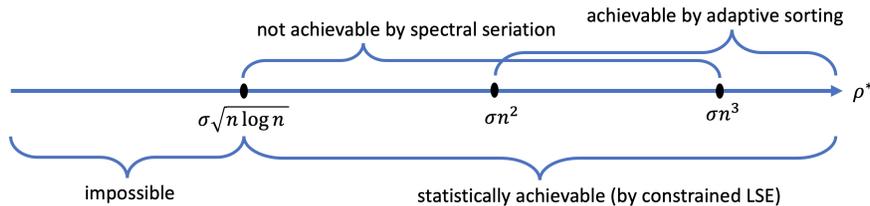


Fig. 2. A graphical illustration of the main theoretical results. The rates under the horizontal line are thresholds for $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n)$ for three different matrix reordering algorithms, constrained LSE, spectral seriation, and adaptive sorting, which are defined in Sections II-A, III, and IV, respectively.

Although the constrained LSE is rate-optimal, it requires solving an optimization over a potentially large permutation set \mathcal{S}'_n and is in general computationally infeasible for large n . Alternatively, polynomial-time algorithms have been used in practice. Among them, spectral seriation [22, 23, 24] is arguably the current state-of-art polynomial-time matrix reordering algorithm¹. We analyze the performance of the spectral seriation algorithm and show that it is suboptimal for reordering Toeplitz matrices as in \mathcal{T}_n .

We then develop a novel adaptive sorting algorithm, which runs in polynomial time, and show that it has strictly weaker signal strength requirement compared to the spectral seriation algorithm. This result indicates the advantage of the proposed adaptive sorting algorithm over the current state-of-art matrix reordering method, and explains its overall superior empirical performance over various existing methods across a wide range of simulation settings and real data examples. On the other hand, despite its numerical advantages, the adaptive sorting algorithm is still statistically suboptimal. We conjecture in Section V that there is a fundamental gap between statistical optimality and computational efficiency.

D. Related Works

Many statistical seriation problems that in one way or another aim to find an element in the discrete permutation set optimizing certain objective function have been studied from various aspects under different settings. These include the well-known consecutive one's problem [26, 27, 28] that dates back to the 1960s; the feature matching problem [29, 30, 31] and the noisy ranking problem [32, 33, 34, 35, 36, 37, 38, 39]; the matrix seriation problem for various shape-constrained matrices including the monotone or bi-monotone matrices [40, 41, 9, 42], the Robinson matrices [23, 24, 43, 44], and the Monge matrices [45]; and more recently, the seriation problem under the latent space models [46, 47].

Many of the existing works have focused on recovering the underlying permutations, estimation of the (disordered) signal structures, or both. However, about the matrix seriation problems, statistical limit and optimal procedures for the permutation recovery problem are relatively less-understood, compared to the estimation of the signal matrices

¹See Section VI for empirical evidences of the superiority of the spectral seriation over other existing methods.

[40, 41, 45, 9]. In particular, Bagaria et al. [8] considered the hidden Hamiltonian cycle recovery model, which is related to our model (I.1) under a tridiagonal signal matrix, and Ding et al. [48] considered a hidden $2k$ -nearest neighbor graph recovery model, corresponding to our model (I.1) under a band signal matrix. Both papers established the information-theoretical threshold for exact recovery and/or almost exact recovery, but under some related but in general different signal-to-noise ratio parameter. Moreover, these papers considered exact recovery over a special set of signal matrices (say, tridiagonal Toeplitz matrices) and the full permutation set \mathcal{S}_n alone, whereas our work considers exact recovery over all possible combinations of subsets of signal matrices and permutations, which may reveal fundamentally more difficult scenarios. Besides, for the general monotone Toeplitz matrix reordering problem as formulated in the current paper, most of the existing results only concern permutation recovery in the noiseless setting [23, 24, 43], thus leaving the fundamental behavior of the problem under the more realistic noisy observations largely unexplored.

E. Organization

The rest of the paper is organized as follows. We finish this section with notation that will be used throughout the paper. Section II establishes the fundamental statistical limit for exact matrix reordering, including the minimal signal strength required by the constrained LSE, and the matching fundamental information threshold. Section III analyzes the spectral seriation algorithm and shows its fundamental suboptimality. In Section IV, the adaptive sorting algorithm is proposed and its theoretical properties are investigated. Section V discusses the potential tradeoff between computational efficiency and statistical optimality. Section VI contains simulation studies that compare the empirical performances of several matrix reordering algorithms in various settings. Section VII presents the analyses of two real datasets, showing the advantage of the adaptive sorting in real-world applications. Possible extensions of the current work are discussed in Section VIII. The proofs of main results are given in Section IX and the proofs of other technical results are given in X.

F. Notation

For a vector $\mathbf{a} = (a_1, \dots, a_n)^\top \in \mathbb{R}^n$, $\text{diag}(a_1, \dots, a_n) \in \mathbb{R}^{n \times n}$ denotes the diagonal matrix whose i -th diagonal entry is a_i , and define the ℓ_p norm $\|\mathbf{a}\|_p = (\sum_{i=1}^n |a_i|^p)^{1/p}$. For a matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n \times n}$, its Frobenius norm is $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2}$; its i -th column is denoted by $\mathbf{A}_{\cdot i} \in \mathbb{R}^n$ and its i -th row by $\mathbf{A}_i \in \mathbb{R}^n$. Moreover, we denote $\mathbf{A}_{i,-j}$ as a subvector of \mathbf{A}_i with its j -th component removed. For any integer $n > 0$, we denote the set $[n] = \{1, 2, \dots, n\}$. For any $a > 0$, $\lfloor a \rfloor$ denotes the largest integer no greater than a , and $\lceil a \rceil$ denotes the smallest integer no less than a . For a finite set S , its cardinality is denoted by $|S|$. A random variable X is sub-Gaussian if there are positive constants C, v such that for every $t > 0$, we have $P(|X| > t) \leq C e^{-vt^2}$. For sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n = o(b_n)$ or $a_n \ll b_n$ if $\lim_n a_n/b_n = 0$, and write $a_n = O(b_n)$, $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if

there exists a constant C such that $a_n \leq Cb_n$ for all n . We write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $a_n \gtrsim b_n$. Throughout, C, C_1, C_2, \dots are universal constants independent of n , and can vary from place to place.

II. FUNDAMENTAL STATISTICAL LIMIT FOR MATRIX REORDERING

Our main result on the statistical limit for matrix reordering consists of two parts: a fundamental information threshold that benchmarks all the matrix reordering algorithms, and a constrained least square permutation estimator, whose performance is rate-optimal among all the estimators. We start with the rate-optimal estimator.

A. The Constrained Least Square Estimator

For any given parameter space $\mathcal{T}'_n \times \mathcal{S}'_n$ where $\mathcal{T}'_n \subseteq \mathcal{T}_n$ and $\mathcal{S}'_n \subseteq \mathcal{S}_n$, suppose one observes \mathbf{Y} from (I.1) for some $(\Theta^*, \Pi^*) \in \mathcal{T}'_n \times \mathcal{S}'_n$. A natural estimator for the unknown permutation is the constrained least square estimator (or the maximum likelihood estimator in the *i.i.d.* Gaussian case) over $\mathcal{T}'_n \times \mathcal{S}'_n$ defined through

$$(\hat{\Theta}^{lse}, \hat{\Pi}^{lse}) = \arg \min_{(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n} \|\mathbf{Y} - \Pi\Theta\Pi^\top\|_F^2. \quad (\text{II.1})$$

The following result provides the theoretical guarantee of $\hat{\Pi}^{lse}$ over $\mathcal{T}'_n \times \mathcal{S}'_n$.

Theorem 1 (Theoretical guarantee for constrained LSE): Under model (I.1), there exists some absolute constants $C, c > 0$ such that, for sufficiently large n , for any $\mathcal{T}'_n \subseteq \mathcal{T}_n$ and any $\mathcal{S}'_n \subseteq \mathcal{S}_n$ such that $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \geq C\sigma\sqrt{n \log n}$, the permutation estimator $\hat{\Pi}^{lse}$ given by (II.1) satisfies

$$\sup_{(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n} P_{\Theta, \Pi}(\hat{\Pi}^{lse} \Theta (\hat{\Pi}^{lse})^\top \neq \Pi\Theta\Pi^\top) \leq n^{-c}. \quad (\text{II.2})$$

Theorem 1 characterizes the explicit correspondence between the signal strength condition on ρ^* and the final exact recovery error probability; it applies to any subsets of ridged monotone Toeplitz matrices, any subsets of permutations, and the general sub-Gaussian noises. The theorem identifies a sufficient minimal signal strength condition

$$\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \gtrsim \sigma\sqrt{n \log n}, \quad (\text{II.3})$$

under which the constrained LSE $\hat{\Pi}^{lse}$ achieves exact matrix reordering uniformly over $(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n$ with high probability. Importantly, combined with the information lower bound obtained in Section II-B, Theorem 1 essentially implies that $\hat{\Pi}^{lse}$ is rate-optimal.

The proof of Theorem 1, detailed in Section IX-B, relies on analyzing the probability of exact matrix reordering for the constrained LSE at any given parameters $(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n$. To do so, we develop a general reduction scheme that connects the risk of matrix reordering to the risk of matrix denoising, i.e., estimating the permuted matrix $\Pi\Theta\Pi^\top$ from \mathbf{Y} , under model (I.1). We summarize our reduction scheme as the following proposition, proved in Section IX-A.

Proposition 1 (Reduction scheme): For any $\Theta, \hat{\Theta} \in \mathcal{T}_n$ and any permutations $\Pi, \hat{\Pi} \in \mathcal{S}'_n \subseteq \mathcal{S}_n$, it holds that

$$P(\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top \neq \Pi\Theta\Pi^\top) \leq \exp \left\{ 2\mathbb{E}\|\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top - \Pi\Theta\Pi^\top\|_F - \rho(\Theta, \mathcal{S}'_n) \right\}, \quad (\text{II.4})$$

where

$$\rho(\Theta, \mathcal{S}'_n) = \min_{\substack{\Pi_1, \Pi_2 \in \mathcal{S}'_n \\ \Pi_1\Theta\Pi_1^\top \neq \Pi_2\Theta\Pi_2^\top}} \|\Pi_1\Theta\Pi_1^\top - \Pi_2\Theta\Pi_2^\top\|_F.$$

Proposition 1 provides a tool for bounding the probability $P_{\Theta, \Pi}(\hat{\Pi}^{lse}\Theta(\hat{\Pi}^{lse})^\top \neq \Pi\Theta\Pi^\top)$ by analyzing the matrix denoising risk

$$\mathbb{E}\|\hat{\Pi}^{lse}\hat{\Theta}^{lse}(\hat{\Pi}^{lse})^\top - \Pi\Theta\Pi^\top\|_F, \quad (\text{II.5})$$

which is easier to handle. This reduction step paves the way for an asymptotically sharp risk analysis of $(\hat{\Theta}^{lse}, \hat{\Pi}^{lse})$ using powerful tools developed for general shape-constrained least square estimators. The key ingredients of our proof, which generalizes the idea for proving Theorem 3.1 of [40], include Chatterjee's variational formula (Lemma 3), an improved Dudley's integral inequality (Lemma 4), and a nontrivial calculation of the metric entropy of a set of permuted Toeplitz matrices (Lemma 5).

B. Fundamental Information Threshold and the Planted Path Reconstruction Problem

We investigate the necessity of the minimal signal strength condition (II.3), and uncover the fundamental information threshold underlying the matrix reordering problem. The following theorem provides a lower bound for the minimum signal strength.

Theorem 2 (Fundamental information threshold): Suppose $n \geq 48$. Then there exist subsets $\mathcal{T}'_n \subset \mathcal{T}_n$ and $\mathcal{S}'_n \subset \mathcal{S}_n$ satisfying $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) = 0.02\sigma\sqrt{n\log n}$ such that

$$\inf_{\hat{\Pi}} \sup_{(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n} P_{\Theta, \Pi}(\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top \neq \Pi\Theta\Pi^\top) \geq 0.6. \quad (\text{II.6})$$

Theorem 2 shows that, under the Gaussian noise there exists a certain parameter space $\mathcal{T}'_n \times \mathcal{S}'_n$ with minimal signal strength $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \asymp \sigma\sqrt{n\log n}$ such that no method could achieve exact matrix reordering uniformly with high probability. Theorems 1 and 2 together show that the condition (II.3) is asymptotically sharp and the rate

$$\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \asymp \sigma\sqrt{n\log n}, \quad (\text{II.7})$$

is the fundamental information threshold for the matrix reordering problem and the permutation estimator $\hat{\Pi}^{lse}$ given by (II.1) is minimax rate-optimal.

To prove the information lower bound in Theorem 2, we connect the matrix reordering problem with the following planted path reconstruction problem in graph-information theory, and obtain the information lower bound of the

former problem by analyzing that of the latter problem.

Definition 1 (Planted Hamiltonian path reconstruction): Consider a weighted undirected graph $G(V, E)$ with an adjacency matrix \mathbf{Z} , where \mathbf{Z} has *i.i.d.* standard normal entries up to symmetry. Suppose an arbitrary Hamiltonian path $P(\theta)$ of constant edge weight θ and length $|V| - 1$ connecting all the vertices in V , is added to $G(V, E)$, resulting to a new graph $G'(V, E')$. Then we refer problem of reconstructing P , or equivalently the recovery of the support of its adjacency matrix, from $G'(V, E')$, as the planted Hamiltonian path reconstruction problem.

To study the statistical limit of the above path reconstruction problem, we construct a least favorable class \mathcal{H} of Hamiltonian paths, each with a constant edge weight $\theta > 0$, such that (i) the class is sufficiently large in the sense that $\log |\mathcal{H}| \gtrsim n \log n$; and (ii) any two paths in \mathcal{H} are sufficiently distinct from each other under the Hamming distance (defined in Lemma 1). In particular, we show that whenever $\theta \lesssim \sqrt{\log n}$, there is no way to tell with confidence from the new graph $G'(V, E')$ which path in \mathcal{H} is planted in $G'(V, E')$. To construct the set \mathcal{H} , we introduce to the current context a useful result due to [49] concerning the Hamming packing in the permutation space [50, 51, 52].

Lemma 1 (Deza's bound on permutation packing): For any $\pi_1, \pi_2 \in \mathcal{S}_n$, we define their Hamming distance $d_H(\pi_1, \pi_2) = |\{i : \pi_1(i) \neq \pi_2(i)\}|$. A d -packing in the finite metric space (\mathcal{S}_n, d_H) is a subset $M \subset \mathcal{S}_n$ such that its elements are at a distance of at least d from each other. Then the largest cardinality of a d -packing $\beta_n(d)$ satisfies

$$\beta_n(d) \geq \frac{n!}{V_d}, \quad (\text{II.8})$$

where

$$V_d = \sum_{k=0}^d \binom{n}{k} k! \sum_{x=0}^k \frac{(-1)^x}{x!}. \quad (\text{II.9})$$

Consequently, for $2 \leq d \leq n - 1$, we have

$$\beta_n(d) \geq \frac{n!}{(n-d)!} \cdot \frac{(n-d+1)}{2(n-d)}. \quad (\text{II.10})$$

The detailed proof of Theorem 2 is provided in Section IX-C. The proof techniques developed there may be applied to obtain information lower bounds for other permutation related problems.

In this study, we define the fundamental information threshold to be the minimum separation $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n)$ required for any possible subset \mathcal{T}'_n of ridged monotone matrices, and any subset \mathcal{S}'_n of permutations, in order for the exact recovery to be achievable. This is different from the existing work such as [8] and [48] where the separation is considered for a specific subset of signal matrices such as tridiagonal Toeplitz matrices, and the full set \mathcal{S}_n of the permutations. In particular, from our proof of Theorem 2, one can show that when restricted to the settings of [8], a smaller separation condition $\rho^* \gtrsim \sigma \sqrt{\log n}$ may be obtained. The stronger requirement on the minimum separation

obtained by Theorem 2, is essentially due to the greater variety of scenarios allowed by our framework, some of which can be more difficult than those considered in the existing work. In this sense, our specific construction mainly reveals the fundamental difficulty caused by the unknown underlying permutation set. Finally, we remark that due to the nature of our minimax lower bound argument, it remains unclear if there exist many other hard scenarios, concerning possibly different signal matrices, under which a $\sigma\sqrt{n\log n}$ -order separation is needed. Nevertheless, we believe this is an important problem that deserves further investigation.

Remark 1: Our proof of Theorems 1 and 2 does not involve the “ridge” condition $\theta_1 - \theta_{n/2} \geq \theta_{n/2} - \theta_{n-1}$. However, such a condition plays an important role in our analysis of the proposed adaptive sorting algorithm (Section IV), and is satisfied by our suboptimality argument of the spectral seriation algorithm (Section III). Therefore, we include it in our definition of the parameter space for the integrity of our theoretical statements.

III. SUBOPTIMALITY OF SPECTRAL SERIATION

The constrained LSE introduced in Section II-A is rate-optimal. However, obtaining such an estimator requires solving an optimization over the discrete permutation set $\mathcal{S}'_n \subseteq \mathcal{S}_n$, which could be computationally infeasible, either because the subset \mathcal{S}'_n is unknown, or \mathcal{S}'_n contains a large number of permutations that grows exponentially fast as the matrix size n increases.

Alternatively, spectral approaches have been widely used for matrix reordering tasks. Among them, a spectral seriation method based on the Fiedler vector is particularly popular and has been extensively studied in the literature [22, 23, 24, 53, 43, 46]. In this section, we show that, despite the success of such a spectral seriation algorithm in many applications, it is nonetheless suboptimal for reordering Toeplitz matrices compared to the constrained LSE.

To formally introduce the spectral seriation estimator $\check{\Pi}$, we define the following ranking function.

Definition 2 (Ranking function): The ranking operator $\tau : \mathbb{R}^n \rightarrow \mathcal{S}_n$ is defined such that for any vector $\mathbf{x} \in \mathbb{R}^n$, $\tau(\mathbf{x})$ contains the ranks of the components of \mathbf{x} in increasing order. Whenever there are ties, increasing orders are assigned from left to right.

As an example, for a vector $\mathbf{x} = (2, 5, 1, 6, 2)^\top$, we have $\tau(\mathbf{x}) = (2, 4, 1, 5, 3)$. Following [23, 24, 43], the spectral seriation estimator $\check{\Pi}$ is then defined in Algorithm 1 below.

Algorithm 1 Spectral seriation

Input: Observed matrix $\mathbf{Y} = (Y_{ij})_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}$.

1. Compute the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{Y}$ where $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$ and $d_i = \sum_{j=1}^n Y_{ij}$.
2. Obtain the Fiedler eigenvector $\hat{\mathbf{v}} \in \mathbb{R}^n$ corresponding to the smallest nonzero eigenvalue of \mathbf{L} .

Output: $\check{\Pi} = [\tau(\hat{\mathbf{v}})]^{-1}$, where the inverse $[\cdot]^{-1}$ means the reversion of a permutation.

For matrix reordering, Algorithm 1 has been shown in [23] to achieve exact recovery in the noiseless case $\mathbf{Z} = 0$

for all Robinson matrices $\mathbf{R} = (r_{ij})_{1 \leq i, j \leq n}$ satisfying

$$\mathbf{R} = \mathbf{R}^\top, \quad r_{ij} \leq \min\{r_{ik}, r_{kj}\} \text{ for all } 1 \leq i < k < j \leq n, \quad (\text{III.1})$$

which contain Toeplitz matrices in \mathcal{T}_n as a special case. This result is summarized in the following proposition.

Proposition 2: For any $\mathbf{R} \in \mathbb{R}^{n \times n}$ satisfying (III.1) and $\Pi \in \mathcal{S}_n$, suppose the Fiedler eigenvector $\mathbf{v} \in \mathbb{R}^n$ associated to the smallest nonzero eigenvalue of $\Pi \mathbf{R} \Pi^\top$ contains n distinct components, then we have $\Pi = [\mathbf{v}]^{-1}$.

Proposition 2 essentially implies that, in the noiseless setting, the spectral seriation algorithm is able to achieve exact reorder recovery of any matrix $\Theta \in \mathcal{T}_n$ based on $\Pi \Theta \Pi^\top$.

However, the story is different in the noisy settings. Our analysis shows that the spectral seriation can be sensitive to the eigen-structure of the signal matrices and the noises. As a consequence, it may suffer from inconsistent estimation, and therefore significant suboptimality, due to insufficient separation between the Laplacian eigenvalues. Similar results concerning the suboptimality of spectral methods have been obtained in [8] under a hidden Hamiltonian cycle recovery model, but with a slightly different formulation and signal strength measure.

Theorem 3: Suppose the noise matrix has *i.i.d.* entries up to symmetry generated from $N(0, \sigma^2)$. Then there exists some $\mathcal{T}'_n \subseteq \mathcal{T}_n$ and $\mathcal{S}'_n \subseteq \mathcal{S}_n$ with $\log |\mathcal{S}'_n| \gtrsim n \log n$ satisfying $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) = C\sigma n^3$ for some absolute constant $C > 0$, such that

$$\lim_{n \rightarrow \infty} \inf_{(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n} P_{\Theta, \Pi}(\check{\Pi} \check{\Theta} \check{\Pi}^\top \neq \Pi \Theta \Pi^\top) \geq 1/2.$$

Theorem 3 shows that even when the minimal signal strength ρ^* is of order σn^3 , there still exists nontrivial cases over which the exact matrix reordering using the spectral seriation is impossible. The proof of Theorem 3, given in Section IX-D relies on a delicate eigenvector analysis of a deformed random Laplacian matrix. To this end, we develop a novel triangulation argument inspired by [54] and [55] that allows us to show inconsistency of the sample Fiedler vector $\hat{\mathbf{v}}$ in relation to the underlying true Fiedler vector, in the so-called subcritical regime [56, 57, 58]. Our analytic framework can be useful for other lower bound problems in statistics and random matrix theory, especially when they involve characterizing the asymptotic behavior of Laplacian eigenvectors associated to the bulk eigenvalues.

IV. EFFICIENT MATRIX REORDERING VIA ADAPTIVE SORTING

The suboptimality of the spectral seriation motivates us to develop an alternative algorithm with improved performance. We propose in this section a novel polynomial-time matrix reordering algorithm. The method, referred as the adaptive sorting, is summarized below in Algorithm 2.

Step 1 of Algorithm 2 identifies the location of the first (or equivalently, the last) row of the signal matrix after permutation. It uses the fact that the sum of the first or the last row of the original signal matrix Θ is the smallest

Algorithm 2 Adaptive sorting

Input: Observed matrix $\mathbf{Y} = (Y_{ij}) \in \mathbb{R}^{n \times n}$.

1. **Locating the initial element.** For given matrix \mathbf{Y} ,

(i) calculate $S_i = \sum_{j \in [n] \setminus \{i\}} Y_{ij}$ for $i \in [n]$;

(ii) set $\tilde{\pi}(1) = \arg \min_{i \in [n]} S_i$.

2. **Iterative sorting.** For $i = 1, 2, \dots, n-1$, set

$$\tilde{\pi}(i+1) = \arg \min_{j \in [n] \setminus \{\tilde{\pi}(1), \dots, \tilde{\pi}(i)\}} \|\mathbf{Y}_{\tilde{\pi}(i), -\tilde{\pi}(i)} - \mathbf{Y}_{j, -j}\|_1,$$

where $\mathbf{Y}_{i, -i} \in \mathbb{R}^{n-1}$ is the i -th row of \mathbf{Y} with i -th component removed.

Output: $\tilde{\pi} = (\tilde{\pi}(1), \dots, \tilde{\pi}(n))$, and $\tilde{\Pi}$ as the corresponding permutation matrix.

among all the row sums. Once a beginning point of n elements in their original order has been identified, Step 2 starts from there and builds up the complete permutation map $\tilde{\pi}$ iteratively based on the following rationale: after removing the main diagonals, each row of a monotone Toeplitz matrix is more similar to its nearby rows than the more distant rows. In particular, it can be shown that any reduced row (main diagonal removed) of the correctly ordered signal matrix, say $\Theta_{i, -i}$, has its nearest ℓ_1 -neighbors the rows $\Theta_{j, -j}$ for $j \in \{i-1, i+1\}$. Thus, for given $\tilde{\pi}(i)$, to identify $\tilde{\pi}(i+1)$, we look among the remaining rows $\{\mathbf{Y}_{j, -j}\}_{j \in [n] \setminus \{\tilde{\pi}(1), \dots, \tilde{\pi}(i)\}}$ for the one that minimizes $\|\mathbf{Y}_{\tilde{\pi}(i), -\tilde{\pi}(i)} - \mathbf{Y}_{j, -j}\|_1$. This is applied iteratively until all the rows (or columns) are properly ordered.

Related to the above reasoning, the application of the ℓ_1 -norm for comparing the reduced rows $\{\mathbf{Y}_{i, -i}\}_{1 \leq i \leq n}$ in Step 2 of Algorithm 2 is rooted in the bias-variance tradeoff. For example, although the nearest neighbor of a given reduced row $\Theta_{i, -i}$ can be invariably determined based on either the ℓ_1 - or the ℓ_2 -norm, the variance of the observed ℓ_2 -distance $\|\mathbf{Y}_{\tilde{\pi}(i), -\tilde{\pi}(i)} - \mathbf{Y}_{j, -j}\|_2$ can be much larger than the variance of $\|\mathbf{Y}_{\tilde{\pi}(i), -\tilde{\pi}(i)} - \mathbf{Y}_{j, -j}\|_1$. From our analysis in Section IX-E, it can be seen that an inflated variability may significantly deteriorate the performance of the sorting algorithm.

The following theorem provides the theoretical guarantee for the adaptive sorting algorithm, whose proof can be found in Section IX-E.

Theorem 4 (Theoretical guarantee for adaptive sorting): Under model (I.1), there exists some absolute constants $C, c > 0$ such that, for sufficiently large n , for any $\mathcal{T}'_n \subseteq \mathcal{T}_n$ and any $\mathcal{S}'_n \subseteq \mathcal{S}_n$ such that $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \geq C\sigma n^2$, we have

$$\sup_{(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n} P_{\Theta, \Pi}(\tilde{\Pi} \tilde{\Theta} \tilde{\Pi}^\top \neq \Pi \Theta \Pi^\top) \leq n^{-c}.$$

Compared with the results from the previous sections, although the adaptive sorting algorithm requires a condition

$$\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \gtrsim \sigma n^2, \tag{IV.1}$$

that is till stronger than the optimal condition (II.3), it is nevertheless much weaker than that required by the spectral seriation, by a factor at least of order n . In general, the adaptive sorting algorithm takes advantage of the Toeplitz

structure to achieve better performance in reordering matrices in \mathcal{T}_n , whereas the spectral seriation may have wider applicability when signal-to-noise ratio is sufficiently large. As for the fundamental limit of the adaptive sorting, in Section X-F below, we show that there exist $\mathcal{T}' \subset \mathcal{T}_n$ and $\mathcal{S}'_n \subset \mathcal{S}_n$ satisfying $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \asymp \sigma n^{3/2}$ over which the adaptive sorting algorithm does not always work, that is,

$$\sup_{(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n} P_{\Theta, \Pi}(\tilde{\Pi} \Theta \tilde{\Pi}^\top \neq \Pi \Theta \Pi^\top) \geq 0.2. \quad (\text{IV.2})$$

This information lower bound suggests that a signal condition stronger than (II.3) is also necessary for the adaptive sorting to perform well.

Remark 2: From our theoretical analysis, it can be seen that the adaptive sorting algorithm may actually perform well in cases beyond the ridged monotone Toeplitz class \mathcal{T}_n considered in this paper. For example, it is shown in Section IX-E (Proposition 4) that exact matrix reordering can be achieved for all the monotone Toeplitz matrices of the form (I.2) satisfying $\theta_1 - \theta_{\lceil n/2 \rceil} \geq C\sigma n$ for some positive constant C , regardless what is the underlying permutation Π . See also Section VIII for possible extensions.

V. INTERPLAY BETWEEN COMPUTATIONAL EFFICIENCY AND STATISTICAL ACCURACY

Theorems 1 to 4 altogether suggest a fundamental information gap between the statistically optimal procedure and the computationally efficient algorithms. On the one hand, the constrained LSE has asymptotically the weakest signal strength requirement, but can be computationally infeasible. On the other hand, the adaptive sorting or the spectral seriation algorithm has polynomial running time but requires a strictly stronger signal strength. Thus, it remains unclear whether the existing gap between the statistical optimal procedure and the computationally efficient procedure is essential and unsurpassable, or it can be reduced, or even closed by devising a better polynomial-time algorithm.

The tradeoff between computational efficiency and statistical accuracy has been observed in other permutation-related statistical problems such as sparse/submatrix detection [59, 60], structured PCA [61, 62], permuted isotonic regression [41, 42], tensor spectral clustering [63], among many others. In particular, assuming the computational hardness of the well-known planted clique problem, many of these problems [59, 62, 60, 42, 63] have been shown to preserve a regime with fundamental computational barrier; that is, any randomized polynomial-time algorithm must be statistically suboptimal.

In light of these existing work, it is of interest to prove or disprove the existence of any polynomial-time algorithm that succeeds over the region

$$\sigma \sqrt{n \log n} \lesssim \rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \lesssim \sigma n^2. \quad (\text{V.1})$$

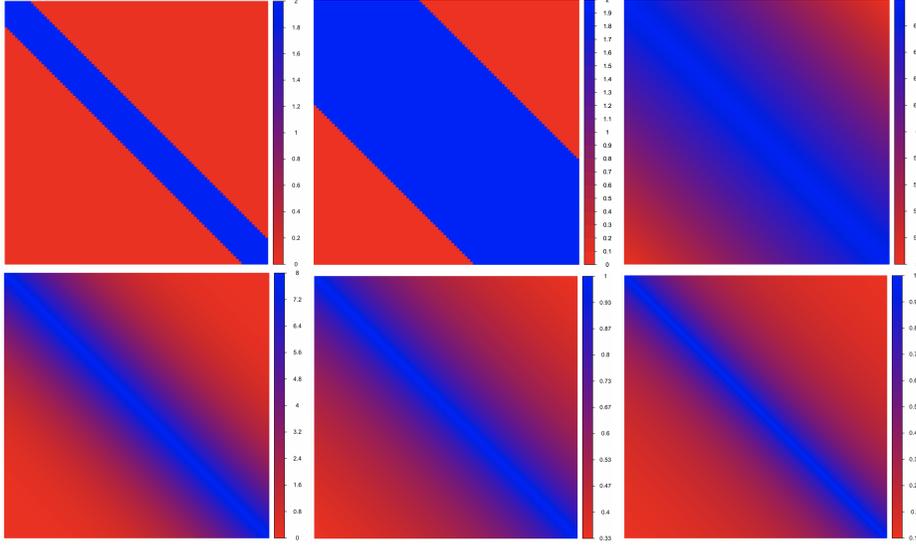


Fig. 3. Visualizations of six true signal matrices. Settings 1 to 6 arranged by row.

Solving this problem requires a quite different set of tools and we leave this fundamental and challenging problem for future investigation.

VI. NUMERICAL STUDIES

In this section, we evaluate the empirical performance of the proposed adaptive sorting algorithm and compare it with several existing matrix reordering methods.

We first set the dimensionality $n = 100$ for each matrix, and generate the noise matrix with *i.i.d.* entries from either a Gaussian distribution $N(0, \sigma^2)$, or a heavier-tailed Laplace distribution $Lap(0, \sigma)$. To better assess the range of applicability of different methods, we consider the following six settings for the underlying true signal matrix (Figure 3), containing band, linear decaying and nonlinear decaying monotone Toeplitz matrices:

(1) Narrow-band matrix: $\Theta \in \mathbb{R}^{n \times n}$ is Toeplitz of the form (I.2), where $\theta_1 = \theta_2 = \dots = \theta_{10} = 2$, and $\theta_i = 0$ for all $i \geq 11$; (2) Wide-band matrix: $\Theta \in \mathbb{R}^{n \times n}$ is Toeplitz of the form (I.2), where $\theta_1 = \theta_2 = \dots = \theta_{40} = 2$, and $\theta_i = 0$ for all $i \geq 41$; (3) Linear decaying matrix: $\Theta \in \mathbb{R}^{n \times n}$ is Toeplitz of the form (I.2), where $\theta_i = 5 + 0.02 \cdot (n - i)$ for all $i \in [n - 1]$; (4) Polynomial decaying matrix: $\Theta \in \mathbb{R}^{n \times n}$ is Toeplitz of the form (I.2), where $\theta_i = [(n - i) \cdot 0.02]^3$ for all $i \in [n - 1]$; (5) Inverse linear decaying matrix: $\Theta \in \mathbb{R}^{n \times n}$ is Toeplitz of the form (I.2), where $\theta_i = [1 + 0.02 \cdot i]^{-1}$ for all $i \in [n - 1]$; (6) Inverse polynomial decaying matrix: $\Theta \in \mathbb{R}^{n \times n}$ is Toeplitz of the form (I.2), where $\theta_i = [1 + 0.02 \cdot i]^{-2}$ for all $i \in [n - 1]$.

For each setting, we evaluate the performance of various methods by comparing their ability to achieve exact order recovery across a range of noise levels. The underlying permutations are generated uniformly from the permutation group \mathcal{S}_n . The performance of each algorithm is measured by the empirical proportion of failures in exact order

recovery (as determined by the loss function $\tau_{\Theta}(\Pi, \Pi')$, where we do not distinguish between the true permutation and its complete reversal) over 500 rounds of simulations at each noise level σ .

In addition to the adaptive sorting algorithm ("AS") proposed in Section IV and the spectral seriation algorithm ("SS") defined in Section III, we also evaluate the following five existing matrix reordering methods:

- Best permutation analysis ("BP") proposed by [64].
- The multidimensional scaling based method ("MDS") implemented by the function `seriate` in the R package `seriation`, with the argument option `method="MDS"` [25].
- The rank-two ellipse seriation algorithm ("R2E") proposed by [65], and implemented by the function `seriate` in the R package `seriation`, with the option `method="R2E"`.
- The "VAT" (visual assessment of tendency) algorithm proposed by [66], and implemented by the function `seriate` in the R package `seriation`, with the option `method="VAT"`.
- The normalized spectral seriation algorithm ("SS.n"), which differs from the spectral seriation algorithm only in its definition of the Laplacian matrix $\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-1}\mathbf{Y}$.

The numerical results for both noise settings are presented in Figures 4 and 5. Overall, the proposed AS algorithm demonstrates the best performance across all settings, followed by SS in most cases. In particular, AS outperforms other methods most significantly for the band Toeplitz matrices (Settings 1 and 2). Under the other four settings with linear or nonlinear decaying diagonals, AS, SS, and often BP exhibit relatively better performance compared to the other methods. Among these three methods, AS consistently performs better than SS and BP in the nonlinear decaying cases (Settings 4 to 6), while in the strict linear decaying case (Setting 3), SS demonstrates the best performance, followed by AS.

To further assess the relative performance for large matrices, we repeated the experiments under Gaussian noise with $n = 1000$. Notably, BP was excluded from this comparison due to its lack of scalability for large n , as the algorithm requires n^2 iterations, and in each iteration, the determinant of a matrix up to size $n \times n$ must be evaluated. Our results in Figure 6 suggest similar trends to those observed for $n = 100$, highlighting the superior performance of AS in most settings and the best performance achieved by SS in Setting 3. Comparing Figures 4 and 6, particularly the values of σ at which the phase transition occurs, we observe that, in line with our theory, the performance of AS and SS deteriorates as n increases from 100 to 1000.

The simulation studies demonstrate the overall superiority and adaptivity of the proposed AS algorithm over the other six alternative methods. In particular, the numerical results indicate SS to be the overall best existing method, whose empirical performance is in turn dominated by the proposed AS algorithm in most cases. This phenomenon is consistent with our theoretical analysis of the two methods, showing the strength and practical relevance of the theoretical results developed in the preceding sections.

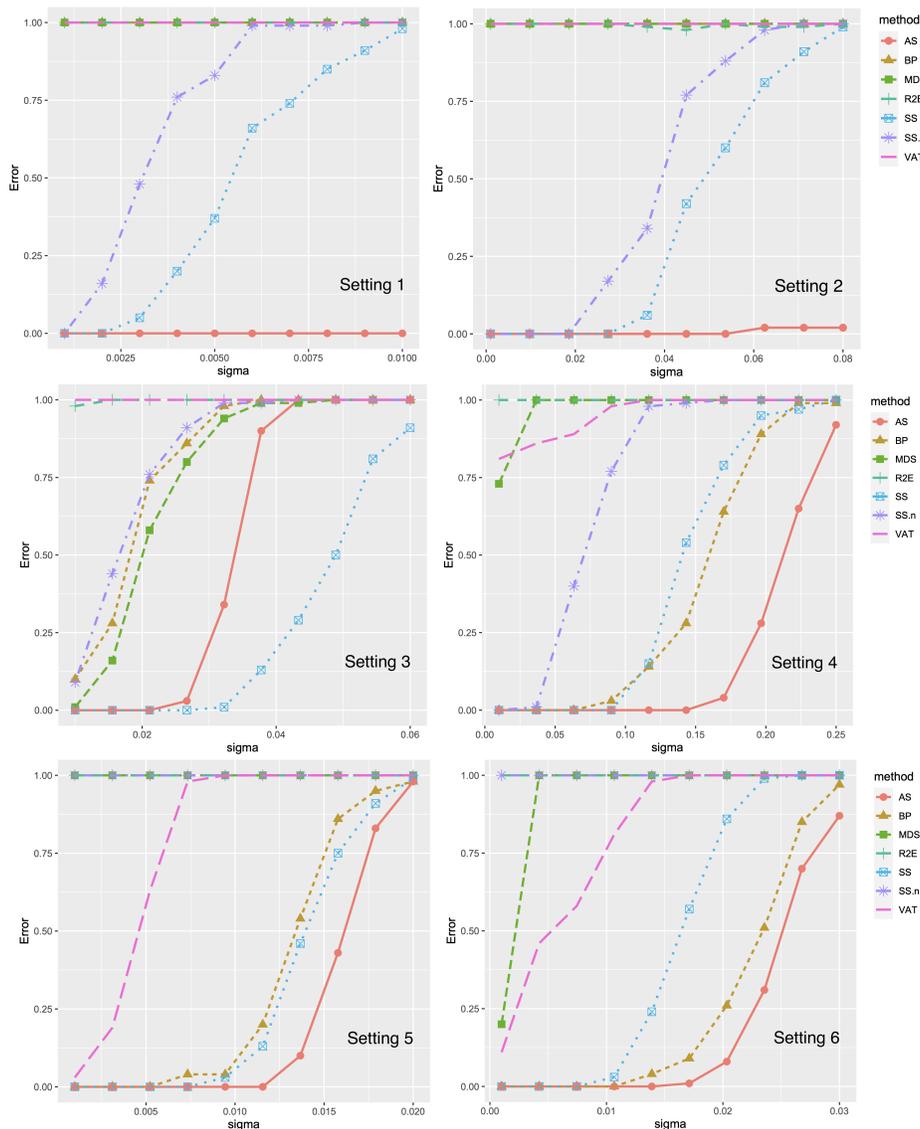


Fig. 4. Comparison of seven matrix reordering methods under the Gaussian noise with $n = 100$ and variance σ^2 . The errors quantify the empirical probability of failures in exact order recovery over 500 rounds of simulations. AS: adaptive sorting, BP: best permutation analysis, MDS: multidimensional scaling based seriation, R2E: rank-two ellipse seriation, SS: spectral seriation, SS.n: normalized spectral seriation, VAT: visual assessment of tendency.

VII. APPLICATION TO TWO REAL DATASETS

We analyze two real single-cell RNA sequencing datasets, and compare the performance of the adaptive sorting and the spectral seriation algorithms for inferring the latent pseudotemporal orders of single cells.

The first dataset contains single-cell mRNA sequencing reads for 372 primary human skeletal muscle myoblasts undergoing differentiation [4]. Specifically, primary human myoblasts were cultured in high-serum medium; after switching to low-serum medium that induces differentiation, the cells were dissociated and individually captured at 24-h intervals (0, 24, 48 and 72 h), and each cell was sequenced to obtain the final mRNA reads. As a result, each of the four time points contains about 90 cells. Due to possible variations in the speed of differentiation across

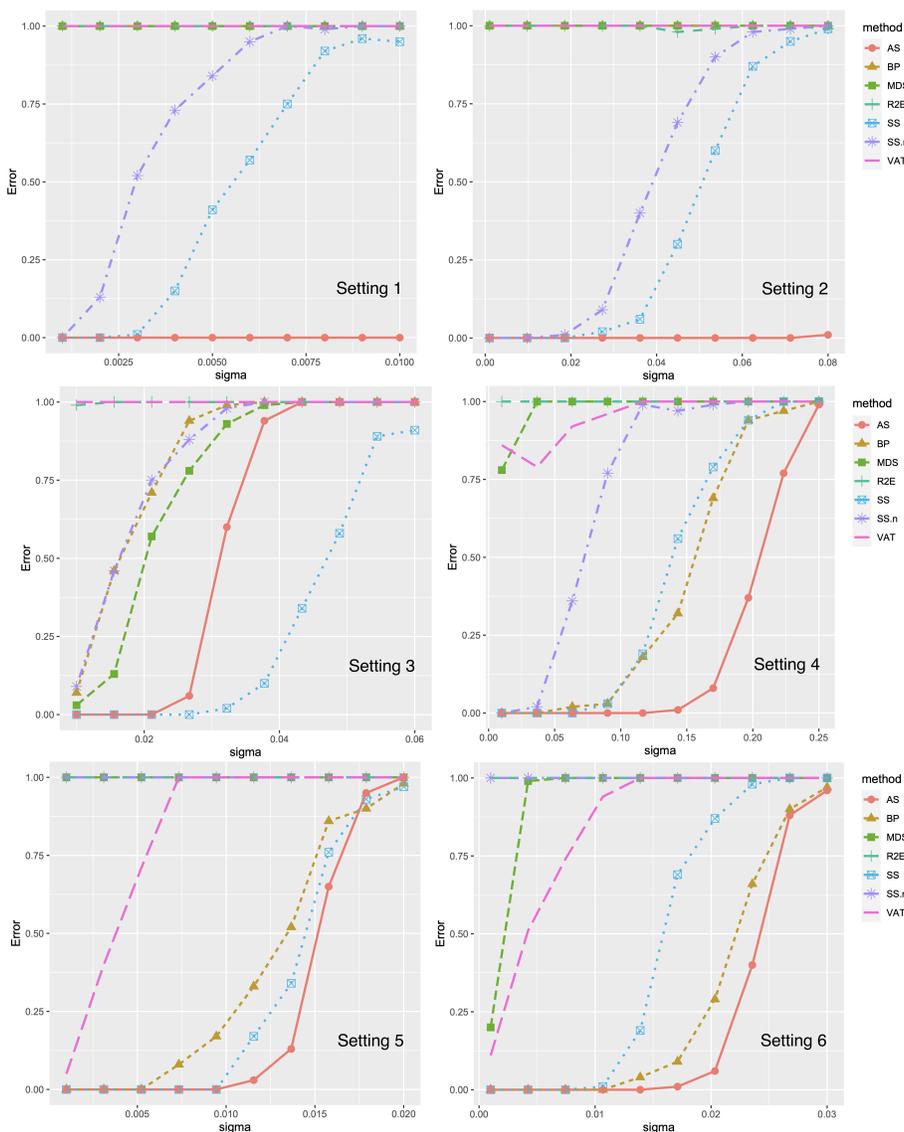


Fig. 5. Comparison of seven matrix reordering methods under the Laplacian noise with $n = 100$ and scale parameter σ (i.e., with variance $2\sigma^2$).

the cells, we expect the cells to be approximately uniform-distributed along the progression path, whose order may be recovered by the matrix reordering algorithm. The raw count data were preprocessed and normalized using the functions `CreateSeuratObject` and `NormalizeData` in the R package `Seurat`² under default settings. We applied the functions `FindVariableFeatures` and `ScaleData` in `Seurat` to identify and standardize the levels of $p = 1500$ most variable genes for subsequent analysis. Then a pairwise similarity matrix across the 372 cells was calculated as $\mathbf{Y} = c\mathbf{J}_n - \mathbf{D}$, where \mathbf{J}_n is an all-one matrix, c is a sufficiently large constant making \mathbf{Y} nonnegative (e.g., $c = \|\mathbf{D}\|_\infty$), and \mathbf{D} is the pairwise Euclidean distance matrix of the r -dimensional spectral embeddings of the cells, based on the singular value decomposition of the standardized data. Specifically, after

²<https://cran.r-project.org/web/packages/Seurat/index.html>

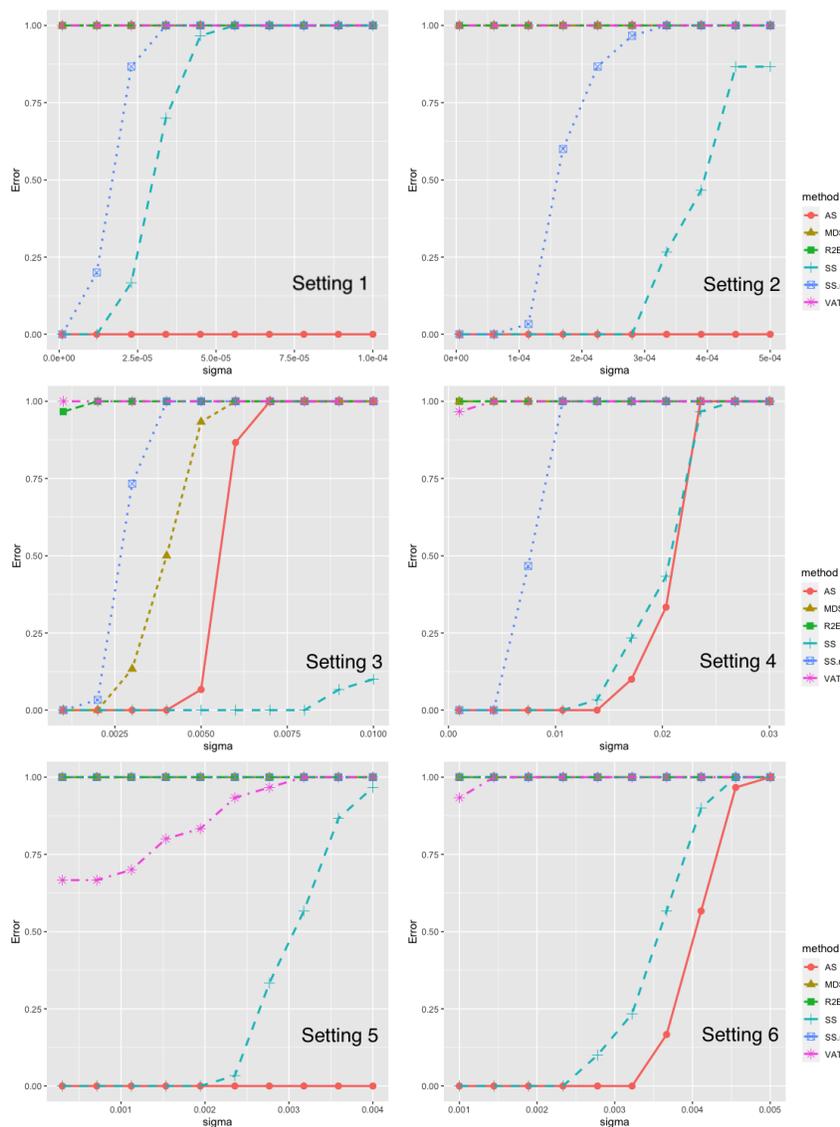


Fig. 6. Comparison of six matrix reordering methods (“BP” removed due to lack of scalability) under the Gaussian noise with $n = 1000$ and variance σ^2 .

obtaining the standardized dataset, which contains expression levels of p genes for n cells, we apply SVD to the data matrix, say $X \in \mathbb{R}^{n \times p}$, and define the r -dimensional embedding as the leading r left singular vectors of X weighted by their associated singular values. Finally, we applied the adaptive sorting (AS) and the spectral seriation (SS) to reorder the similarity matrix Y , which gives the inferred temporal order of cells, up to a possible reversion. To evaluate the performance of two methods, we compared the inferred temporal orders with the true order using Spearman’s rho statistic. On the left of Figure 7, we show a boxplot of Spearman’s rho statistics evaluated over various values for $r \in \{2, 3, \dots, 20\}$. Our AS algorithm shows clear advantages over SS in terms of the preciseness of the inferred temporal orders, even though the thus constructed similarity matrices have possibly dependent entries. In particular, the evaluation results were consistent for different choices of $p \in \{500, 1500, 2500\}$.

The second dataset consists of single-cell RNA sequencing reads for 149 human primordial germ cells ranging from 4 weeks to 19 weeks old [67]. Specifically, there were 6 cells of 4 weeks old, 37 cells of 7 weeks old, 20 cells of 10 weeks old, 27 cells of 11 weeks old, and 57 cells of 19 weeks old. The RNA-seq data were preprocessed and normalized using the same procedure, leading to a similarity matrix across the 149 cells calculated from the r -dimensional spectral embeddings of the cells. On the right of Figure 7, we have a boxplot for the Spearman’s rho statistics evaluated for different r values ($r \in \{2, 3, \dots, 20\}$), again indicating AS to be overall much better than SS for inferring temporal orders of single cells. Like the previous example, we also found the results to be consistent over different choices of $p \in \{500, 1500, 2500\}$.

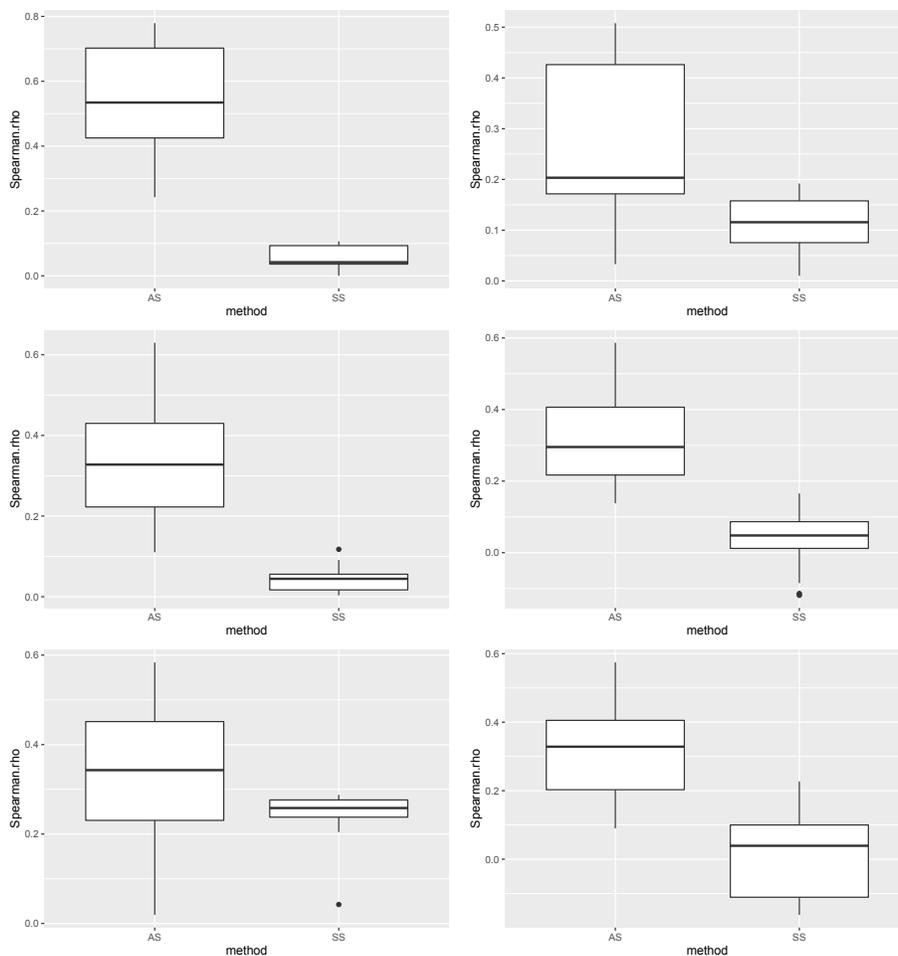


Fig. 7. Comparison of adaptive sorting (AS) and spectral seriation (SS) for inferring temporal orders of single cells using Spearman’s rho when $p = 500$ (top), $p = 1500$ (middle), and $p = 2500$ (bottom). Left: study of primary human myoblasts [4]. Right: study of human primordial germ cells [67].

VIII. DISCUSSION

This paper studied the matrix reordering problem for a large class of monotone Toeplitz matrices characterized by \mathcal{T}_n , where the diagonals decay as they deviate from the main diagonal. The analysis can be easily extended to

the class of monotone Toeplitz matrices where the diagonals increase as they deviate from the main diagonal (i.e., $\theta_1 \leq \theta_2 \leq \dots \leq \theta_{n-1}$). Specifically, on the one hand, since the direction of monotonicity is not essential in both the lower and the upper bound arguments, the analysis of the fundamental limit of matrix reordering as presented in Section II may be adopted with minor changes to obtain similar results. On the other hand, a computationally efficient algorithm can be constructed by slightly modifying the adaptive sorting algorithm in Section IV with $\tilde{\pi}(1)$ in the initialization step given by $\tilde{\pi}(1) = \arg \max_{i \in [n]} S_i$.

In Section II-A, we essentially reduced the problem of matrix reordering to an estimation problem, about which we obtained sharp upper bound for the estimation risk. In particular, our proofs of Theorem 1 and Proposition 1 together yielded the risk upper bound

$$\sup_{\mathcal{T}'_n \times \mathcal{S}'_n \subseteq \mathcal{T}_n \times \mathcal{S}_n} \sup_{(\Theta^*, \Pi^*) \in \mathcal{T}'_n \times \mathcal{S}'_n} \mathbb{E} \|\hat{\Pi}^{lse} \Theta^* (\hat{\Pi}^{lse})^\top - \Pi^* \Theta^* (\Pi^*)^\top\|_F \leq C\sigma\sqrt{n \log n}. \quad (\text{VIII.1})$$

The above rate of convergence is in fact minimax optimal as one can show using a similar argument as in the proof of Theorem 2 that

$$\inf_{\hat{\Pi}} \sup_{\mathcal{T}'_n \times \mathcal{S}'_n \subseteq \mathcal{T}_n \times \mathcal{S}_n} \sup_{(\Theta^*, \Pi^*) \in \mathcal{T}'_n \times \mathcal{S}'_n} \mathbb{E} \|\hat{\Pi} \Theta^* \hat{\Pi}^\top - \Pi^* \Theta^* (\Pi^*)^\top\|_F \geq C'\sigma\sqrt{n \log n}. \quad (\text{VIII.2})$$

In particular, to prove (VIII.1) there is no requirement on the minimum separation, unlike Theorem 1 which requires $\rho^* \gtrsim \sigma\sqrt{n \log n}$. Nevertheless, Equations (VIII.1) and (VIII.2) imply a minimax rate of convergence of order $\sigma\sqrt{n \log n}$, indicating a close connection between the exact recovery problem under the loss τ_Θ and the estimation problem under the loss $|\hat{\Pi} \Theta \hat{\Pi}^\top - \Pi \Theta \Pi^\top|_F$. However, our analyses of the spectral seriation algorithm and the adaptive sorting algorithm cannot be directly adapted to obtain an estimation result.

In addition to monotone Toeplitz matrices, it is important to extend the current theoretical framework to other matrix classes that are relevant in practice, such as bimonotone matrices [41], circulant matrices [47], Robinson matrices [23], and Monge matrices [45]. Furthermore, compared to the exact recovery problem considered in this study, evaluating performance under partial recovery criteria [10] can be less restrictive and potentially more interesting, albeit more challenging. We plan to investigate this problem systematically in a future study.

This paper considered the prototypical setting where the noise matrix has independent, homoscedastic, and sub-Gaussian entries. However, noise structures in real-world applications can be more complex. It is intriguing to consider matrix reordering in settings with dependent, heteroscedastic, and/or heavy-tailed observations. Furthermore, in terms of computationally efficient algorithms, in the absence of a computational lower bound, there may exist other algorithms with weaker separation requirements than the proposed AS algorithm. Although we are currently unaware of such algorithms, we recognize this as an interesting follow-up question and plan to explore it in future work. Some of the technical tools and theoretical results developed in this paper may be useful for solving

the problem in more intricate settings.

IX. PROOF OF MAIN RESULTS

In this section, we present the proofs of Proposition 1 and Theorems 1, 2, 3 and 4. The proofs of other theorems, propositions and technical lemmas are given in Section X.

A. Proof of Proposition 1

Note that for any $\Theta \in \mathcal{T}_n$, the simple inequality

$$1_{\{\widehat{\Pi}\Theta\widehat{\Pi}^\top \neq \Pi\Theta\Pi^\top\}} \leq \exp \left\{ \|\widehat{\Pi}\Theta\widehat{\Pi}^\top - \Pi\Theta\Pi^\top\|_F - \rho(\Theta, \mathcal{S}'_n) \right\}, \quad (\text{IX.1})$$

holds trivially for

$$\rho(\Theta, \mathcal{S}'_n) = \min_{\substack{\Pi_1, \Pi_2 \in \mathcal{S}'_n \\ \Pi_1\Theta\Pi_1^\top \neq \Pi_2\Theta\Pi_2^\top}} \|\Pi_1\Theta\Pi_1^\top - \Pi_2\Theta\Pi_2^\top\|_F.$$

On the other hand, we note that for any $\Theta_1, \Theta_2 \in \mathcal{T}_n$, and any permutation matrices $\Pi_1, \Pi_2 \in \mathcal{S}'_n$, it holds that

$$\|\Theta_1 - \Theta_2\|_F \leq \|\Pi_1\Theta_1\Pi_1^\top - \Pi_2\Theta_2\Pi_2^\top\|_F. \quad (\text{IX.2})$$

To see this, note that by the Toeplitz structure of Θ_1 and Θ_2 , one can identify Θ_1 and Θ_2 with two monotonic nondecreasing vectors of dimension n^2 , denoted as $\text{vec}(\Theta_1)$ and $\text{vec}(\Theta_2)$, respectively, by arranging the matrix entries in the following order: the main diagonals, the first principal diagonals, the second principal diagonals, etc.

As a consequence, for any $\Pi_1, \Pi_2 \in \mathcal{S}'_n$,

$$\begin{aligned} \|\Theta_1 - \Theta_2\|_F^2 &= \|\text{vec}(\Theta_1) - \text{vec}(\Theta_2)\|_2^2 \\ &= \|\text{vec}(\Theta_1)\|_2^2 + \|\text{vec}(\Theta_2)\|_2^2 - 2\text{vec}(\Theta_1)^\top \text{vec}(\Theta_2) \\ &\leq \|\text{vec}(\Pi_1\Theta_1\Pi_1^\top)\|_2^2 + \|\text{vec}(\Pi_2\Theta_2\Pi_2^\top)\|_2^2 - 2\text{vec}(\Pi_1\Theta_1\Pi_1^\top)^\top \text{vec}(\Pi_2\Theta_2\Pi_2^\top) \\ &= \|\text{vec}(\Pi_1\Theta_1\Pi_1^\top) - \text{vec}(\Pi_2\Theta_2\Pi_2^\top)\|_F^2 \\ &= \|\Pi_1\Theta_1\Pi_1^\top - \Pi_2\Theta_2\Pi_2^\top\|_F^2, \end{aligned}$$

where the third line follows from the permutation invariance of the ℓ_2 norm and the following elementary rearrangement inequality.

Lemma 2 (Rearrangement Inequality): If $a_1 \geq a_2 \geq \dots \geq a_n$ and $b_1 \geq b_2 \geq \dots \geq b_n$, then $a_n b_1 + \dots + a_1 b_n \leq a_{\sigma(1)} b_1 + \dots + a_{\sigma(n)} b_n \leq a_1 b_1 + \dots + a_n b_n$, where σ is any permutation in \mathcal{S}_n .

Thus for any $\Theta \in \mathcal{T}_n$ and $\Pi, \hat{\Pi} \in \mathcal{S}'_n$, we have

$$\begin{aligned} \|\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top - \Pi\Theta\Pi^\top\|_F &\leq \|\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top - \Pi\Theta\Pi^\top\|_F + \|\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top - \hat{\Pi}\Theta\hat{\Pi}^\top\|_F \\ &= \|\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top - \Pi\Theta\Pi^\top\|_F + \|\hat{\Theta} - \Theta\|_F \\ &\leq 2\|\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top - \Pi\Theta\Pi^\top\|_F, \end{aligned}$$

where the last inequality follows from the property (IX.2). Combining the above inequalities and taking expectations for both sides, we obtain the final inequality

$$P(\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top \neq \Pi\Theta\Pi^\top) \leq \exp\left\{2\mathbb{E}\|\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top - \Pi\Theta\Pi^\top\|_F - \rho(\Theta, \mathcal{S}'_n)\right\}. \quad (\text{IX.3})$$

B. Proof of Theorem 1

In the following, to simplify notation, we write $(\hat{\Theta}^{lse}, \hat{\Pi}^{lse})$ as $(\hat{\Theta}, \hat{\Pi})$, and denote the permutation maps associated to the permutation matrices $\hat{\Pi}$ and Π^* as $\hat{\pi} : [n] \rightarrow [n]$ and $\pi^* : [n] \rightarrow [n]$, respectively. We also identify a permutation matrix with the associated permutation map when there is no confusion. We will prove the following pointwise result, from which the uniform statement in Theorem 1 follows directly.

Theorem 5 (Pointwise guarantee): Under model (I.1), there exist some absolute constants $C, c > 0$ such that, for sufficiently large n , for any $\Theta^* \in \mathcal{T}'_n \subseteq \mathcal{T}_n$ and any $\mathcal{S}'_n \subseteq \mathcal{S}_n$ satisfying $\rho(\Theta^*; \mathcal{S}'_n) \geq C\sigma\sqrt{n\log n}$, the constrained LSE $(\hat{\Pi}, \hat{\Theta})$ defined over $(\mathcal{T}'_n, \mathcal{S}'_n)$ satisfies

$$P_{\Theta^*, \Pi^*}(\hat{\Pi}\hat{\Theta}^*\hat{\Pi}^\top \neq \Pi^*\Theta^*(\Pi^*)^\top) \leq \exp\{-c\sigma\sqrt{n\log n}\}, \quad (\text{IX.4})$$

for each $\Pi^* \in \mathcal{S}'_n$.

To prove Theorem 5, by the general reduction scheme (Proposition 1), it suffices to obtain the upper bound for the matrix denoising risk $\mathbb{E}\|\hat{\Pi}\hat{\Theta}\hat{\Pi}^\top - \Pi^*\Theta^*(\Pi^*)^\top\|_F \leq C\sigma\sqrt{n\log n}$ under fixed parameters (Θ^*, Π^*) . The key ingredient is Chatterjee's variational formula, originally developed in [68]. The following version is proved as Lemma A.1 in [40].

Lemma 3 (Chatterjee's variational formula): Let \mathcal{C} be a closed subset of \mathbb{R}^d . Suppose $y = a^* + z$ where $a^* \in \mathcal{C}$ and $z \in \mathbb{R}^d$. Let $\hat{a} \in \arg \min_{a \in \mathcal{C}} \|y - a\|_2^2$ be a projection of y onto \mathcal{C} . Define the function $f_{a^*} : \mathbb{R}_+ \rightarrow \mathbb{R}$ by

$$f_{a^*}(t) = \sup_{a \in \mathcal{C} \cap \mathbb{B}_d(a^*, t)} \langle a - a^*, z \rangle - \frac{t^2}{2}.$$

Then we have

$$\|\hat{a} - a^*\|_2 \in \arg \min_{t \geq 0} f_{a^*}(t).$$

Moreover, if there exists $t^* > 0$ such that $f_{a^*}(t) < 0$ for all $t \geq t^*$, then $\|\hat{a} - a^*\|_2 \leq t^*$.

We set $a^* = \Pi^* \Theta^* (\Pi^*)^\top$, $z = \mathbf{Z}$, $y = \mathbf{Y}$, $\mathcal{C} = \mathcal{P}(\mathcal{T}'_n, \mathcal{S}'_n) \equiv \{\mathbf{Q} \in \mathbb{R}^{n \times n} : \mathbf{Q} = \Pi \Theta \Pi^\top, \Theta \in \mathcal{T}'_n, \Pi \in \mathcal{S}'_n\}$, $\mathbb{B}(\Theta_0, t) = \{\Theta \in \mathbb{R}^{n \times n} : \|\Theta - \Theta_0\|_F \leq t\}$, and let $\langle \cdot, \cdot \rangle$ be the Hilbert-Schmidt inner product over symmetric matrices in $\mathbb{R}^{n \times n}$. Hence, by Lemma 3, we have

$$\|\widehat{\Pi} \widehat{\Theta} \widehat{\Pi}^\top - \Pi^* \Theta^* (\Pi^*)^\top\|_F \in \arg \min_{t \geq 0} \left\{ \sup_{\mathbf{Q} \in \mathcal{P}(\mathcal{T}'_n, \mathcal{S}'_n) \cap \mathbb{B}(\Pi^* \Theta^* (\Pi^*)^\top, t)} \langle \mathbf{Q} - \Pi^* \Theta^* (\Pi^*)^\top, \mathbf{Z} \rangle - \frac{t^2}{2} \right\}.$$

Now since $(\Pi^*)^\top \Pi^* = \text{Id}_n$, we have

$$\begin{aligned} & \sup_{\mathbf{Q} \in \mathcal{P}(\mathcal{T}'_n, \mathcal{S}'_n) \cap \mathbb{B}(\Pi^* \Theta^* (\Pi^*)^\top, t)} \langle \mathbf{Q} - \Pi^* \Theta^* (\Pi^*)^\top, \mathbf{Z} \rangle \\ &= t \cdot \sup_{\mathbf{Q} \in \mathcal{P}(\mathcal{T}'_n, \mathcal{S}'_n) \cap \mathbb{B}(t^{-1} \Pi^* \Theta^* (\Pi^*)^\top, 1)} \langle \mathbf{Q} - t^{-1} \Pi^* \Theta^* (\Pi^*)^\top, \mathbf{Z} \rangle \\ &= t \cdot \sup_{\mathbf{Q} \in \mathcal{P}(\mathcal{T}'_n, \mathcal{S}'_n) \cap \mathbb{B}(t^{-1} \Pi^* \Theta^* (\Pi^*)^\top, 1)} \langle (\Pi^*)^\top \mathbf{Q} \Pi^* - t^{-1} \Theta^*, (\Pi^*)^\top \mathbf{Z} \Pi^* \rangle \\ &= t \cdot \sup_{\mathbf{Q} \in \mathcal{P}(\mathcal{T}'_n, (\pi^*)^{-1} \circ \mathcal{S}'_n) \cap \mathbb{B}(t^{-1} \Theta^*, 1)} \langle \mathbf{Q} - t^{-1} \Theta^*, (\Pi^*)^\top \mathbf{Z} \Pi^* \rangle \\ &\leq t \cdot \sup_{\mathbf{H} \in \mathcal{M}_n} \langle \mathbf{H}, \mathbf{Z}' \rangle, \end{aligned}$$

where $\mathcal{M}_n \equiv \mathcal{M}(\mathcal{T}'_n, \mathcal{S}'_n) = \{\mathbf{Q} - t^{-1} \Theta^* : \mathbf{Q} \in \mathcal{P}(\mathcal{T}'_n, (\pi^*)^{-1} \circ \mathcal{S}'_n) \cap \mathbb{B}(t^{-1} \Theta^*, 1)\}$, $(\pi^*)^{-1} \circ \mathcal{S}'_n = \{(\Pi^*)^\top \Pi : \Pi \in \mathcal{S}'_n\}$ and $\mathbf{Z}' = (\Pi^*)^\top \mathbf{Z} \Pi^*$. In particular, we have $\mathcal{M}_n \subseteq \mathbb{B}(\mathbf{0}, 1)$ and $\mathbf{0} \in \mathcal{M}_n$.

By the second statement of Lemma 3, we can choose $t^* = 2 \sup_{\mathbf{H} \in \mathcal{M}_n} |\langle \mathbf{H}, \mathbf{Z}' \rangle| + s$ for any constant $s > 0$. Then, it can be easily checked that, for any $t \geq t^*$, we have

$$\sup_{\mathbf{Q} \in \mathcal{P}(\mathcal{T}'_n, \mathcal{S}'_n) \cap \mathbb{B}(\Pi^* \Theta^* (\Pi^*)^\top, t)} \langle \mathbf{Q} - \Pi^* \Theta^* (\Pi^*)^\top, \mathbf{Z} \rangle - \frac{t^2}{2} \leq t \sup_{\mathbf{H} \in \mathcal{M}_n} \langle \mathbf{H}, \mathbf{Z}' \rangle - \frac{t^2}{2} < 0. \quad (\text{IX.5})$$

Thus, we have

$$\mathbb{E} \|\widehat{\Pi} \widehat{\Theta} \widehat{\Pi}^\top - \Pi^* \Theta^* (\Pi^*)^\top\|_F \leq 2 \mathbb{E} \sup_{\mathbf{H} \in \mathcal{M}_n} |\langle \mathbf{H}, \mathbf{Z}' \rangle| + s. \quad (\text{IX.6})$$

To control the right-hand side of the above inequality, we use the following version of Dudley's integral inequality, whose proof is given in Section X below.

Lemma 4 (Dudley's integral inequality): Let $\{X_t\}_{t \in T}$ be a mean zero random process on a metric space (T, d) with sub-Gaussian increments, in the sense that there exists some constant $K > 0$ such that $\inf\{t > 0 : \mathbb{E} \exp((X_t - X_s)^2/t^2) \leq 2\} \leq Kd(t, s)$ for all $t, s \in T$. Then, we have

$$\mathbb{E} \sup_{t \in T} |X_t| \leq CK \int_0^{\text{diam}(T)} \sqrt{\log \mathcal{N}(T, d, \epsilon)} d\epsilon + \inf_{t \in T} \mathbb{E} |X_t|, \quad (\text{IX.7})$$

where $\mathcal{N}(T, d, \epsilon)$ is the ϵ -covering number of T , that is, the smallest number of closed balls with centers in T and radii ϵ whose union covers T .

Since $\{\langle \mathbf{H}, \mathbf{Z}' \rangle\}_{\mathbf{H} \in \mathcal{M}_n}$ is a mean-zero random process with sub-Gaussian increment, and one can check that for

any $\mathbf{H}_1, \mathbf{H}_2 \in \mathcal{M}_n$, there exists some constant $C' > 0$ such that

$$\inf\{t > 0 : \mathbb{E} \exp(\langle \mathbf{H}_1 - \mathbf{H}_2, \mathbf{Z}' \rangle^2 / t^2) \leq 2\} \leq \sigma C' \|\mathbf{H}_1 - \mathbf{H}_2\|_F.$$

From Lemma 4, it follows that

$$\mathbb{E} \sup_{\mathbf{H} \in \mathcal{M}_n} |\langle \mathbf{H}, \mathbf{Z}' \rangle| \leq C\sigma \int_0^{\text{diam}(\mathcal{M}_n)} \sqrt{\log \mathcal{N}(\mathcal{M}_n, d_2, \epsilon)} d\epsilon + \mathbb{E} |\langle \mathbf{H}_0, \mathbf{Z}' \rangle|, \quad (\text{IX.8})$$

for any $\mathbf{H}_0 \in \mathcal{M}_n$, where the metric d_2 is defined as $d_2(\mathbf{G}_1, \mathbf{G}_2) = \|\mathbf{G}_1 - \mathbf{G}_2\|_F$. In particular, we can take $\mathbf{H}_0 = \mathbf{0}$ to get

$$\mathbb{E} |\langle \mathbf{H}_0, \mathbf{Z}' \rangle| = 0. \quad (\text{IX.9})$$

In this way, we further reduced the calculation of the denoising risk $\mathbb{E} \|\widehat{\Pi} \widehat{\Theta} \widehat{\Pi}^\top - \Pi \Theta \Pi^\top\|_F$ to that of the metric entropy of the set \mathcal{M}_n under the Frobenius norm. The following lemma, proved in Section X below, provides an estimate of such a entropy measure.

Lemma 5: Under the conditions of Theorem 1, using the above notations, we have

$$\int_0^{\text{diam}(\mathcal{M}_n)} \sqrt{\log \mathcal{N}(\mathcal{M}_n, d_2, \epsilon)} d\epsilon \leq C\sqrt{n \log n}. \quad (\text{IX.10})$$

The above upper bound estimate of the entropy integral along with inequalities (IX.6) (IX.8) and (IX.9) leads to the upper bound in Theorem 5, as long as we choose s sufficiently small, for example, $s = \sigma$. This completes the proof of the theorem.

C. Proof of Theorem 2

From Lemma 1, we can further simplify the lower bound for $\beta_n(d)$ using the inequality

$$\left(\frac{n}{e}\right)^n e \leq n! \leq \left(\frac{n+1}{e}\right)^{n+1} e,$$

which implies

$$\frac{n!}{(n-d)!} \geq \frac{(n/e)^n}{[(n-d+1)/e]^{n-d+1}} \geq \left(\frac{n}{n-d+1}\right)^{n-d+1} (n/e)^{d-1} \geq (n/e)^{d-1}.$$

Consequently, for $2 \leq d \leq n-2$, we have

$$\log \beta_n(d) \geq (d-1) \log(n/e) - \log(n-d). \quad (\text{IX.11})$$

Set $d = (1 - e^{-1})n$. We have

$$\log \beta_n(d) \geq (d-2) \log(n/e).$$

Thus, for $n \geq 16$, we have

$$\log \beta_n(d) \geq \frac{n}{2} \log(n/e). \quad (\text{IX.12})$$

This is the key packing number inequality that we will use for the proof of the fundamental limit.

The proof of the fundamental limit starts with a careful construction of a set of least favourable scenarios, over which exact matrix reordering is most difficult to achieve. To this end, we consider the following tridiagonal signal matrix, which is also the adjacency matrix of a Hamiltonian path,

$$\Theta_0 = \begin{bmatrix} 0 & \delta & 0 & \dots & 0 \\ \delta & 0 & \delta & \dots & 0 \\ 0 & \delta & 0 & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}. \quad (\text{IX.13})$$

We also consider the subset of \mathcal{S}_n that only permutes the columns and rows in Θ_0 satisfying $i = 3k - 1$ for some $k \in \{1, \dots, \lfloor n/3 \rfloor\}$. The reason we consider such a class of permutations is that, by treating Θ_0 as a concatenation of small blocks of size 3×3 , it will be seen that, for any two permutations π_1, π_2 in the above subset, the distance $\|\Pi_1 \Theta_0 \Pi_1^\top - \Pi_2 \Theta_0 \Pi_2^\top\|_F^2$ is completely determined by the Hamming distance $d_H(\pi_1, \pi_2)$ between the two permutations (equation (IX.14) below).

Let $n' = \lfloor n/3 \rfloor$. Suppose $n \geq 48$, or $n' \geq 16$, and set $d = (1 - e^{-1})n'$. By inequality (IX.12), there exists a subset $\mathcal{S}_{n'}^*$ of $\mathcal{S}_{n'}$ that attains the maximal d -packing number $\beta_{n'}(d)$, where $\log \beta_{n'}(d) \geq \frac{n'}{2} \log(n'/e) \geq \frac{5n}{32} \log(5n/16e)$, as $n' \geq 5n/16$ for $n \geq 48$.

Now we identify the permutations in $\mathcal{S}_{n'}^*$ with the permutations in \mathcal{S}_n that only involve the i -th element for $i = 3k - 1, k \in \{1, \dots, \lfloor n/3 \rfloor\}$. Thus, we have constructed a subset \mathcal{S}_n^* of \mathcal{S}_n where $\log |\mathcal{S}_n^*| \geq \frac{5n}{32} \log(5n/16e)$ elements with mutual distance at least $d \geq (1 - e^{-1})5n/16$.

Now note that for any two permutations $\Pi_1, \Pi_2 \in \mathcal{S}_n^*$ such that $d_H(\pi_1, \pi_2) = d$, we have

$$2d\delta^2 \leq \|\Pi_1 \Theta_0 \Pi_1^\top - \Pi_2 \Theta_0 \Pi_2^\top\|_F^2 \leq 4n\delta^2 \quad (\text{IX.14})$$

To see this, we notice that

$$\|\Pi_1 \Theta_0 \Pi_1^\top - \Pi_2 \Theta_0 \Pi_2^\top\|_F^2 = \|\Pi \Theta_0 \Pi^\top - \Theta_0\|_F^2,$$

where $\Pi = \Pi_2^\top \Pi_1 \in \mathcal{S}_n^*$. We denote $\Theta_0 = (\theta_{ij})_{1 \leq i, j \leq n}$ and calculate that, for any i such that $\pi(i) > i$, we have

$$\sum_{j=1}^n (\theta_{\pi(i), \pi(j)} - \theta_{ij})^2 \geq (\theta_{\pi(i), \pi(i)+1} - \theta_{i, \pi(i)+1})^2 + (\theta_{\pi(i), i-1} - \theta_{i, i-1})^2 \geq 2\delta^2,$$

since the $(\pi(i) + 1)$ -th and the $(i - 1)$ -th components are fixed in the permutation Π . A similar result can be obtained for i such that $\pi(i) < i$. Therefore, we obtain the lower bound

$$\|\Pi_1 \Theta_0 \Pi_1^\top - \Pi_2 \Theta_0 \Pi_2^\top\|_F^2 \geq \sum_{i:i \neq \pi(i)} \sum_{j=1}^n (\theta_{\pi(i)\pi(j)} - \theta_{ij})^2 \geq 2d\delta^2. \quad (\text{IX.15})$$

The upper bound follows from the simple inequality

$$\|\Pi_1 \Theta_0 \Pi_1^\top - \Pi_2 \Theta_0 \Pi_2^\top\|_F^2 \leq 2\|\Theta_0\|_F^2 \leq 4(n-1)\delta^2 \quad (\text{IX.16})$$

In the following, we consider the fundamental limit for $\rho^*(\cdot, \cdot)$ over the subspace $\{\Theta_0\} \times \mathcal{S}_n^*$. The proof relies on the following lemma from [69].

Lemma 6: Assume that for some integer $M \geq 2$ there exist distinct parameters $\theta_0, \dots, \theta_M$ from the parameter space Θ and mutually absolutely continuous probability measures P_0, \dots, P_M with $P_j = P_{\theta_j}$ for $j = 0, 1, \dots, M$, defined on a common probability space (Ω, \mathcal{F}) such that the averaged KL divergence $\frac{1}{M} \sum_{j=1}^M D(P_j, P_0) \leq \frac{1}{8} \log M$. Then, for every measurable mapping $\hat{\theta} : \Omega \rightarrow \Theta$,

$$\max_{j=0, \dots, M} P_j(\hat{\theta} \neq \theta_j) \geq \frac{\sqrt{M}}{\sqrt{M} + 1} \left(\frac{3}{4} - \frac{1}{2\sqrt{\log M}} \right).$$

Applying the above lemma to the parameter subspace $\{\Theta_0\} \times \mathcal{S}_n^*$, we could check that for any $\Pi_1, \Pi_2 \in \mathcal{S}_n^*$, the KL divergence between the probability measures of $\mathbf{Y}_1 = \Pi_1 \Theta_0 \Pi_1^\top + \mathbf{Z}$ and that of $\mathbf{Y}_2 = \Pi_2 \Theta_0 \Pi_2^\top + \mathbf{Z}$ can be bounded by

$$D(P_{\mathbf{Y}_1}, P_{\mathbf{Y}_2}) = \frac{\|\Pi_1 \Theta_0 \Pi_1^\top - \Pi_2 \Theta_0 \Pi_2^\top\|_F^2}{2\sigma^2} \leq \frac{2n\delta^2}{\sigma^2},$$

where the last inequality follows from (IX.14). Set $\delta = 0.06\sigma\sqrt{\log(5n/16e)}$. By the lower bound on $|\mathcal{S}_n^*|$, we have

$$D(P_{\mathbf{Y}_1}, P_{\mathbf{Y}_2}) \leq 0.019n \log(5n/16e) \leq \frac{1}{8} \log |\mathcal{S}_n^*|.$$

Thus, by Lemma 6, it follows that, for any $n \geq 48$,

$$\inf_{\hat{\Pi}} \max_{(\Theta, \Pi) \in \{\Theta_0\} \times \mathcal{S}_n^*} P(\hat{\Pi} \Theta \hat{\Pi}^\top \neq \Pi \Theta \Pi^\top) = \inf_{\hat{\pi}} \max_{(\Theta, \pi) \in \{\Theta_0\} \times \mathcal{S}_n^*} P(\hat{\pi} \neq \pi) \geq 0.6. \quad (\text{IX.17})$$

Lastly, by the lower bound (IX.14) on the mutual distance between the elements in the set \mathcal{S}_n^* , we have

$$\begin{aligned} \rho^*(\{\Theta_0\}, \mathcal{S}_n^*) &= \min_{\substack{\Pi_1, \Pi_2 \in \mathcal{S}_n^* \\ \Pi_1 \neq \Pi_2}} \|\Pi_1 \Theta_0 \Pi_1 - \Pi_2 \Theta_0 \Pi_2\|_F \\ &\geq \sqrt{2d\delta^2} > 0.037\sigma\sqrt{n \log(5n/16e)} > 0.02\sigma\sqrt{n \log n}, \end{aligned}$$

where the last inequality follows from $\log(5n/16e) \geq 0.44 \log n$.

In other words, we have found a subset $\{\Theta_0\} \times \mathcal{S}_n^*$ of $\mathcal{T}_n \times \mathcal{S}_n$ with $\rho^*(\{\Theta_0\}, \mathcal{S}_n^*) \geq 0.02\sigma\sqrt{n\log n}$ such that uniform exact matrix ordering over $\{\Theta_0\} \times \mathcal{S}_n^*$ with high probability is not possible. In particular, from the above argument, we can see that for any $\rho \leq 0.02\sigma\sqrt{n\log n}$, we can always construct similar subsets by choosing δ smaller such that $\rho^*(\{\Theta_0\}, \mathcal{S}_n^*) = \rho$, and show that uniform exact matrix ordering over $\{\Theta_0\} \times \mathcal{S}_n^*$ with high probability is not possible. This completes the proof of the theorem.

D. Proof of Theorem 3

We first define the degree and the Laplacian operators as follows.

Definition 3 (Degree & Laplacian Operators): For a symmetric matrix $\mathbf{A} = (a_{ij})_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}$, define the degree operator $\mathbf{D} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ by $\mathbf{D}(\mathbf{A}) = \text{diag}(\sum_{i=1}^n a_{i1}, \dots, \sum_{i=1}^n a_{in})$, and the Laplacian operator $\mathbf{L} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ by $\mathbf{L}(\mathbf{A}) = \mathbf{D}(\mathbf{A}) - \mathbf{A}$.

Without loss of generality, we set $\Pi = \text{Id}_n$. We define $\mathcal{T}'_n = \{\Theta_0\}$ where Θ_0 is the tridiagonal matrix defined in (IX.13), with $\delta = C_0\sigma n^2\sqrt{n}$, and let \mathcal{S}'_n in the theorem be the permutation set \mathcal{S}_n^* constructed in the proof of Theorem 2. By inequalities (IX.15) and (IX.16), we have $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) = C\sigma n^3$ for some absolute constant $C > 0$. Note that $\mathbf{L} \equiv \mathbf{L}(\mathbf{Y}) = \mathbf{L}(\Theta) + \mathbf{L}(\mathbf{Z})$. We denote $\mathbf{B} = \mathbf{L}(\Theta)$ and $\mathbf{E} = \mathbf{L}(\mathbf{Z})$. Let $\mathbf{B} = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^\top$ be the eigendecomposition of \mathbf{B} , with $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. In particular, by Lemma 7 below, \mathbf{B} has simple eigenvalues, which implies $\mathbf{u}_2 = \mathbf{v}$, the Fiedler vector, up to a change of sign. The following lemma is well-known, and can be found, for example, on page 3234 of [70].

Lemma 7: The Laplacian matrix of any tridiagonal matrix in the form (IX.13) has eigenvalues $\lambda_k = 4\delta \sin^2\left(\frac{\pi(k-1)}{2n}\right)$, for $k = 1, \dots, n$, and eigenvectors $\mathbf{u}_k = (u_{k1}, u_{k2}, \dots, u_{kn})^\top$, $k = 1, \dots, n$ where $u_{kj} = \frac{1}{\sqrt{n/2}} \cos\left(\frac{\pi(k-1)(j-1/2)}{n}\right)$, for $j = 1, \dots, n$. In particular, the Fiedler vector \mathbf{u}_2 is $\mathbf{u}_2 = \frac{1}{\sqrt{n/2}} \left(\cos\left(\frac{\pi}{2n}\right), \cos\left(\frac{3\pi}{2n}\right), \dots, \cos\left(\frac{(2n-1)\pi}{2n}\right) \right)^\top$.

In order to show $\Pi\Theta_0\Pi^\top \neq \check{\Pi}\Theta_0\check{\Pi}$, by the definition of Θ_0 , it is equivalent to showing $\Pi \neq \check{\Pi}$, or $\tau(\hat{\mathbf{v}}) \neq \tau(\mathbf{v})$. The rest of the proof is devoted to

$$\liminf_{n \rightarrow \infty} \inf_{\Pi \in \mathcal{S}'_n} P_{\Theta_0, \Pi}(\tau(\hat{\mathbf{v}}) \neq \tau(\mathbf{v})) \geq 1/2. \quad (\text{IX.18})$$

Let $\mathbf{B} + \mathbf{E} = \sum_{i=1}^n \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^\top$ be the eigendecomposition of $\mathbf{B} + \mathbf{E}$ with $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \dots \leq \hat{\lambda}_n$. In other words, we have $\hat{\mathbf{v}} = \hat{\mathbf{u}}_2$. We also define $\hat{\mathbf{v}}_-$ as the eigenvector associated to the second smallest eigenvalue of $\mathbf{B} - \mathbf{E}$.

In order to show $\tau(\hat{\mathbf{v}}) \neq \tau(\mathbf{v})$, we note that by Lemma 7, the minimal distance between any two consecutive components in \mathbf{v} is bounded by

$$\min_{1 \leq i \neq j \leq n} |u_{2i} - u_{2j}| \leq \sqrt{\frac{2}{n}} \cdot \frac{\pi}{n} \cdot \sin\left(\frac{1}{2n}\right) \leq \frac{\pi}{\sqrt{2}n^{5/2}}.$$

If we are able to show that $\|\hat{\mathbf{v}} - \mathbf{v}\|_\infty > \frac{\pi}{\sqrt{2}n^{5/2}}$, then it follows that $\tau(\hat{\mathbf{v}}) \neq \tau(\mathbf{v})$. Since $\|\hat{\mathbf{v}} - \mathbf{v}\|_\infty \geq \frac{1}{\sqrt{n}} \|\hat{\mathbf{v}} - \mathbf{v}\|_2$, it suffices to show that $\|\hat{\mathbf{v}} - \mathbf{v}\|_2 > \frac{\pi}{\sqrt{2}n^2}$. To this end, we obtain the following proposition, proved in Section X

below.

Proposition 3: Define the probability event $\mathcal{A}_n = \{\|\mathbf{E}\mathbf{v}\|_2 \geq C_1\sigma\sqrt{n}, \|\mathbf{E}\| \leq C_2\sigma\sqrt{n\log n}, |\mathbf{v}^\top \mathbf{E}\mathbf{v}| \leq C_3\sigma\sqrt{\log n}\}$. Then we have $\lim_{n \rightarrow \infty} P(\mathcal{A}_n) = 1$, where $C_1, C_2, C_3 > 0$ are some universal constants. In addition, under the event \mathcal{A}_n , for any constant $c > 0$ we have $\|\hat{\mathbf{v}} - \mathbf{v}\|_2 + \|\hat{\mathbf{v}}_- - \mathbf{v}\|_2 \geq c/n^2$ for sufficiently large n .

Since the distribution of \mathbf{E} is the same as that of $-\mathbf{E}$, the distributions of the corresponding eigenvectors $\hat{\mathbf{v}}$ and $\hat{\mathbf{v}}_-$ should also be identical. In other words, we have $P(\|\hat{\mathbf{v}} - \mathbf{v}\|_2 \geq z) = P(\|\hat{\mathbf{v}}_- - \mathbf{v}\|_2 \geq z)$ for all $z \in \mathbb{R}$. It then follows that, for any constant $c > 0$,

$$\begin{aligned} P(\mathcal{A}_n) &= P(\mathcal{A}_n, \|\hat{\mathbf{v}} - \mathbf{v}\|_2 + \|\hat{\mathbf{v}}_- - \mathbf{v}\|_2 \geq c/n^2) \\ &\leq P((\mathcal{A}_n \cap \{\|\hat{\mathbf{v}} - \mathbf{v}\|_2 \geq \frac{c}{2n^2}\}) \cup (\mathcal{A}_n \cap \{\|\hat{\mathbf{v}}_- - \mathbf{v}\|_2 \geq \frac{c}{2n^2}\})) \\ &\leq P(\|\hat{\mathbf{v}} - \mathbf{v}\|_2 \geq cn^{-2}/2) + P(\|\hat{\mathbf{v}}_- - \mathbf{v}\|_2 \geq cn^{-2}/2) \\ &= 2P(\|\hat{\mathbf{v}} - \mathbf{v}\|_2 \geq cn^{-2}/2), \end{aligned}$$

which implies $\lim_{n \rightarrow \infty} P(\|\hat{\mathbf{v}} - \mathbf{v}\|_2 \geq c'n^{-2}) \geq 1/2$. This completes the proof.

E. Proof of Theorem 4

We define the class of λ -ridged Topelitz matrices as

$$\mathcal{T}_n^R(\lambda) = \left\{ \Theta \in \mathcal{T}_n : \theta_1 - \theta_{\lceil n/2 \rceil} \geq \lambda \right\}. \quad (\text{IX.19})$$

Then we have the following propositions hold.

Proposition 4 (Theoretical guarantee over λ -ridged Topelitz matrices): There exists some absolute constants $C, c > 0$ such that, whenever $\lambda \geq C\sigma n$, we have $\sup_{(\Theta, \Pi) \in \mathcal{T}_n^R(\lambda) \times \mathcal{S}_n} P_{\Theta, \Pi}(\tilde{\Pi}\Theta\tilde{\Pi}^\top \neq \Pi\Theta\Pi^\top) \leq n^{-c}$.

Proposition 5 (Sharp λ - ρ^ correspondence):* For any $\mathcal{T}'_n \subseteq \mathcal{T}_n$ and any $\mathcal{S}'_n \subseteq \mathcal{S}_n$ such that $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) \geq C_1 n \lambda^*$ for some absolute constant $C_1 > 0$, there exists some absolute constant $C_2 > 0$ such that $\mathcal{T}'_n \subseteq \mathcal{T}_n^R(\lambda)$ for $\lambda = C_2 \lambda^*$. On the other hand, the above characterization is asymptotically sharp in the sense that, for any λ^* , there exists some $(\mathcal{T}'_n, \mathcal{S}'_n) \subseteq \mathcal{T}_n \times \mathcal{S}_n$ such that $\mathcal{T}'_n \subseteq \mathcal{T}_n^R(\lambda^*)$ and $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) = C n \lambda^*$.

The proofs of these propositions can be found in Section X below. With the above results, Theorem 4 is then proved by combining Propositions 4 and 5 with $\lambda^* = \sigma n$. In particular, the sharp λ - ρ^* correspondence in Proposition 5 suggests that the argument for proving Theorem 4, and therefore the obtained minimal signal strength condition (IV.1) for the adaptive sorting algorithm, are asymptotically tight.

X. PROOF OF TECHNICAL RESULTS

A. Dudley's Integral Inequality: Proof of Lemma 4

The integral inequality stated in the lemma generalizes the results in [71]. Specifically, by Remark 8.1.5 (supremum of increments) of [71], we have

$$\mathbb{E} \sup_{t \in T} |X_t - X_{t_0}| \leq CK \int_0^{\text{diam}(T)} \sqrt{\log \mathcal{N}(T, d, \epsilon)} d\epsilon,$$

holds for any fixed $t_0 \in T$. Then our lemma follows immediately from the simple inequality $\mathbb{E} \sup_{t \in T} |X_t| \leq \mathbb{E} \sup_{t \in T} |X_t - X_{t_0}| + \mathbb{E} |X_{t_0}|$.

B. Metric Entropy Calculation: Proof of Lemma 5

In order to study the metric entropy of the set \mathcal{M}_n , we consider the permutation set \mathcal{S}'_n and denote $\mathcal{S}'_n = \{\Pi_1, \dots, \Pi_{|\mathcal{S}'_n|}\}$. For any given element $\mathbf{H} \in \mathcal{M}_n$, there exists an element $\Pi_i \in \mathcal{S}'_n$ and a Toeplitz matrix $\mathbf{Q}^* \in \mathcal{T}'_n$ such that

$$\mathbf{H} = (\Pi^*)^\top \Pi_i \mathbf{Q}^* \Pi_i^\top \Pi^* - t^{-1} \Theta^*.$$

In other words, if we define $\bar{\mathcal{F}}_i \equiv \{(\Pi^*)^\top \Pi_i \mathbf{Q} \Pi_i^\top \Pi^* - t^{-1} \Theta^* : \mathbf{Q} \in \mathcal{T}'_n\}$ where \mathcal{T}'_n is the set of all the $n \times n$ symmetric Toeplitz matrices, then there exists an $i \in \{1, \dots, |\mathcal{S}'_n|\}$ such that $\mathbf{H} \in \bar{\mathcal{F}}_i$. Hence, we have

$$\mathcal{M}_n \subseteq \bigcup_{1 \leq i \leq |\mathcal{S}'_n|} \bar{\mathcal{F}}_i. \quad (\text{X.1})$$

Moreover, by definition we also have $\mathcal{M}_n \subseteq \mathbb{B}(\mathbf{0}, 1)$, so that

$$\mathcal{M}_n \subseteq \bigcup_{1 \leq i \leq |\mathcal{S}'_n|} \bar{\mathcal{F}}_i \cap \mathbb{B}(\mathbf{0}, 1). \quad (\text{X.2})$$

By the union bound, we have

$$\begin{aligned} \log \mathcal{N}(\mathcal{M}_n, d_2, \epsilon) &\leq \log \left[\sum_{1 \leq i \leq |\mathcal{S}'_n|} \mathcal{N}(\bar{\mathcal{F}}_i \cap \mathbb{B}(\mathbf{0}, 1), d_2, \epsilon) \right] \\ &\leq \log |\mathcal{S}'_n| + \log \left[\max_{1 \leq i \leq |\mathcal{S}'_n|} \mathcal{N}(\bar{\mathcal{F}}_i \cap \mathbb{B}(\mathbf{0}, 1), d_2, \epsilon) \right] \\ &\leq \log(n!) + \log \left[\max_{1 \leq i \leq m} \mathcal{N}(\bar{\mathcal{F}}_i \cap \mathbb{B}(\mathbf{0}, 1), d_2, \epsilon) \right]. \end{aligned}$$

For the first term in the last inequality, by Stirling's formula, we have

$$\log(n!) \leq (n+1) \log \frac{n+1}{e} + 1 \leq 3n \log n, \quad (\text{X.3})$$

where the last inequality holds for all $n \geq 2$. In the following, we control the second term in the last inequality, to obtain an upper bound for $\log \mathcal{N}(\mathcal{M}_n, d_2, \epsilon)$.

We control the metric entropy $\log \mathcal{N}(\bar{\mathcal{F}}_i \cap \mathbb{B}(\mathbf{0}, 1), d_2, \epsilon)$ for any $i \in \{1, 2, \dots, |\mathcal{S}'_n|\}$. Suppose without loss of generality we consider the set $\bar{\mathcal{F}}_{i_0}$, associated to the permutation $\Pi_{i_0} \in \mathcal{S}'_n$. Recall that we denote $\pi_{i_0} : [n] \rightarrow [n]$ as the permutation map associated to the permutation matrix Π_{i_0} . Now we consider the map $\Phi_{i_0} : (\mathbb{R}^n, d_2) \rightarrow (\bar{\mathcal{F}}_{i_0}, d_2)$, where for each $\gamma = (\gamma_1, \dots, \gamma_n) \in \mathbb{R}^n$, we define a matrix $\mathbf{H} \in \bar{\mathcal{F}}_{i_0}$ as follows:

- 1) Define a symmetric Toeplitz matrix $\mathbf{Q}^*(\gamma)$ such that its first row is γ .
- 2) Set $\Phi_{i_0}(\gamma) = (\Pi^*)^\top \Pi_{i_0} \mathbf{Q}^*(\gamma) \Pi_{i_0}^\top \Pi^* - t^{-1} \Theta^*$.

The map Φ_{i_0} plays a key role in translating the metric entropy of $\bar{\mathcal{F}}_{i_0} \cap \mathbb{B}(\mathbf{0}, 1)$ in $(\mathbb{R}^{n \times n}, d_2)$ to that of the Euclidean ball $\mathbb{B}_n(\mathbf{0}, 1)$ in (\mathbb{R}^n, d_2) . Specifically, we will need the following lemma concerning the property of covering numbers with respect to Lipschitz maps.

Lemma 8 ([72]): Let (M, d) and (M_1, d_1) be metric spaces, $K \subset M$, $\Phi : M \rightarrow M_1$, and let $L > 0$. If Φ satisfies $d_1(\Phi(x), \Phi(y)) \leq Ld(x, y)$ for $x, y \in M$, then, for every $\epsilon > 0$, we have $\mathcal{N}(\Phi(K), d_1, L\epsilon) \leq \mathcal{N}(K, d, \epsilon)$.

Now to use Lemma 8, for any $x, y \in \mathbb{R}^n$, we have, on the one hand,

$$\begin{aligned} \|\Phi_{i_0}(x) - \Phi_{i_0}(y)\|_F^2 &= \|(\Pi^*)^\top \Pi_{i_0} \mathbf{Q}^*(x) \Pi_{i_0}^\top \Pi^* - (\Pi^*)^\top \Pi_{i_0} \mathbf{Q}^*(y) \Pi_{i_0}^\top \Pi^*\|_F^2 \\ &= \|\mathbf{Q}^*(x) - \mathbf{Q}^*(y)\|_F^2 \\ &\leq 2n\|x - y\|_2^2, \end{aligned} \tag{X.4}$$

where the last inequality follows from the property of symmetric Toeplitz matrices, and on the other hand

$$\begin{aligned} \|\Phi_{i_0}(x) - \Phi_{i_0}(y)\|_F^2 &= \|(\Pi^*)^\top \Pi_{i_0} \mathbf{Q}^*(x) \Pi_{i_0}^\top \Pi^* - (\Pi^*)^\top \Pi_{i_0} \mathbf{Q}^*(y) \Pi_{i_0}^\top \Pi^*\|_F^2 \\ &= \|\mathbf{Q}^*(x) - \mathbf{Q}^*(y)\|_F^2 \\ &\geq \|x - y\|_2^2, \end{aligned} \tag{X.5}$$

where the last inequality follows by only considering the first rows of $\mathbf{Q}^*(x) - \mathbf{Q}^*(y)$. By the above lower bound (X.5), it follows that

$$\{\gamma \in \mathbb{R}^n : \|\Phi_{i_0}(\gamma)\|_F \leq 1\} \subseteq \{\gamma \in \mathbb{R}^n : \|\gamma\|_2 \leq 1\}. \tag{X.6}$$

In addition, for any $\mathbf{H} \in \bar{\mathcal{F}}_{i_0} \cap \mathbb{B}(\mathbf{0}, 1)$, there exists a vector $\gamma \in \{\gamma \in \mathbb{R}^n : \|\Phi_{i_0}(\gamma)\|_F \leq 1\}$ such that $\Phi_{i_0}(\gamma) = \mathbf{H}$.

Then we also have

$$\bar{\mathcal{F}}_{i_0} \cap \mathbb{B}(\mathbf{0}, 1) \subseteq \Phi_{i_0}(\{\gamma \in \mathbb{R}^n : \|\Phi_{i_0}(\gamma)\|_F \leq 1\}). \tag{X.7}$$

By taking the images of both sides of (X.6) under the map Φ_{i_0} and applying (X.7), we have

$$\bar{\mathcal{F}}_{i_0} \cap \mathbb{B}(\mathbf{0}, 1) \subseteq \Phi_{i_0}(\mathbb{B}_n(\mathbf{0}, 1)),$$

which implies

$$\log \mathcal{N}(\bar{\mathcal{F}}_i \cap \mathbb{B}(\mathbf{0}, 1), d_2, \epsilon) \leq \log \mathcal{N}(\Phi_{i_0}(\mathbb{B}_n(\mathbf{0}, 1)), d_2, \epsilon).$$

The right-hand side of the inequality can be further bounded by Lemma 8 and the upper bound (X.4) as

$$\log \mathcal{N}(\Phi_{i_0}(\mathbb{B}_n(\mathbf{0}, 1)), d_2, \epsilon) \leq \log \mathcal{N}(\mathbb{B}_n(\mathbf{0}, 1), d_2, \epsilon/\sqrt{2n}) \leq n \log \left(\frac{2\sqrt{2n}}{\epsilon} + 1 \right), \quad (\text{X.8})$$

where the last inequality is a direct consequence of the following entropy bound for the Euclidean unit ball

$$\mathcal{N}(\mathbb{B}_n(\mathbf{0}, 1), d_2, \epsilon) \leq (2/\epsilon + 1)^n,$$

which can be found, for example, in Corollary 4.2.13 (covering numbers of the Euclidean ball) of [71]. Hence, combining inequalities (X.3) and (X.8), we have

$$\log \mathcal{N}(\mathcal{M}_n, d_2, \epsilon) \lesssim n \log n + n \log \left(\frac{2\sqrt{2n}}{\epsilon} + 1 \right). \quad (\text{X.9})$$

Finally, since $\mathcal{M}_n \subseteq \mathbb{B}(0, 1)$, we have for $n \geq 2$,

$$\begin{aligned} \int_0^{\text{diam}(\mathcal{M}_n)} \sqrt{\log \mathcal{N}(\mathcal{M}_n, d_2, \epsilon)} d\epsilon &\leq C \int_0^2 \sqrt{n \log n + n \log \left(\frac{2\sqrt{2n}}{\epsilon} + 1 \right)} d\epsilon \\ &\leq C \sqrt{n \log n} + C \sqrt{n} \int_0^2 \sqrt{\log \left(\frac{2\sqrt{2n}}{\epsilon} + 1 \right)} d\epsilon \leq C \sqrt{n \log n}, \end{aligned}$$

where the last inequality follows from

$$\begin{aligned} \int_0^2 \sqrt{\log \left(\frac{2\sqrt{2n}}{\epsilon} + 1 \right)} d\epsilon &\leq \int_0^2 \sqrt{\log \left(\frac{4\sqrt{n}}{\epsilon} \right)} d\epsilon \leq C \sqrt{\log n} + C \int_0^2 \sqrt{\log \left(\frac{1}{\epsilon} \right)} d\epsilon \\ &\leq C \sqrt{\log n} + C \int_0^\infty t^2 e^{-t^2} dt \leq C \sqrt{\log n}. \end{aligned}$$

This completes the proof of the lemma.

C. Inconsistency of Sample Eigenvectors: Proof of Proposition 3

We define $\mathbf{B} - \mathbf{E} = \sum_{i=1}^n \tilde{\lambda}_i \tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_i^\top$ where $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \dots \leq \tilde{\lambda}_n$ are the eigenvalues, and $\{\tilde{\mathbf{u}}_i\}$ are the eigenvectors. We denote $\angle(\mathbf{g}, \mathbf{h})$ as the angle between two vectors \mathbf{h} and \mathbf{g} . Because the eigenvectors are identifiable up to a change of sign, we assume without loss of generality that all the angles are between 0 and $\pi/2$. The proof of this proposition is separated into the following steps:

- 1) We show that under event \mathcal{A}_n , if instead $\|\hat{\mathbf{v}} - \mathbf{v}\|_2 \leq c/n^2$ and $\|\hat{\mathbf{v}}_- - \mathbf{v}\|_2 \leq c/n^2$ for some constant $c > 0$, then the angle $\angle(\hat{\mathbf{v}}, (\mathbf{B} - \mathbf{E})\hat{\mathbf{v}}) \geq \frac{\pi}{2} - \epsilon_1$ for some sufficiently small constant $\epsilon_1 \in (0, \pi/2)$.
- 2) We show that the above statement implies that $\|\hat{\mathbf{v}} - \hat{\mathbf{v}}_-\|_2 \geq \epsilon_2$ for some constant $\epsilon_2 > 0$. A contradiction to the first statement by triangle inequality. Thus, we conclude that under event \mathcal{A}_n , we must have $\|\hat{\mathbf{v}} - \mathbf{v}\|_2 + \|\hat{\mathbf{v}}_- - \mathbf{v}\|_2 \geq c/n^2$.
- 3) We finish the proof by showing that event \mathcal{A}_n holds in probability.

The detailed proofs are presented in order.

a) *Step I:* Note that $\hat{\lambda}_2 \hat{\mathbf{v}} = (\mathbf{B} + \mathbf{E})\hat{\mathbf{v}}$. Then

$$\angle(\hat{\mathbf{v}}, (\mathbf{B} - \mathbf{E})\hat{\mathbf{v}}) = \angle((\mathbf{B} + \mathbf{E})\hat{\mathbf{v}}, (\mathbf{B} - \mathbf{E})\hat{\mathbf{v}}) \geq \angle((\mathbf{B} + \mathbf{E})\hat{\mathbf{v}}, \mathbf{B}\hat{\mathbf{v}}). \quad (\text{X.10})$$

Now consider the triangle with sides $\mathbf{B}\hat{\mathbf{v}}$, $\mathbf{E}\hat{\mathbf{v}}$ and $(\mathbf{B} + \mathbf{E})\hat{\mathbf{v}}$. By the sine rule, we have

$$\sin \angle(\mathbf{B}\hat{\mathbf{v}}, (\mathbf{B} + \mathbf{E})\hat{\mathbf{v}}) = \frac{\|\mathbf{E}\hat{\mathbf{v}}\|_2}{\|\mathbf{B}\hat{\mathbf{v}}\|_2} \cdot \sin \angle(\mathbf{E}\hat{\mathbf{v}}, \hat{\mathbf{v}}) \quad (\text{X.11})$$

Note that

$$\|\mathbf{B}\hat{\mathbf{v}}\|_2 \leq \|\mathbf{B}\mathbf{v}\|_2 + \|\mathbf{B}(\mathbf{v} - \hat{\mathbf{v}})\|_2 \leq \lambda_2 + \lambda_n \cdot \|\mathbf{v} - \hat{\mathbf{v}}\|_2 \lesssim \frac{\delta}{n^2} + \frac{\delta}{n^2} \lesssim \sigma\sqrt{n}, \quad (\text{X.12})$$

where the second last inequality follows from the assumption that $\|\mathbf{v} - \hat{\mathbf{v}}\|_2 \lesssim n^{-2}$, and Lemma 7. On the other hand, we have

$$\|\mathbf{E}\hat{\mathbf{v}}\|_2 \geq \|\mathbf{E}\mathbf{v}\|_2 - \|\mathbf{E}(\mathbf{v} - \hat{\mathbf{v}})\|_2 \geq \|\mathbf{E}\mathbf{v}\|_2 - \|\mathbf{E}\| \cdot \|\mathbf{v} - \hat{\mathbf{v}}\|_2 \geq C_1\sigma\sqrt{n} - C_2\frac{\sigma\sqrt{n\log n}}{n^2} \gtrsim \sigma\sqrt{n}, \quad (\text{X.13})$$

where the second last inequality follows from the definition of \mathcal{A}_n (i.e., E_1 and E_2) and the assumption that $\|\mathbf{v} - \hat{\mathbf{v}}\|_2 \lesssim n^{-2}$. Finally, note that $\cos \angle(\mathbf{E}\hat{\mathbf{v}}, \hat{\mathbf{v}}) = \frac{\hat{\mathbf{v}}^\top \mathbf{E}\hat{\mathbf{v}}}{\|\mathbf{E}\hat{\mathbf{v}}\|_2}$. For the numerator, we have

$$|\hat{\mathbf{v}}^\top \mathbf{E}\hat{\mathbf{v}}| \leq |\mathbf{v}^\top \mathbf{E}\mathbf{v}| + |(\mathbf{v} - \hat{\mathbf{v}})^\top \mathbf{E}\mathbf{v}| + |\hat{\mathbf{v}}^\top \mathbf{E}(\mathbf{v} - \hat{\mathbf{v}})| \lesssim \sigma\sqrt{\log n} + \sigma\sqrt{\frac{\log n}{n^3}} \lesssim \sigma\sqrt{\log n},$$

where the second last inequality follows from the definition of \mathcal{A}_n (i.e., E_2 and E_3) and the assumption $\|\mathbf{v} - \hat{\mathbf{v}}\|_2 \lesssim n^{-2}$. For the denominator, we already have $\|\mathbf{E}\hat{\mathbf{v}}\|_2 \gtrsim \sigma\sqrt{n}$ for sufficiently large n . Hence, it follows that

$$\cos \angle(\mathbf{E}\hat{\mathbf{v}}, \hat{\mathbf{v}}) \leq C\sqrt{\frac{\log n}{n}}, \quad (\text{X.14})$$

which implies that, for any small constant $c > 0$, we have

$$\sin \angle(\mathbf{E}\hat{\mathbf{v}}, \hat{\mathbf{v}}) \geq 1 - c \quad (\text{X.15})$$

for all sufficiently large n . Plugging in the above results back to (X.11), by the above arguments (X.12) and (X.13), we can choose $\delta = C_0 \sigma n^2 \sqrt{n}$ for a sufficiently small $C_0 > 0$, such that $\|\mathbf{E}\hat{\mathbf{v}}\|_2 / \|\mathbf{B}\hat{\mathbf{v}}\|_2 > 1$. Therefore,

$$\sin \angle(\mathbf{B}\hat{\mathbf{v}}, (\mathbf{B} + \mathbf{E})\hat{\mathbf{v}}) \geq 1 - c \quad (\text{X.16})$$

for some small constant $c > 0$. This proves the first statement.

b) *Step II:* Suppose the first statement hold. Then for any small constant $c > 0$, we have

$$\cos(\hat{\mathbf{v}}, (\mathbf{B} - \mathbf{E})\hat{\mathbf{v}}) = \frac{\hat{\mathbf{v}}^\top (\mathbf{B} - \mathbf{E})\hat{\mathbf{v}}}{\|(\mathbf{B} - \mathbf{E})\hat{\mathbf{v}}\|_2} < c. \quad (\text{X.17})$$

For the numerator, we have

$$\hat{\mathbf{v}}^\top (\mathbf{B} - \mathbf{E})\hat{\mathbf{v}} = \sum_{i=1}^n \tilde{\lambda}_i (\tilde{\mathbf{u}}_i^\top \hat{\mathbf{v}})^2. \quad (\text{X.18})$$

Under the assumption $\|\mathbf{v} - \hat{\mathbf{v}}_-\|_2 \leq cn^{-2}$ and event \mathcal{A}_n (i.e., E_2 and E_3), we have

$$|\tilde{\lambda}_2 - \lambda_2| \leq C\sigma\sqrt{\log n}. \quad (\text{X.19})$$

To see this, note that $\tilde{\lambda}_2 = \hat{\mathbf{v}}_-^\top (\mathbf{B} - \mathbf{E})\hat{\mathbf{v}}_- = \lambda_2 (\mathbf{v}^\top \hat{\mathbf{v}}_-)^2 + \sum_{i \neq 2} \lambda_i (\mathbf{u}_i^\top \hat{\mathbf{v}}_-)^2 - \hat{\mathbf{v}}_-^\top \mathbf{E}\hat{\mathbf{v}}_-$. Then we have

$$\begin{aligned} |\tilde{\lambda}_2 - \lambda_2| &\leq \lambda_2 |1 - (\mathbf{v}^\top \hat{\mathbf{v}}_-)^2| + \lambda_n \sum_{i \neq 2} (\mathbf{u}_i^\top \hat{\mathbf{v}}_-)^2 + |\mathbf{v}^\top \mathbf{E}\mathbf{v}| + 2|(\mathbf{v} - \hat{\mathbf{v}}_-)^\top \mathbf{E}\mathbf{v}| \\ &\lesssim \frac{\lambda_2}{n^4} + \frac{\lambda_n}{n^3} + \sigma\sqrt{\log n} + \frac{\sigma\sqrt{n \log n}}{n^2}, \end{aligned}$$

where the last inequality follows from

$$|\hat{\mathbf{v}}_-^\top \hat{\mathbf{v}}| = |1 - \frac{1}{2} \|\hat{\mathbf{v}}_- - \hat{\mathbf{v}}\|_2^2| \geq 1 - c_1 n^{-4}, \quad (\text{X.20})$$

and

$$\sum_{i \neq 2} (\mathbf{u}_i^\top \hat{\mathbf{v}}_-)^2 \leq \sum_{i \neq 2} (\mathbf{u}_i^\top \mathbf{v}_-)^2 + \sum_{i \neq 2} [\mathbf{u}_i^\top (\mathbf{v}_- - \hat{\mathbf{v}}_-)]^2 \leq \frac{c_2}{n^3}, \quad (\text{X.21})$$

for sufficiently large n and some constants $c_1, c_2 > 0$. Thus, by (X.19), it holds that

$$\tilde{\lambda}_2 \geq \lambda_2 - |\tilde{\lambda}_2 - \lambda_2| \gtrsim \frac{\delta}{n^2} - \sigma\sqrt{\log n} > 0, \quad (\text{X.22})$$

for sufficiently large n . As $\mathbf{B} - \mathbf{E}$ is a Laplacian matrix, $\{\tilde{\lambda}_i\}_{1 \leq i \leq n}$ must contain an eigenvalue 0, it follows that $\tilde{\lambda}_1 = 0$, so that $\min_{1 \leq i \leq n} \tilde{\lambda}_i \geq 0$. This implies that

$$\hat{\mathbf{v}}^\top (\mathbf{B} - \mathbf{E})\hat{\mathbf{v}} \geq \tilde{\lambda}_2 (\tilde{\mathbf{u}}_2^\top \hat{\mathbf{v}})^2 = \tilde{\lambda}_2 (\hat{\mathbf{v}}_-^\top \hat{\mathbf{v}})^2 \geq (\lambda_2 - |\lambda_2 - \tilde{\lambda}_2|) \cdot |\hat{\mathbf{v}}_-^\top \hat{\mathbf{v}}| \cdot (1 - cn^{-4}) \gtrsim \frac{\delta |\hat{\mathbf{v}}_-^\top \hat{\mathbf{v}}|}{n^2}, \quad (\text{X.23})$$

where the second last inequality follows from the assumption $\|\mathbf{v} - \hat{\mathbf{v}}\|_2 + \|\mathbf{v} - \hat{\mathbf{v}}_-\|_2 \lesssim n^{-2}$, (X.20) and (X.22).

For the denominator of (X.17), we have

$$\begin{aligned} \|(\mathbf{B} - \mathbf{E})\hat{\mathbf{v}}\|_2 &\leq \|\mathbf{B}\mathbf{v}\|_2 + \|\mathbf{B}(\hat{\mathbf{v}} - \mathbf{v})\|_2 + \|\mathbf{E}\mathbf{v}\|_2 + \|\mathbf{E}(\hat{\mathbf{v}} - \mathbf{v})\|_2 \\ &\leq \lambda_2 + \frac{\lambda_n}{n^2} + C\sigma\sqrt{n} + \frac{C\sigma\sqrt{n\log n}}{n^2} \lesssim \frac{\delta}{n^2} \end{aligned}$$

Thus, we have

$$c > \frac{\hat{\mathbf{v}}^\top (\mathbf{B} - \mathbf{E}) \hat{\mathbf{v}}}{\|(\mathbf{B} - \mathbf{E})\hat{\mathbf{v}}\|_2} \geq C|\hat{\mathbf{v}}_-^\top \hat{\mathbf{v}}|. \quad (\text{X.24})$$

Since c can be chosen arbitrarily small, we conclude that $|\hat{\mathbf{v}}_-^\top \hat{\mathbf{v}}| \leq c'$ for some sufficiently small constant $c' > cC^{-1}$.

In other words, $\|\hat{\mathbf{v}} - \hat{\mathbf{v}}_-\|_2 \geq \epsilon_2$ for some constant $\epsilon_2 > 0$. This contradicts the first statement that $\|\hat{\mathbf{v}} - \hat{\mathbf{v}}_-\|_2 \leq cn^{-2}$.

This completes the proof of Step II.

c) *Step III*: Finally, we show each of the following events holds with probability at least $1 - n^{-c}$ for some constant $c > 0$:

$$E_1 = \left\{ \|\mathbf{E}\mathbf{v}\|_2 \geq C_1\sigma\sqrt{n} \right\}, \quad E_2 = \left\{ \|\mathbf{E}\| \leq C_2\sigma\sqrt{n\log n} \right\}, \quad E_3 = \left\{ |\mathbf{v}^\top \mathbf{E}\mathbf{v}| \leq C_3\sigma\sqrt{\log n} \right\}.$$

We start with E_3 .

For E_3 , note that

$$\mathbf{v}^\top \mathbf{E}\mathbf{v} = 2 \sum_{i>j} z_{ij}(v_i - v_j)^2,$$

is a normal random variable with mean 0 and variance

$$\mathbb{E} \left[2 \sum_{i>j} z_{ij}(v_i - v_j)^2 \right]^2 = 2 \sum_{i>j} \mathbb{E} z_{ij}^2 (v_i - v_j)^4 = 2\sigma^2 \sum_{i>j} (v_i - v_j)^4,$$

where

$$\mathbf{v} = (v_1, \dots, v_n)^\top = \frac{1}{\sqrt{n/2}} \left(\cos\left(\frac{\pi}{2n}\right), \cos\left(\frac{3\pi}{2n}\right), \dots, \cos\left(\frac{(2n-1)\pi}{2n}\right) \right)^\top. \quad (\text{X.25})$$

Calculate that

$$\sum_{i>j} (v_i - v_j)^4 = \sum_{i=1}^n [(v_i - v_{i+1})^4 + \dots + (v_i - v_n)^4] \leq n^2 \cdot \frac{c}{n^2} \leq c,$$

where the second last inequality follows from

$$\max_{i \neq j} |v_i - v_j| \leq \frac{c}{\sqrt{n}}. \quad (\text{X.26})$$

We know that $\mathbf{v}^\top \mathbf{E}\mathbf{v}$ is a normal random variable with bounded variance. Thus, by the tail bound of normal random variable, we have E_3 holds with probability at least $1 - n^{-c}$.

For E_1 , we denote $\mathbf{E}\mathbf{v} = (\xi_1, \dots, \xi_n)^\top$, so that

$$\xi_i = \sum_{j \neq i} z_{ij}(v_i - v_j), \quad i = 1, \dots, n.$$

In particular, we know that ξ_i is a normal random variable with mean 0 and variance

$$\mathbb{E}\xi_i^2 = \sum_{k, j \neq i} \mathbb{E}z_{ij}z_{ik}(v_i - v_j)(v_i - v_k) = \sigma^2 \sum_{j \neq i} (v_i - v_j)^2.$$

By (X.25), we have

$$\sum_{j \neq i} (v_i - v_j)^2 \leq n \cdot \frac{c}{n^2} \leq c.$$

If $i < n/2$, then

$$\sum_{j \neq i} (v_i - v_j)^2 \geq \sum_{j=i+n/4}^{i+n/2} (v_i - v_j)^2 \geq \frac{n}{4} \cdot \frac{c_1}{n} \geq c_2,$$

for some constants $c_1, c_2 > 0$. By a similar argument, the above inequality $\sum_{j \neq i} (v_i - v_j)^2 \geq c_2$ also holds if $i \geq n/2$. Thus we obtain $\mathbb{E}\xi_i^2 \asymp 1$ for all $i = 1, \dots, n$. Moreover, for any $i \neq j$, we have

$$\mathbb{E}\xi_i\xi_j = \sum_{\ell \neq j} \sum_{k \neq i} \mathbb{E}z_{ik}z_{j\ell}(v_i - v_k)(v_j - v_\ell) = \mathbb{E}z_{ij}^2(v_i - v_j)^2 = \sigma^2(v_i - v_j)^2.$$

In particular, by (X.26), we have $(v_i - v_j)^2 \leq c/n$. Therefore, $\mathbf{E}\mathbf{v}$ is a multivariate normal random vector with mean 0 and covariance matrix Σ_n , whose diagonal entries are of order σ^2 , and the off-diagonal entries are bounded in absolute value by $O(\sigma^2/n)$. By the property of multivariate normal distribution, there exists a matrix $\Sigma_n^{1/2}$ such that, for *i.i.d.* standard normal random variables $\{\zeta_i\}_{1 \leq i \leq n}$,

$$\mathbf{E}\mathbf{v} = (\xi_1, \dots, \xi_n)^\top = \Sigma_n^{1/2}(\zeta_1, \dots, \zeta_n)^\top.$$

Hence, the quadratic form $\mathbf{v}^\top \mathbf{E}^2 \mathbf{v}$ can be written as $\mathbf{v}^\top \mathbf{E}^2 \mathbf{v} = \zeta^\top \Sigma_n \zeta$, where $\zeta = (\zeta_1, \dots, \zeta_n)^\top$. Note that $\|\Sigma_n\|_F^2 \lesssim \sigma^2 n$. By the decomposition $\Sigma_n = \Sigma_n^D + \Sigma_n^O$ of Σ_n into a diagonal matrix Σ_n^D and an off-diagonal matrix Σ_n^O , we also have

$$\|\Sigma_n\| \leq \|\Sigma_n^D\| + \|\Sigma_n^O\| \leq \max_{1 \leq i \leq n} [\Sigma_n]_{ii} + \|\Sigma_n^O\|_F = O(\sigma^2).$$

Now we could apply Hanson-Wright inequality [73], to obtain the concentration inequality

$$P(|\zeta^\top \Sigma_n \zeta - \mathbb{E}\zeta^\top \Sigma_n \zeta| > t) \leq 2 \exp(-c \min\{t^2/(n\sigma^2), t/\sigma^4\}), \quad t \geq 0. \quad (\text{X.27})$$

By choosing $t = C\sigma^2\sqrt{n \log n}$, we have

$$P(|\zeta^\top \Sigma_n \zeta - \mathbb{E}\zeta^\top \Sigma_n \zeta| > \sigma^2\sqrt{n \log n}) \leq n^{-c}. \quad (\text{X.28})$$

Finally, since $\mathbb{E}\zeta^\top \Sigma_n \zeta = \sum_{i=1}^n [\Sigma_n]_{ii} \asymp \sigma^2 n$, we have

$$P(\zeta^\top \Sigma_n \zeta \leq \mathbb{E}\zeta^\top \Sigma_n \zeta - C\sigma^2 \sqrt{n \log n}) \leq n^{-c},$$

or

$$P(\mathbf{v}^\top \mathbf{E}^2 \mathbf{v} \geq C\sigma^2 n) \geq 1 - n^{-c}.$$

This proves E_1 holds with probability at least $1 - n^{-c}$, by noting that $\mathbf{v}^\top \mathbf{E}^2 \mathbf{v} = \|\mathbf{E}\mathbf{v}\|_2^2$.

For E_2 , to obtain an upper bound for $\|\mathbf{E}\|$, we note that

$$\|\mathbf{E}\| = \|\mathbf{L}(\mathbf{Z})\| \leq \|\mathbf{D}(\mathbf{Z})\| + \|\mathbf{Z}\| \leq \max_{1 \leq i \leq n} \sum_{j=1}^n Z_{ij} + \|\mathbf{Z}\|.$$

Since the Laplacian operator $\mathbf{L}(\cdot)$ is invariant to the diagonal entries of the input, we assume without loss of generality that the diagonal entries of \mathbf{Z} is *i.i.d.* Gaussian $N(0, 2\sigma^2)$, so that \mathbf{Z}/σ is Gaussian Orthogonal Ensemble (GOE). The upper bounds for $\max_{1 \leq i \leq n} \sum_{j=1}^n Z_{ij}$ and $\|\mathbf{Z}\|$ are obtained previously. Specifically, the well-celebrated Bai-Yin theorem [74] implies that for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left(\frac{\|\mathbf{Z}\|}{\sigma\sqrt{n}} \leq 2 + \epsilon\right) = 1, \quad (\text{X.29})$$

whereas the standard tail bound for the Gaussian random variable and an union bound argument imply that for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left(\max_{1 \leq i \leq n} \sum_{j=1}^n Z_{ij} \leq \sigma\sqrt{(2 + \epsilon)(n + 1) \log n}\right) = 1. \quad (\text{X.30})$$

These results imply

$$\lim_{n \rightarrow \infty} P(\|\mathbf{E}\| \leq 8\sigma\sqrt{n \log n}) = 1. \quad (\text{X.31})$$

D. Theoretical Guarantee for λ -Ridged Toeplitz Matrices: Proof of Proposition 4

The proof of this proposition is separated into two parts, corresponding to the initialization (Step 1) and the iterative sorting part (Step 2) of the algorithm, respectively.

a) *Part I:* We start by showing that the initialization step successfully identifies the first or the last row of Θ . Set $\delta_n = n^{-c}$ for some $c > 0$. Without loss of generality, we assume $\Pi = \text{Id}_n$. In other words, we need to show that,

$$P_{\Theta, \Pi}(\tilde{\pi}(1) \in \{1, n\}) \geq 1 - \delta_n/n, \quad (\text{X.32})$$

for any $(\Theta, \Pi) \in \mathcal{R}_n^R(\lambda) \times \mathcal{S}_n$. In fact, (X.32) is implied by

$$P_{\Theta, \Pi}(S_1 < \min_{2 \leq i \leq n-1} S_i) \geq 1 - \delta_n/n. \quad (\text{X.33})$$

Then it suffices to show that, for some τ_n ,

$$P_{\Theta, \Pi}(S_1 < \tau_n) \geq 1 - \frac{\delta_n}{2n}, \quad (\text{X.34})$$

and

$$P_{\Theta, \Pi}\left(\min_{2 \leq i \leq n-1} S_i > \tau_n\right) \geq 1 - \frac{\delta_n}{2n}. \quad (\text{X.35})$$

The rest of the proof is devoted to (X.34) and (X.35).

On the one hand, let $S_1 = \sum_{i \neq 1} Y_{1i} = \sum_{1 \leq i \leq n-1} \theta_i + \sum_{i \neq 1} Z_{1i}$ where Z_{1i} 's are independent subgaussian random variables. By the concentration inequality for subgaussian random variables [75], we have

$$P\left(S_1 - \sum_{i=1}^{n-1} \theta_i < C\sigma\sqrt{nt}\right) \geq 1 - e^{-t^2}. \quad (\text{X.36})$$

On the other hand, for any $2 \leq i \leq n-1$, we have $S_i = \sum_{j \neq i} Y_{ij} = \sum_{j=1}^{i-1} \theta_j + \sum_{j=1}^{n-i} \theta_j + \sum_{j \neq i} Z_{ij}$ where Z_{ij} 's are independent subgaussian random variables, so that

$$P\left(S_i - \sum_{j=1}^{i-1} \theta_j - \sum_{j=1}^{n-i} \theta_j > -C\sigma\sqrt{nt}\right) \geq 1 - e^{-t^2}. \quad (\text{X.37})$$

By setting $t = \sqrt{\log(4n/\delta_n)}$ in (X.36) and $t = \sqrt{\log(4n^2/\delta_n)}$ in (X.37), we have

$$P\left(S_1 < \sum_{i=1}^{n-1} \theta_i + C\sigma\sqrt{n \log(4n/\delta_n)}\right) \geq 1 - \frac{\delta_n}{4n}, \quad (\text{X.38})$$

$$P\left(S_i > \sum_{j=1}^{i-1} \theta_j + \sum_{j=1}^{n-i} \theta_j - C\sigma\sqrt{n \log(4n^2/\delta_n)}\right) \geq 1 - \frac{\delta_n}{4n^2}. \quad (\text{X.39})$$

Since for any $2 \leq i \leq n-1$, we have

$$\sum_{j=1}^{i-1} \theta_j + \sum_{j=1}^{n-i} \theta_j \geq \theta_1 + \sum_{j=1}^{n-2} \theta_j,$$

then we have

$$P\left(S_i > \theta_1 + \sum_{j=1}^{n-2} \theta_j - C\sigma\sqrt{n \log(4n^2/\delta_n)}\right) \geq 1 - \frac{\delta_n}{4n^2}$$

for all $2 \leq i \leq n-1$. A union bound applied to (X.39) implies

$$P\left(\min_{2 \leq i \leq n} S_i > \theta_1 + \sum_{j=1}^{n-2} \theta_j - C\sigma\sqrt{n \log(4n^2/\delta_n)}\right) \geq 1 - \frac{\delta_n}{4n}.$$

Now note that by assumption we have

$$\theta_1 + \sum_{j=1}^{n-2} \theta_j - C\sigma\sqrt{n \log(4n^2/\delta_n)} \geq \sum_{i=1}^{n-1} \theta_i + C\sigma\sqrt{n \log(4n/\delta_n)},$$

as for $\lambda \geq 4C\sigma n$ we have

$$\theta_1 - \theta_{n-1} \geq \lambda \geq C\sigma\sqrt{n \log(4n/\delta_n)} + C\sigma\sqrt{n \log(4n^2/\delta_n)},$$

for sufficiently large n . Therefore, we have shown (X.34) and (X.35) with $\tau_n = C\sigma\sqrt{n \log(4n/\delta_n)} + C\sigma\sqrt{n \log(4n^2/\delta_n)}$.

This proves (X.32).

b) *Part II:* In this part, we show that, for each $1 \leq k \leq n-2$,

$$P_{\Theta, \Pi}(\tilde{\pi}(k+1) = k+1 | \tilde{\pi}(i) = i, \forall i \leq k) \geq 1 - \delta_n/n, \quad (\text{X.40})$$

$$P_{\Theta, \Pi}(\tilde{\pi}(k+1) = n-k | \tilde{\pi}(i) = n-i+1, \forall i \leq k) \geq 1 - \delta_n/n. \quad (\text{X.41})$$

This along with Part I implies that

$$P_{\Theta, \Pi}(\tilde{\pi} = id) = P(\tilde{\pi}(1) = 1) \prod_{k=1}^{n-2} P_{\Theta, \Pi}(\tilde{\pi}(k+1) = k+1 | \tilde{\pi}(i) = i, \forall i \leq k) \geq P(\tilde{\pi}(1) = 1)(1 - \delta_n/n)^{n-2},$$

$$\begin{aligned} P_{\Theta, \Pi}(\tilde{\pi} = id^{-1}) &= P(\tilde{\pi}(1) = n) \prod_{k=1}^{n-2} P_{\Theta, \Pi}(\tilde{\pi}(k+1) = n-k | \tilde{\pi}(i) = n-i+1, \forall i \leq k) \\ &\geq P(\tilde{\pi}(1) = n)(1 - \delta_n/n)^{n-2}, \end{aligned}$$

for $n \geq 1$. Combining the above two inequalities, by the result from Part I we have

$$\begin{aligned} P_{\Theta, \Pi}(\tilde{\pi} \in \{id, id^{-1}\}) &\geq [P(\tilde{\pi}(1) = 1) + P(\tilde{\pi}(1) = n)](1 - \delta_n/n)^{n-2} \geq P(\tilde{\pi}(1) \in \{1, n\})(1 - \delta_n/n)^{n-2} \\ &\geq (1 - \delta_n/n)^{n-1} \geq 1 - \delta_n, \end{aligned}$$

which in turn proves the proposition.

The rest of the proof is devoted to (X.40), as the proof of (X.41) follows by symmetry. Suppose $\tilde{\pi}(i) = i, \forall i \leq k$. Then the event $\tilde{\pi}(k+1) = k+1$ is equivalent to

$$\|\mathbf{Y}_{k, -k} - \mathbf{Y}_{k+1, -(k+1)}\|_1 < \min_{j \in [n] \setminus [k+1]} \|\mathbf{Y}_{k, -k} - \mathbf{Y}_{j, -j}\|_1,$$

or

$$\|\mathbf{Y}_{k, -k} - \mathbf{Y}_{k+1, -(k+1)}\|_1 < \|\mathbf{Y}_{k, -k} - \mathbf{Y}_{j, -j}\|_1, \quad \text{for all } j \in [n] \setminus [k+1].$$

In the following, we show that, for any $j \in [n] \setminus [k+1]$, it holds that

$$P(\|\mathbf{Y}_{k, -k} - \mathbf{Y}_{k+1, -(k+1)}\|_1 < \|\mathbf{Y}_{k, -k} - \mathbf{Y}_{j, -j}\|_1) \geq 1 - \frac{\delta_n}{2n^2}. \quad (\text{X.42})$$

Thus, by applying the union bound, we have

$$P(\|\mathbf{Y}_{k,-k} - \mathbf{Y}_{k+1,-(k+1)}\|_1 \leq \min_{j \in [n] \setminus [k+1]} \|\mathbf{Y}_{k,-k} - \mathbf{Y}_{j,-j}\|_1) \geq 1 - \frac{\delta_n}{2n}. \quad (\text{X.43})$$

To obtain (X.42), we will show that there exists some τ_n such that

$$P(\|\mathbf{Y}_{k,-k} - \mathbf{Y}_{k+1,-(k+1)}\|_1 \leq \tau_n) \geq 1 - \frac{\delta_n}{4n^2}, \quad (\text{X.44})$$

$$P(\tau_n \leq \|\mathbf{Y}_{k,-k} - \mathbf{Y}_{j,-j}\|_1) \geq 1 - \frac{\delta_n}{4n^2}. \quad (\text{X.45})$$

On the one hand, if we denote $\mathbf{Y}_{k,-k} - \mathbf{Y}_{k+1,-(k+1)} = (\xi_1^{(k+1)}, \xi_2^{(k+1)}, \dots, \xi_{n-1}^{(k+1)})$, it follows that $\{\xi_i^{(k+1)}\}$ are independent subgaussian variables with variances bounded by $C\sigma^2$, and means satisfying

$$\mathbb{E}\xi_i^{(k+1)} = \theta_{k-i} - \theta_{k-i+1}, \quad \text{for } i < k, \quad (\text{X.46})$$

$$\mathbb{E}\xi_i^{(k+1)} = \theta_{i-k+1} - \theta_{i-k}, \quad \text{for } i > k, \quad (\text{X.47})$$

and $\mathbb{E}\xi_k^{(k+1)} = 0$. To see this, note that the k -th row of the signal matrix Θ with the main diagonal removed is

$$\underbrace{[\theta_{k-1} \ \theta_{k-2} \ \dots \ \theta_2 \ \theta_1]}_{k-1} \quad \underbrace{[\theta_1 \ \theta_2 \ \dots \ \theta_{n-k}]}_{n-k}$$

Then, we can write $\|\mathbf{Y}_{k,-k} - \mathbf{Y}_{k+1,-(k+1)}\|_1 = \sum_{i=1}^{n-1} |\xi_i^{(k+1)}| = \sum_{i=1}^{n-1} |\mathbb{E}\xi_i^{(k+1)} + w_i|$ for some independent subgaussian random variables w_i with mean 0 and variance bounded by $C\sigma^2$. By the concentration inequality for subgaussian random variables, we have

$$P\left(\sum_{i=1}^{n-1} |\xi_i^{(k+1)}| \leq \sum_{i=1}^{n-1} \mathbb{E}\xi_i^{(k+1)} + \sum_{i=1}^{n-1} |w_i| \leq \tau_1\right) \geq 1 - \frac{\delta_n}{2n^2},$$

where $\tau_1 = \sum_{i < k} |\theta_{k-i} - \theta_{k-i+1}| + \sum_{i > k} |\theta_{i-k+1} - \theta_{i-k}| + C\sigma n + C\sigma\sqrt{n \log(2n^2/\delta_n)}$. On the other hand, for any $j \in [n] \setminus [k+1]$, if we denote $\mathbf{Y}_{k,-k} - \mathbf{Y}_{j,-j} = (\xi_1^{(j)}, \xi_2^{(j)}, \dots, \xi_{n-1}^{(j)})$, it follows that $\{\xi_i^{(j)}\}$ are subgaussian variables with

$$\mathbb{E}\xi_i^{(j)} = \theta_{k-i} - \theta_{j-i}, \quad \text{for } i < k,$$

$$\mathbb{E}\xi_i^{(j)} = \theta_{i-k+1} - \theta_{j-i}, \quad \text{for } k \leq i < j,$$

$$\mathbb{E}\xi_i^{(j)} = \theta_{i-k+1} - \theta_{i-j+1}, \quad \text{for } i \geq j.$$

In addition, out of the $(n-1)$ elements in $\{\xi_i^{(j)}\}$, $(n-3)$ elements are mutually independent, whereas the rest two elements, corresponding to the entries Y_{kj} and Y_{jk} , are correlated with covariance bounded by $C\sigma^2$, but independent from the others. By the similar argument for the $(n-3)$ independent variables and by the sub-Gaussian property

of the two dependent variables, we still have

$$P\left(\sum_{i=1}^{n-1} |\xi_i^{(j)}| \geq \tau_2\right) \geq 1 - \frac{\delta_n}{2n^2},$$

where $\tau_2 = \sum_{i < k} |\theta_{k-i} - \theta_{j-i}| + \sum_{k \leq i < j} |\theta_{i-k+1} - \theta_{j-i}| + \sum_{i \geq j} |\theta_{i-k+1} - \theta_{i-j+1}| - Cn\sigma - C\sigma\sqrt{n \log(2n^2/\delta_n)}$.

Now we claim that, if $\lambda \geq 4C\sigma n$, we have, $\tau_2 \geq \tau_1$ for sufficiently large n , or

$$\begin{aligned} & \sum_{i < k} |\theta_{k-i} - \theta_{j-i}| + \sum_{k \leq i < j} |\theta_{i-k+1} - \theta_{j-i}| + \sum_{i \geq j} |\theta_{i-k+1} - \theta_{i-j+1}| \\ & \geq \sum_{i < k} |\theta_{k-i} - \theta_{k-i+1}| + \sum_{i > k} |\theta_{i-k+1} - \theta_{i-k}| + 2Cn\sigma + 2C\sigma\sqrt{n \log(4n^2/\delta_n)}. \end{aligned} \quad (\text{X.48})$$

To see this, it suffices to show that, by the definition of $\mathcal{T}_n^R(\lambda)$, it holds that

$$\begin{aligned} & \sum_{i < k} |\theta_{k-i} - \theta_{j-i}| + \sum_{k \leq i < j} |\theta_{i-k+1} - \theta_{j-i}| + \sum_{i \geq j} |\theta_{i-k+1} - \theta_{i-j+1}| \\ & - \sum_{i < k} |\theta_{k-i} - \theta_{k-i+1}| - \sum_{i > k} |\theta_{i-k+1} - \theta_{i-k}| \geq \sum_{i=1}^{\lceil n/2 \rceil - 1} |\theta_i - \theta_{i+1}| \geq \lambda. \end{aligned} \quad (\text{X.49})$$

Specifically, if we denote $\delta_k = \theta_k - \theta_{k+1}$ for $1 \leq k \leq n-2$, then

$$\sum_{i < k} |\theta_{k-i} - \theta_{k-i+1}| + \sum_{i > k} |\theta_{i-k+1} - \theta_{i-k}| = \sum_{i < k} \delta_{k-i} + \sum_{i > k} \delta_{i-k} = \sum_{i=1}^{k-1} \delta_i + \sum_{i=1}^{n-k-1} \delta_i,$$

and

$$\begin{aligned} & \sum_{i < k} |\theta_{k-i} - \theta_{j-i}| + \sum_{k \leq i < j} |\theta_{i-k+1} - \theta_{j-i}| + \sum_{i \geq j} |\theta_{i-k+1} - \theta_{i-j+1}| \\ & \geq \sum_{i < k} |\theta_{k-i} - \theta_{j-i}| + 2\delta_1 + \sum_{i \geq j} |\theta_{i-k+1} - \theta_{i-j+1}| \\ & = \sum_{i < k} (\delta_{k-i} + \delta_{k-i+1} + \dots + \delta_{j-i-1}) + \sum_{i \geq j} (\delta_{i-j+1} + \delta_{i-j+2} + \dots + \delta_{i-k}) + 2\delta_1 \\ & = \sum_{i=1}^{k-1} (\delta_i + \delta_{i+1} + \dots + \delta_{i+j-k-1}) + \sum_{i=1}^{n-j+1} (\delta_i + \delta_{i+1} + \dots + \delta_{i+j-k-1}) + 2\delta_1. \end{aligned} \quad (\text{X.50})$$

Thus, we have

$$\begin{aligned} & \sum_{i < k} |\theta_{k-i} - \theta_{j-i}| + \sum_{k \leq i < j} |\theta_{i-k+1} - \theta_{j-i}| + \sum_{i \geq j} |\theta_{i-k+1} - \theta_{i-j+1}| - \sum_{i < k} |\theta_{k-i} - \theta_{k-i+1}| - \sum_{i > k} |\theta_{i-k+1} - \theta_{i-k}| \\ & \geq \sum_{i=1}^{k-1} (\delta_i + \delta_{i+1} + \dots + \delta_{i+j-k-1}) + \sum_{i=1}^{n-j+1} (\delta_i + \delta_{i+1} + \dots + \delta_{i+j-k-1}) + 2\delta_1 - \sum_{i=1}^{k-1} \delta_i - \sum_{i=1}^{n-k-1} \delta_i \\ & \geq \sum_{i=1}^k \delta_i + \sum_{i=1}^{n-k} \delta_i \geq \sum_{i=1}^{\lceil n/2 \rceil - 1} \delta_i, \end{aligned}$$

where in the second last inequality we used the fact that

$$\sum_{i=1}^{n-j+1} (\delta_i + \delta_{i+1} + \dots + \delta_{i+j-k-1}) - \sum_{i=1}^{n-k-1} \delta_i \geq \sum_{i=2}^{n-k} \delta_i. \quad (\text{X.51})$$

To see this, note that

$$\begin{aligned} & \sum_{i=1}^{n-j+1} (\delta_i + \delta_{i+1} + \dots + \delta_{i+j-k-1}) - \sum_{i=1}^{n-k-1} \delta_i \\ &= (\delta_1 + \delta_2 + \dots + \delta_{j-k}) + (\delta_2 + \delta_3 + \dots + \delta_{j-k+1}) + \dots + (\delta_{n-j+1} + \dots + \delta_{n-k}) - \sum_{i=1}^{n-k-1} \delta_i. \end{aligned}$$

If $n-j+1 < j-k$, then we have

$$(\delta_1 + \delta_2 + \dots + \delta_{j-k}) + (\delta_{n-j+1} + \dots + \delta_{n-k}) - \sum_{i=1}^{n-k-1} \delta_i \geq \delta_{n-k},$$

so that

$$\begin{aligned} & (\delta_1 + \delta_2 + \dots + \delta_{j-k}) + (\delta_2 + \delta_3 + \dots + \delta_{j-k+1}) + \dots + (\delta_{n-j+1} + \dots + \delta_{n-k}) - \sum_{i=1}^{n-k-1} \delta_i \\ & \geq (\delta_2 + \delta_3 + \dots + \delta_{j-k+1}) + \dots + (\delta_{n-j} + \dots + \delta_{n-k-1}) + \delta_{n-k} \geq \sum_{i=2}^{n-k} \delta_i. \end{aligned}$$

If $n-j+1 \geq j-k$, then we can "extract" the first terms in each of the first $n-j$ sums, and all but the last term δ_{n-k} in the $(n-j+1)$ -th sum, and get

$$\begin{aligned} & (\delta_1 + \delta_2 + \dots + \delta_{j-k}) + (\delta_2 + \delta_3 + \dots + \delta_{j-k+1}) + \dots + (\delta_{n-j+1} + \dots + \delta_{n-k}) - \sum_{i=1}^{n-k-1} \delta_i \\ & \geq (\delta_2 + \delta_3 + \dots + \delta_{j-k}) + (\delta_3 + \delta_4 + \dots + \delta_{j-k+1}) + \dots + (\delta_{n-j+1} + \dots + \delta_{n-k-1}) + \delta_{n-k} \geq \sum_{i=2}^{n-k} \delta_i. \end{aligned}$$

This proves (X.49) or (X.48) under the condition that $\Theta \in \mathcal{T}_n^R(\lambda)$. Thus, we can take any $\tau_n \in [\tau_1, \tau_2]$, to get (X.44) and (X.45), which imply (X.42) and (X.43). This proves (X.40) under the conditions of the proposition.

E. Sharp λ - ρ^* Correspondence: Proof of Proposition 5

a) *Sufficiency*: By the conditions of Proposition 5, for all $\Theta \in \mathcal{T}'_n$, we have $\rho(\Theta, \mathcal{S}'_n) \geq Cn\lambda^*$. Denote the diagonal values of Θ as $\{\theta_0, \dots, \theta_{n-1}\}$. By the definition of $\rho(\Theta, \mathcal{S}'_n)$, we have

$$[\rho(\Theta, \mathcal{S}'_n)]^2 = \min_{\Pi_1, \Pi_2 \in \mathcal{S}'_n} \|\Pi_1 \Theta \Pi_1^\top - \Pi_2 \Theta \Pi_2^\top\|_F^2 \leq n(n-1) \cdot 4(\theta_1 - \theta_{\lceil n/2 \rceil})^2,$$

where the last inequality follows from the fact that $(\theta_j - \theta_k)^2 \leq (\theta_1 - \theta_n)^2$ for any $j \neq k$, and the "ridge" condition $\theta_1 - \theta_{\lceil n/2 \rceil} \geq \theta_{\lceil n/2 \rceil} - \theta_{n-1}$. Thus, $\rho(\Theta, \mathcal{S}'_n) \geq Cn\lambda^*$ implies $\theta_1 - \theta_{\lceil n/2 \rceil} \geq \frac{C\lambda^*}{2}$, that is, $\Theta \in \mathcal{T}_n^R(\lambda)$ for $\lambda = C_2\lambda^*$ where $C_2 = C/2$. Since this holds for any such $\Theta \in \mathcal{T}'_n$, we have proven the first part of Proposition 5.

b) *Necessity*: To begin with, we first show that, there exists some $\mathcal{T}'_n \subseteq \mathcal{T}_n$ and $\mathcal{S}'_n \subseteq \mathcal{S}$ such that

$$\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) = C\sigma n\lambda, \quad \mathcal{T}'_n \subseteq \mathcal{T}_n^R(\lambda), \quad (\text{X.52})$$

for some absolute constant $C > 0$. To prove this, without loss of generality, we assume $n = 4m$ for some integer $m > 0$, and consider the class of symmetric Toeplitz matrices

$$\mathcal{T}'_n = \left\{ \Theta \in \mathcal{T}_n : \theta_0 = 0, \theta_k = \alpha + (n-k)\beta, \forall k \in \{1, 2, \dots, n-1\}, \alpha \geq 0 \right\},$$

where $\beta(n/2-1) = \lambda$. As a result, we can easily check that $\mathcal{T}'_n \subseteq \mathcal{T}_n^R(\lambda)$. On the other hand, we consider the permutation set $\mathcal{S}'_n \subseteq \mathcal{S}_n$ including only two permutations $\pi_0 = id$ and $\pi_1 = (n, \dots, 3n/4+1, n/4+1, \dots, 3n/4, n/4, \dots, 1)$. In other words, π_1 exchanges the first $n/4$ elements with the last $n/4$ elements, and arranges them in the reversed order. Let Π_0 and Π_1 be the permutation matrices associated with π_0 and π_1 , respectively. In the following, we show that for any $\Theta \in \mathcal{T}'_n$, we have

$$\|\Pi_0 \Theta \Pi_0^\top - \Pi_1 \Theta \Pi_1^\top\|_F^2 = C_0 n^2 \lambda^2, \quad (\text{X.53})$$

which implies (X.52). To obtain (X.53), we denote $\mathbf{A} = (a_{ij}) = \Pi_0 \Theta \Pi_0^\top$ and $\mathbf{B} = (b_{ij}) = \Pi_1 \Theta \Pi_1^\top$, and calculate the differences $\|\mathbf{A}_{\cdot i} - \mathbf{B}_{\cdot i}\|_2^2$ carefully.

Due to the invariance of the difference with respect to a translation of Θ , we assume without loss of generality that $\alpha = 0$. For $i = 1$, we have

$$\begin{aligned} \mathbf{A}_{\cdot 1} &= \Theta_{\cdot 1} = (0, (n-1)\beta, (n-2)\beta, \dots, 2\beta, \beta)^\top, \\ \mathbf{B}_{\cdot 1} &= \Pi_1 \Theta_{\cdot n} = \Pi_1 \cdot (\beta, 2\beta, \dots, (n-1)\beta, 0)^\top \\ &= \underbrace{(0, (n-1)\beta, (n-2)\beta, \dots, (3n/4+1)\beta)}_{n/4}, \underbrace{(n/4+1)\beta, \dots, (3n/4)\beta)}_{n/2}, \underbrace{(n/4)\beta, \dots, 2\beta, \beta)^\top}_{n/4}. \end{aligned}$$

Thus,

$$\|\mathbf{A}_{\cdot 1} - \mathbf{B}_{\cdot 1}\|_2^2 = 2\beta^2 \sum_{k=0}^{n/4-1} (n/2 - 2k - 1)^2 = 2\beta^2 \sum_{k=1}^{n/4} (2k-1)^2 = \frac{\beta^2 n(n/2-1)(n/2+1)}{6}.$$

For $i = 2$, we have

$$\begin{aligned} \mathbf{A}_{\cdot 2} &= \Theta_{\cdot 2} = ((n-1)\beta, 0, (n-1)\beta, (n-2)\beta, \dots, 3\beta, 2\beta)^\top \\ \mathbf{B}_{\cdot 2} &= \Pi_1 \Theta_{\cdot (n-1)} = \Pi_1 \cdot (2\beta, 3\beta, \dots, (n-1)\beta, 0, (n-1)\beta)^\top \\ &= \underbrace{((n-1)\beta, 0, (n-1)\beta, \dots, (3n/4+2)\beta)}_{n/4}, \underbrace{(n/4+2)\beta, \dots, (3n/4+1)\beta)}_{n/2}, \underbrace{(n/4+1)\beta, \dots, 3\beta, 2\beta)^\top}_{n/4}. \end{aligned}$$

Thus,

$$\|\mathbf{A}_{.2} - \mathbf{B}_{.2}\|_2^2 = 2\beta^2 \sum_{k=1}^{n/4} (2k-1)^2 = \frac{\beta^2 n(n/2-1)(n/2+1)}{6}.$$

Similarly, one can show that,

$$\|\mathbf{A}_{.i} - \mathbf{B}_{.i}\|_2^2 = \frac{\beta^2 n(n/2-1)(n/2+1)}{6}, \quad \text{for all } i \in \{1, 2, \dots, n/4\}. \quad (\text{X.54})$$

Hence, by adding up the above differences, we have

$$\|\mathbf{A} - \mathbf{B}\|_F^2 \geq \frac{\beta^2 n^2(n/2-1)(n/2+1)}{24} \geq \frac{\beta^2 n^4}{24 \times 8}$$

for $n \geq 3$. On the other hand, we have

$$\|\mathbf{A} - \mathbf{B}\|_F^2 \leq 2\|\mathbf{A}\|_F^2 + 2\|\mathbf{B}\|_F^2 = 4\|\Theta\|_F^2 \leq 2\beta^2 \sum_{k=1}^{n-1} k^3 = \frac{\beta^2 n^2(n-1)^2}{2} \leq \beta^2 n^4.$$

Combining the above results, we have shown $\|\mathbf{A} - \mathbf{B}\|_F^2 \asymp \beta^2 n^4 \asymp \lambda^2 n^2$, or (X.53).

F. An Information Lower Bound for Adaptive Sorting

In this part, we show that, for the adaptive sorting to achieve exact matrix reordering, a minimal signal strength condition that is strictly stronger than that required by the constrained LSE is in fact needed. The result is summarized as the following theorem.

Theorem 6: Suppose the noise matrix has *i.i.d.* entries up to symmetry generated from $N(0, \sigma^2)$. Then there exists some $\mathcal{T}'_n \subset \mathcal{T}_n$ and $\mathcal{S}'_n \subseteq \mathcal{S}_n$ satisfying $\rho^*(\mathcal{T}'_n, \mathcal{S}'_n) = C\sigma n^{3/2}$ for some absolute constant $C > 0$, such that $\sup_{(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n} P_{\Theta, \Pi}(\tilde{\Pi}\Theta\tilde{\Pi}^\top \neq \Pi\Theta\Pi^\top) \geq 0.2$.

In particular, in light of the second part of Proposition 5, it suffices to show the following proposition.

Proposition 6 (Lower bound for λ -ridged Toeplitz matrices): Suppose the noise matrix \mathbf{Z} has *i.i.d.* entries up to symmetry generated from $N(0, \sigma^2)$. Then there exists some $\mathcal{T}'_n \subseteq \mathcal{T}_n^R(\lambda)$ where $\lambda = c\sigma\sqrt{n}$ for some absolute constant $c > 0$, and $\mathcal{S}'_n \subseteq \mathcal{S}_n$, such that $\sup_{(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n} P_{\Theta, \Pi}(\tilde{\Pi}\Theta\tilde{\Pi}^\top \neq \Pi\Theta\Pi^\top) \geq 0.2$.

Proof. To prove this proposition, we start with the set of symmetric Toeplitz matrices

$$\mathcal{T}'_n = \left\{ \Theta \in \mathcal{T}_n : \theta_0 = 0, \theta_k = \alpha + (n-k)\beta, \forall k \in \{1, 2, \dots, n-1\}, \alpha \geq 0 \right\},$$

with $\beta = \lambda/(n-2)$, and the permutation set $\mathcal{S}'_n \subseteq \mathcal{S}_n$ including only two permutations $\pi_0 = id$ and $\pi_1 = (n, \dots, 3n/4+1, n/4+1, \dots, 3n/4, n/4, \dots, 1)$, constructed in the proof of the second statement of Proposition 5.

In particular, we can easily check that $\mathcal{T}'_n \subseteq \mathcal{T}_n^R(\lambda)$, and that, for any $\Theta \in \mathcal{T}'_n$, we have

$$\|\Pi_0\Theta\Pi_0^\top - \Pi_1\Theta\Pi_1^\top\|_F^2 = C_0 n^2 \lambda^2, \quad (\text{X.55})$$

In what follows, we will show that, for $\lambda = c\sigma\sqrt{n}$, we have $P_{\Theta, \Pi}(\tilde{\Pi}\Theta\tilde{\Pi}^\top \neq \Pi\Theta\Pi^\top) \geq 0.2$ for any $(\Theta, \Pi) \in \mathcal{T}'_n \times \mathcal{S}'_n$. In fact, we only need to show that, in such a case,

$$P_{\Theta, \Pi}(\tilde{\pi}(1) \notin \{1, n\}) \geq 0.2. \quad (\text{X.56})$$

To show this, we show that

$$P_{\Theta, \Pi}(S_1 > S_2) \geq 0.49, \quad P_{\Theta, \Pi}(S_n > S_{n-1}) \geq 0.49, \quad (\text{X.57})$$

where $S_j = \sum_{i \neq j} Y_{ji}$. If we are able to show that event $\{S_1 > S_2\}$ is independent of the event $\{S_n > S_{n-1}\}$, then we have

$$\begin{aligned} P_{\Theta, \Pi}(\tilde{\pi}(1) \notin \{1, n\}) &\geq P_{\Theta, \Pi}(\min\{S_1, S_n\} > \min_{2 \leq i \leq n-1} S_i) \geq P_{\Theta, \Pi}(S_1 > S_2, S_n > S_{n-1}) \\ &\geq P_{\Theta, \Pi}(S_1 > S_2) \cdot P_{\Theta, \Pi}(S_n > S_{n-1}) \geq 0.2. \end{aligned}$$

The rest of the proof is devoted to (X.57) and the independence between $\{S_1 > S_2\}$ and $\{S_n > S_{n-1}\}$.

On the one hand, note that $S_1 - S_2 = \sum_{i \neq 1} Y_{1i} - \sum_{i \neq 2} Y_{2i} = \sum_{i=1}^{n-1} \theta_i + \sum_{i \neq 1} Z_{1i} - \theta_1 - \sum_{j=1}^{n-2} \theta_j - \sum_{j \neq 2} Z_{2j}$ where Z_{ki} 's are independent Gaussian random variables for $k = 1, 2$. By the tail bound of the Gaussian random variable $\sum_{i \neq 1} Z_{1i} - \sum_{j \neq 2} Z_{2j} \sim N(0, 2(n-1)\sigma^2)$, we have

$$P(S_1 - S_2 \geq 0) = P\left(\theta_{n-1} - \theta_1 + \sum_{i \neq 1} Z_{1i} - \sum_{j \neq 2} Z_{2j} \geq 0\right) = \Phi\left(\frac{\theta_{n-1} - \theta_1}{\sigma\sqrt{2(n-1)}}\right) \geq 0.49$$

if $\lambda = c\sigma\sqrt{n}$ for some small constant $c > 0$. Similarly, we can also show $P(S_n - S_{n-1} \geq 0) \geq 0.49$. This completes the proof of (X.57).

On the other hand, we write

$$\begin{aligned} S_1 - S_2 &= \sum_{i=1}^{n-1} \theta_i - \theta_1 - \sum_{j=1}^{n-2} \theta_j + \sum_{i \neq 1} Z_{1i} - \sum_{j \neq 2} Z_{2j} \\ &= \sum_{i=1}^{n-1} \theta_i - \theta_1 - \sum_{j=1}^{n-2} \theta_j + \sum_{i \notin \{1, 2, n-1, n\}} Z_{1i} - \sum_{j \notin \{1, 2, n-1, n\}} Z_{2j} \\ &\quad + Z_{1, n-1} + Z_{1, n} - Z_{2, n-1} - Z_{2, n}, \end{aligned}$$

and

$$\begin{aligned}
S_n - S_{n-1} &= \sum_{i=1}^{n-1} \theta_i - \theta_1 - \sum_{j=1}^{n-2} \theta_j + \sum_{i \neq n} Z_{n,i} - \sum_{j \neq n-1} Z_{n-1,j} \\
&= \sum_{i=1}^{n-1} \theta_i - \theta_1 - \sum_{j=1}^{n-2} \theta_j + \sum_{i \notin \{1,2,n-1,n\}} Z_{n,i} - \sum_{j \notin \{1,2,n-1,n\}} Z_{n-1,j} \\
&\quad + Z_{1,n-1} + Z_{2,n-1} - Z_{1,n} - Z_{2,n}.
\end{aligned}$$

By the property of normal random variables, we have that $Z_{1,n-1} + Z_{1,n}$ is independent of $Z_{1,n-1} - Z_{1,n}$, and $Z_{2,n-1} + Z_{2,n}$ is independent of $Z_{2,n-1} - Z_{2,n}$. Therefore, noting that $\sum_{i \notin \{1,2,n-1,n\}} Z_{1i} - \sum_{j \notin \{1,2,n-1,n\}} Z_{2j}$ is independent of $\sum_{i \notin \{1,2,n-1,n\}} Z_{n,i} - \sum_{j \notin \{1,2,n-1,n\}} Z_{n-1,j}$, we conclude that the event $S_1 - S_2 \geq 0$ is independent of $S_n - S_{n-1} \geq 0$. This completes the proof of the proposition. ■

ACKNOWLEDGEMENT

We would like to thank the Editor, the Associate Editor, and three anonymous referees for their helpful comments and suggestions on the previous version of the manuscript, which led to significant improvement of the paper. This work was partially done when R.M. was a PhD candidate in biostatistics at the University of Pennsylvania, supported by Professor Hongzhe Li.

REFERENCES

- [1] I. Liiv, “Seriation and matrix reordering methods: An historical overview,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 3, no. 2, pp. 70–91, 2010.
- [2] M. Behrisch, B. Bach, N. Henry Riche, T. Schreck, and J.-D. Fekete, “Matrix reordering methods for table and network visualization,” in *Computer Graphics Forum*, vol. 35, no. 3. Wiley Online Library, 2016, pp. 693–716.
- [3] H.-M. Wu, S. L. Tzeng, and C.-H. Chen, “Matrix visualization,” in *Handbook of Data Visualization*. Springer, 2008, pp. 681–708.
- [4] C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, and J. L. Rinn, “The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells,” *Nature Biotechnology*, vol. 32, no. 4, pp. 381–386, 2014.
- [5] C. Zeng, F. Mulas, Y. Sui, T. Guan, N. Miller, Y. Tan, F. Liu, W. Jin, A. C. Carrano, M. O. Huising *et al.*, “Pseudotemporal ordering of single cells reveals metabolic control of postnatal β cell proliferation,” *Cell Metabolism*, vol. 25, no. 5, pp. 1160–1175, 2017.
- [6] W. Saelens, R. Cannoodt, H. Todorov, and Y. Saeys, “A comparison of single-cell trajectory inference methods,” *Nature Biotechnology*, vol. 37, no. 5, pp. 547–554, 2019.

- [7] J. Karin, Y. Bornfeld, and M. Nitzan, “scprisma: inference, filtering and enhancement of periodic signals in single-cell data using spectral template matching,” *bioRxiv*, 2022.
- [8] V. Bagaria, J. Ding, D. Tse, Y. Wu, and J. Xu, “Hidden hamiltonian cycle recovery via linear programming,” *Operations Research*, vol. 68, no. 1, pp. 53–70, 2020.
- [9] R. Ma, T. T. Cai, and H. Li, “Optimal estimation of bacterial growth rates based on a permuted monotone matrix,” *Biometrika*, vol. 108, no. 3, pp. 693–708, 2021.
- [10] —, “Optimal permutation recovery in permuted monotone matrix model,” *J. Amer. Statist. Assoc.*, vol. 116, no. 535, pp. 1358–1372, 2021.
- [11] M. Friendly, “Corrgrams: Exploratory displays for correlation matrices,” *The American Statistician*, vol. 56, no. 4, pp. 316–324, 2002.
- [12] D. Murdoch and E. Chow, “A graphical display of large correlation matrices,” *The American Statistician*, vol. 50, no. 2, pp. 178–180, 1996.
- [13] R. Kolde, *pheatmap: Pretty Heatmaps*, 2019, r package version 1.0.12. [Online]. Available: <https://CRAN.R-project.org/package=pheatmap>
- [14] T. Wei and V. Simko, *R package 'corrplot': Visualization of a Correlation Matrix*, 2021, (Version 0.90). [Online]. Available: <https://github.com/taiyun/corrplot>
- [15] G. Meurant, “A review on the inverse of symmetric tridiagonal and block tridiagonal matrices,” *SIAM Journal on Matrix Analysis and Applications*, vol. 13, no. 3, pp. 707–728, 1992.
- [16] S. Noschese, L. Pasquini, and L. Reichel, “Tridiagonal toeplitz matrices: properties and novel applications,” *Numerical Linear Algebra with Applications*, vol. 20, no. 2, pp. 302–326, 2013.
- [17] C. M. da Fonseca, V. Kowalenko, and L. Losonczi, “Ninety years of k-tridiagonal matrices,” *Studia Scientiarum Mathematicarum Hungarica*, vol. 57, no. 3, pp. 298–311, 2020.
- [18] A. Böttcher and S. M. Grudsky, *Spectral properties of banded Toeplitz matrices*. SIAM, 2005.
- [19] F. Bünger, “Inverses, determinants, eigenvalues, and eigenvectors of real symmetric toeplitz matrices with linearly increasing entries,” *Linear Algebra and its Applications*, vol. 459, pp. 595–619, 2014.
- [20] K. S. Berenhaut and D. Bandyopadhyay, “Monotone convex sequences and cholesky decomposition of symmetric toeplitz matrices,” *Linear Algebra and Its Applications*, vol. 403, pp. 75–85, 2005.
- [21] T. T. Cai, Z. Ren, and H. H. Zhou, “Optimal rates of convergence for estimating toeplitz covariance matrices,” *Probability Theory and Related Fields*, vol. 156, no. 1-2, pp. 101–143, 2013.
- [22] S. T. Barnard, A. Pothen, and H. Simon, “A spectral algorithm for envelope reduction of sparse matrices,” *Numerical Linear Algebra with Applications*, vol. 2, no. 4, pp. 317–334, 1995.
- [23] J. E. Atkins, E. G. Boman, and B. Hendrickson, “A spectral algorithm for seriation and the consecutive ones

- problem,” *SIAM Journal on Computing*, vol. 28, no. 1, pp. 297–310, 1998.
- [24] F. Fogel, A. d’Aspremont, and M. Vojnovic, “Serialrank: Spectral ranking using seriation,” *Advances in Neural Information Processing Systems*, vol. 27, pp. 900–908, 2014.
- [25] M. Hahsler, C. Buchta, and K. Hornik, *seriation: Infrastructure for Ordering Objects Using Seriation*, 2021, r package version 1.3.1. [Online]. Available: <https://CRAN.R-project.org/package=seriation>
- [26] D. R. Fulkerson and O. A. Gross, “Incidence matrices with the consecutive 1’s property,” *Bulletin of the American Mathematical Society*, vol. 70, no. 5, pp. 681–684, 1964.
- [27] D. G. Kendall, “A statistical approach to flinders petries sequence-dating,” *Bulletin of the International Statistical Institute*, vol. 40, no. 2, pp. 657–681, 1963.
- [28] —, “A mathematical approach to seriation,” *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 269, no. 1193, pp. 125–134, 1970.
- [29] O. Collier and A. S. Dalalyan, “Minimax rates in permutation estimation for feature matching,” *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 162–192, 2016.
- [30] M. Jeong, A. Dytso, M. Cardone, and H. V. Poor, “Recovering data permutations from noisy observations: The linear regime,” *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 3, pp. 854–869, 2020.
- [31] T. Galstyan, A. Minasyan, and A. Dalalyan, “Optimal detection of the feature matching map in presence of noise and outliers,” *arXiv preprint arXiv:2106.07044*, 2021.
- [32] M. Braverman and E. Mossel, “Sorting from noisy information,” *arXiv preprint arXiv:0910.1191*, 2009.
- [33] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye, “Statistical ranking and combinatorial hodge theory,” *Mathematical Programming*, vol. 127, no. 1, pp. 203–244, 2011.
- [34] S. Negahban, S. Oh, and D. Shah, “Iterative ranking from pair-wise comparisons,” in *NIPS*, 2012, pp. 2474–2482.
- [35] S. Chatterjee and S. Mukherjee, “On estimation in tournaments and graphs under monotonicity constraints,” *arXiv preprint arXiv:1603.04556*, 2016.
- [36] N. B. Shah, S. Balakrishnan, A. Guntuboyina, and M. J. Wainwright, “Stochastically transitive models for pairwise comparisons: Statistical and computational issues,” *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 934–959, 2016.
- [37] C. Mao, J. Weed, and P. Rigollet, “Minimax rates and efficient algorithms for noisy sorting,” *arXiv preprint arXiv:1710.10388*, 2017.
- [38] S. Negahban, S. Oh, K. K. Thekumparampil, and J. Xu, “Learning from comparisons and choices,” *J. Mach. Learn. Res.*, vol. 19, no. 40, 2018.
- [39] Y. Chen, J. Fan, C. Ma, and K. Wang, “Spectral method and regularized mle are both optimal for top-k

- ranking,” *Ann. Statist.*, vol. 47, no. 4, p. 2204, 2019.
- [40] N. Flammarion, C. Mao, and P. Rigollet, “Optimal rates of statistical seriation,” *Bernoulli*, vol. 25, no. 1, pp. 623–653, 2019.
- [41] C. Mao, A. Pananjady, and M. J. Wainwright, “Towards optimal estimation of bivariate isotonic matrices with unknown permutations,” *Ann. Statist.*, vol. 48, no. 6, pp. 3183–3205, 2020.
- [42] A. Pananjady and R. J. Samworth, “Isotonic regression with unknown permutations: Statistics, computation, and adaptation,” *arXiv preprint arXiv:2009.02609*, 2020.
- [43] A. Recanati, T. Kerdreux, and A. d’Aspremont, “Reconstructing latent orderings by spectral clustering,” *arXiv preprint arXiv:1807.07122*, 2018.
- [44] S. Armstrong, C. Guzmán, and C. A. Sing-Long, “An optimal algorithm for strict circular seriation,” *arXiv preprint arXiv:2106.05944*, 2021.
- [45] J.-C. Hütter, C. Mao, P. Rigollet, and E. Robeva, “Estimation of monge matrices,” *Bernoulli*, vol. 26, no. 4, pp. 3051–3080, 2020.
- [46] C. Giraud, Y. Issartel, and N. Verzelen, “Localization in 1d non-parametric latent space models from pairwise affinities,” *arXiv preprint arXiv:2108.03098*, 2021.
- [47] Y. Issartel, C. Giraud, and N. Verzelen, “Optimal embedding on the sphere in non-parametric latent space models,” 2021.
- [48] J. Ding, Y. Wu, J. Xu, and D. Yang, “Consistent recovery threshold of hidden nearest neighbor graphs,” *IEEE Transactions on Information Theory*, vol. 67, no. 8, pp. 5211–5229, 2021.
- [49] M. Deza, “Matrices dont deux lignes quelconque coïncident dans un nombre donné de positions communes,” *Journal of Combinatorial Theory, Series A*, vol. 20, no. 3, pp. 306–318, 1976.
- [50] P. J. Cameron and I. M. Wanless, “Covering radius for sets of permutations,” *Discrete Mathematics*, vol. 293, no. 1-3, pp. 91–109, 2005.
- [51] J. Quistorff, “A survey on packing and covering problems in the hamming permutation space,” *The Electronic Journal of Combinatorics*, vol. 13, no. 1, p. A1, 2006.
- [52] K. Hendrey and I. M. Wanless, “Covering radius in the hamming permutation space,” *European Journal of Combinatorics*, vol. 84, p. 103025, 2020.
- [53] A. Recanati, T. Brüls, and A. d’Aspremont, “A spectral algorithm for fast de novo layout of uncorrected long nanopore reads,” *Bioinformatics*, vol. 33, no. 20, pp. 3188–3194, 2017.
- [54] A. Y. Lu, “Sparse principal component analysis for functional data,” Ph.D. dissertation, 2002, stanford University. [Online]. Available: <https://proxy.library.upenn.edu/login?url=https://www.proquest.com/dissertations-Btheses/sparse-Bprincipal-Bcomponent-Banalysis-Bfunctional/docview/>

[305571180/se-B2?accountid=14707](https://doi.org/10.1007/978-1-4939-9832-7_305571180/se-B2?accountid=14707)

- [55] I. M. Johnstone and A. Y. Lu, “Sparse principal components analysis,” *arXiv preprint arXiv:0901.4392*, 2009.
- [56] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin, “On the principal components of sample covariance matrices,” *Probability Theory and Related Fields*, vol. 164, no. 1-2, pp. 459–552, 2016.
- [57] Z. Bao, X. Ding, and K. Wang, “Singular vector and singular subspace distribution for the matrix denoising model,” *Ann. Statist.*, vol. 49, no. 1, pp. 370–392, 2021.
- [58] T. T. Cai, H. Li, and R. Ma, “Optimal structured principal subspace estimation: Metric entropy and minimax rates,” *J. Mach. Learn. Res.*, vol. 22, no. 46, pp. 1–45, 2021.
- [59] Z. Ma and Y. Wu, “Computational barriers in minimax submatrix detection,” *Ann. Statist.*, vol. 43, no. 3, pp. 1089–1116, 2015.
- [60] T. T. Cai and Y. Wu, “Statistical and computational limits for sparse matrix detection,” *Ann. Statist.*, vol. 48, no. 3, pp. 1593–1614, 2020.
- [61] T. T. Cai, Z. Ma, and Y. Wu, “Sparse PCA: Optimal rates and adaptive estimation,” *Ann. Statist.*, vol. 41, no. 6, pp. 3074–3110, 2013.
- [62] T. Wang, Q. Berthet, and R. J. Samworth, “Statistical and computational trade-offs in estimation of sparse principal components,” *Ann. Statist.*, vol. 44, no. 5, pp. 1896–1930, 2016.
- [63] Y. Luo and A. R. Zhang, “Tensor clustering with planted structures: Statistical optimality and computational limits,” *arXiv preprint arXiv:2005.10743*, 2020.
- [64] B. Rajaratnam and J. Salzman, “Best permutation analysis,” *Journal of Multivariate Analysis*, vol. 121, pp. 193–223, 2013.
- [65] C.-H. Chen, “Generalized association plots: Information visualization via iteratively generated correlation matrices,” *Statistica Sinica*, pp. 7–29, 2002.
- [66] J. C. Bezdek and R. J. Hathaway, “Vat: A tool for visual assessment of (cluster) tendency,” in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN’02 (Cat. No. 02CH37290)*, vol. 3. IEEE, 2002, pp. 2225–2230.
- [67] F. Guo, L. Yan, H. Guo, L. Li, B. Hu, Y. Zhao, J. Yong, Y. Hu, X. Wang, Y. Wei *et al.*, “The transcriptome and dna methylome landscapes of human primordial germ cells,” *Cell*, vol. 161, no. 6, pp. 1437–1452, 2015.
- [68] S. Chatterjee, “A new perspective on least squares under convex constraint,” *Ann. Statist.*, vol. 42, no. 6, pp. 2340–2381, 2014.
- [69] A. B. Tsybakov, *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009.
- [70] Y. Nakatsukasa, N. Saito, and E. Woei, “Mysteries around the graph laplacian eigenvalue 4,” *Linear Algebra*

and its Applications, vol. 438, no. 8, pp. 3231–3246, 2013.

- [71] R. Vershynin, *High-dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018, vol. 47.
- [72] S. Szarek, “Metric entropy of homogeneous spaces,” *Banach Center Publications*, vol. 43, no. 1, pp. 395–410, 1998.
- [73] M. Rudelson and R. Vershynin, “Hanson-wright inequality and sub-gaussian concentration,” *Electronic Communications in Probability*, vol. 18, 2013.
- [74] Z.-D. Bai and Y.-Q. Yin, “Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix,” *The Annals of Probability*, pp. 1729–1741, 1988.
- [75] R. Vershynin, “Introduction to the non-asymptotic analysis of random matrices,” *arXiv preprint arXiv:1011.3027*, 2010.