# Detection and Recovery of Hidden Submatrices

Marom Dadon, Wasim Huleihel, *Member, IEEE*, and Tamir Bendory, *Senior Member, IEEE*

***Abstract*—In this paper, we study the problems of detection and recovery of hidden submatrices with elevated means inside a large Gaussian random matrix. We consider two different structures for the planted submatrices. In the first model, the planted matrices are disjoint, and their row and column indices can be arbitrary. Inspired by scientific applications, the second model restricts the row and column indices to be consecutive. In the detection problem, under the null hypothesis, the observed matrix is a realization of independent and identically distributed standard normal entries. Under the alternative, there exists a set of hidden submatrices with elevated means inside the same standard normal matrix. Recovery refers to the task of locating the hidden submatrices. For both problems, and for both models, we characterize the statistical and computational barriers by deriving information-theoretic lower bounds, designing and analyzing algorithms matching those bounds, and proving computational lower bounds based on the low-degree polynomials conjecture. In particular, we show that the space of the model parameters (i.e., number of planted submatrices, their dimensions, and elevated mean) can be partitioned into three regions: the *impossible* regime, where all algorithms fail; the *hard* regime, where while detection or recovery are statistically possible, we give some evidence that polynomial-time algorithm do not exist; and finally the *easy* regime, where polynomial-time algorithms exist.**

***Index Terms*—Hidden structures, random matrices, statistical and computational limits, statistical inference.**

## I. Introduction

THIS paper studies the detection and recovery problems of hidden submatrices inside a large Gaussian random matrix. In the *detection problem*, under the null hypothesis, the observed matrix is a realization of an independent and identically distributed random matrix with standard normal entries. Under the alternative, there exists a set of hidden submatrices with elevated means inside the same standard normal matrix. Our task is to design a statistical test (i.e., an algorithm) to decide which hypothesis is correct. The *recovery task* is the problem of locating the hidden submatrices. In this case, the devised algorithm estimates the location of the submatrices.

We consider two statistical models for the planted submatrices. In the first model, the planted matrices are disjoint, and their row and column indices can be arbitrary. The detection and recovery variants of this model are well-known as the *submatrix detection* and *submatrix recovery* (or localization) problems, respectively, and received significant attention in the last few years, e.g., [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [15], [16], [17], and references therein. Specifically, for the case of a *single* planted submatrix, the task is to detect the presence of a small $k \times k$ submatrix with entries sampled from a distribution $\mathcal{P}$ in an $n \times n$ matrix of samples from a distribution $\mathcal{Q}$. In the special case where $\mathcal{P}$ and $\mathcal{Q}$ are Gaussians, the statistical and computational barriers, i.e., information-theoretic lower bounds, algorithms, and computational lower bounds, were studied in great detail and were characterized in [1], [2], [3], [4], [7], [9], [17]. When $\mathcal{P}$ and $\mathcal{Q}$ are Bernoulli random variables, the detection task is well-known as the planted dense subgraph problem, which has also been studied extensively in the literature, e.g., [4], [5], [6], [8], [16]. Most notably, for both the Gaussian and Bernoulli problems, it is well understood by now that there appears to be a statistical-computational gap between the minimum value of $k$ at which detection can be solved, and the minimum value of $k$ at which detection can be solved in polynomial time (i.e., with an efficient algorithm). The statistical and computational barriers to the recovery problem have also received significant attention in the literature, e.g., [13], [14], [16], [18], [19], [20], [21], covering several types of distributions, as well as single and (non-overlapping) multiple planted submatrices.

The submatrix model above, where the planted column and row indices are arbitrary, might be less realistic in certain scientific and engineering applications. Accordingly, we also analyze a second model that restricts the row and column indices to be consecutive. One important motivation for this model stems from single-particle cryo-electron microscopy (cryo-EM): a leading technology to elucidate the three-dimensional atomic structure of macromolecules, such as proteins [22], [23]. At the beginning of the algorithmic pipeline of cryo-EM, it is required to locate multiple particle images (tomographic projections of randomly oriented copies of the sought molecular structure) in a highly noisy, large image [24], [25]. This task is dubbed particle picking. While many particle picking algorithms were designed, e.g., [26], [27], [28], [29], this work can be seen as a first attempt to unveil the statistical and computational properties of this task that were not analyzed heretofore.

*a) Main contributions:* To present our results, let us introduce a few notations. In our models, we have $m$ disjoint $k \times k$ submatrices planted in an $n \times n$ matrix. We denote the observed matrix by $\mathsf{X}$. We consider the Gaussian setting, where the entries of the planted submatrices are independent Gaussian random

Manuscript received 30 October 2023; accepted 21 December 2023. Date of publication 10 January 2024; date of current version 22 January 2024. The work of Wasim Huleihel was supported in part by the IISF under Grant 1734/21. The work of Tamir Bendory was supported in part by the NSF-BSF under Grant 2019752, in part by the BSF under Grant 2020159, and in part by the ISF under Grant 1924/21. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shunsuke Ono. *(Corresponding author: Wasim Huliehel.)*

The authors are with the Department of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv 6997801, Israel (e-mail: marom.dadon@gmail.com; wasimh@tauex.tau.ac.il; bendory@tauex.tau.ac.il).

Digital Object Identifier 10.1109/TSIPN.2024.3352264

TABLE I
STATISTICAL AND COMPUTATIONAL THRESHOLDS FOR SUBMATRIX DETECTION (SD), SUBMATRIX RECOVERY (SR), CONSECUTIVE SUBMATRIX DETECTION (CSD), AND CONSECUTIVE SUBMATRIX RECOVERY (CSR), UP TO POLY-LOG FACTORS

| **Type** | **Impossible** | **Hard** | **Easy** |
|---|---|---|---|
| SD | $\lambda \ll \frac{n}{mk^2} \wedge \frac{1}{\sqrt{k}}$ | $\frac{n}{mk^2} \wedge \frac{1}{\sqrt{k}} \ll \lambda \ll 1 \wedge \frac{n}{mk^2}$ | $\lambda \gg 1 \wedge \frac{n}{mk^2}$ |
| SR | $\lambda \ll \frac{1}{\sqrt{k}}$ | $\frac{1}{\sqrt{k}} \ll \lambda \ll 1 \wedge \frac{\sqrt{n}}{k}$ | $\lambda \gg 1 \wedge \frac{\sqrt{n}}{k}$ |
| CSD | $\lambda \ll \frac{1}{k}$ | NO | $\lambda \gg \frac{1}{k}$ |
| CSR | $\lambda \ll \frac{1}{\sqrt{k}}$ | NO | $\lambda \gg \frac{1}{\sqrt{k}}$ |

The bounds in the first row for the special case of $m = 1$ and the second row, are known in the literature (e.g., [4], [9], [13], [14]).

variables with mean $\lambda > 0$ and unit variance, while the entries of the other entries in $\mathsf{X}$ are independent standard normal random variables. This falls under the general "signal+noise" model, in the sense that $\mathsf{X} = \lambda \cdot \mathsf{S} + \mathsf{Z}$, with $\mathsf{S}$ being the signal of interest, $\mathsf{Z}$ is a standard normal noise matrix, and $\lambda$ parameterize the signal-to-noise ratio (SNR) of the problem. We consider two models for $\mathsf{S}$; in the first, the placement of the $m$ planted submatrices is arbitrary, while in the second each of the $m$ planted submatrices have consecutive row and column indices. We will refer to the detection/recovery of the former model as *submatrix detection/recovery*, and for the later as *consecutive submatrix detection/recovery*.

The submatrix detection and recovery problems received significant attention in the literature. The recovery task was analyzed in [13], [14], for any number $m \geq 1$ of planted submatrices, while the detection task [4], [9] was analyzed for $m = 1$ only; our contribution to this literature is the analysis of the detection task of any (possibly growing) number of planted submatrices. Our consecutive model is completely new. In current literature, the elements of the structure/submatrix are typically unconstrained, while in our consecutive model, they must appear in a specific form. This changes both the statistical and computational aspects of the inference problems.

We now discuss our contributions in more detail. We study the computational and statistical boundaries and derive information-theoretic lower bounds, algorithmic upper bounds, and computational lower bounds. In particular, we show that the space of the model parameters $(k, m, \lambda)$ can be partitioned into different disjoint regions: the *impossible* regime, where all algorithms fail; the *hard* regime, where while detection or recovery are statistically possible, we give some evidence that polynomial-time algorithms do not exist; and finally the *easy* regime, where polynomial-time algorithms exist. Table I summarizes the statistical and computational thresholds for the detection and recovery problems discussed above. Note that the bounds in the second row of Table I (submatrix recovery), as well as the first row (submatrix detection) for $m = 1$, are known results as mentioned above.

Interestingly, while it is well-known that the number of planted submatrices $m$ does not play any significant role in the statistical and computational barriers in the submatrix recovery problem, it can be seen that this is not the case for the submatrix detection problem. Similarly to the submatrix recovery problem, the submatrix detection problem undergoes a statistical-computational gap. To provide evidence for this phenomenon, we follow a recent line of work [30], [31], [32], [33], [34], and show that the class of low-degree polynomials fail to solve the detection problem in this conjecturally hard regime. Furthermore, it can be seen that the consecutive submatrix detection and recovery problems are either impossible or easy to solve, namely, there is no hard regime. Here, for both the detection and recovery problems, the number of planted submatrices $m$ does not play an inherent role. We note that there is a statistical gap between consecutive detection and recovery; the former is statistically easier. This is true as long as exact recovery is the performance criterion. We also analyze the correlated recovery (also known as weak recovery) criterion, where recovery is successful if only a fraction of planted entries are recovered. For this weaker criterion, we show that recovery and detection are asymptotically equivalent.

*b) Notation:* Given a distribution $\mathbb{P}$, let $\mathbb{P}^{\otimes n}$ denote the distribution of the $n$-dimensional random vector $(X_1, X_2, \ldots, X_n)$, where the $X_i$ are i.i.d. according to $\mathbb{P}$. Similarly, $\mathbb{P}^{\otimes m \times n}$ denotes the distribution on $\mathbb{R}^{m \times n}$ with i.i.d. entries distributed as $\mathbb{P}$. Given a measurable set $\mathcal{X}$, let $\mathsf{Unif}[\mathcal{X}]$ denote the uniform distribution on $\mathcal{X}$. The relation $X \perp\!\!\!\perp Y$ means that the random variables $X$ and $Y$ are statistically independent. The Hadamard and inner product between two $n \times n$ matrices $\mathsf{A}$ and $\mathsf{B}$ are denoted, respectively, by $\mathsf{A} \odot \mathsf{B}$ and $\langle \mathsf{A}, \mathsf{B} \rangle = \mathsf{trace}(\mathsf{A}^T \mathsf{B})$. For $x \in \mathbb{R}$, we define $[x]_+ = \max(x, 0)$. The nuclear norm of a symmetric matrix $\mathsf{A}$ is denoted by $\|\mathsf{A}\|_\star$, and equals the summation of the absolute values of the eigenvalues of $\mathsf{A}$. We let $\mathbf{I}$ and $\mathbf{J}$ denote the identity and all-one matrices, respectively.

Let $\mathcal{N}(\mu, \sigma^2)$ denote a normal random variable with mean $\mu$ and variance $\sigma^2$, when $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{\geq 0}$. Let $\mathcal{N}(\mu, \Sigma)$ denote a multivariate normal random vector with mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma$, where $\Sigma$ is a $d \times d$ positive semidefinite matrix. Let $\Phi$ denote the cumulative distribution of a standard normal random variable with $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$. For probability measures $\mathbb{P}$ and $\mathbb{Q}$, let $d_{\mathsf{TV}}(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} \int |\mathrm{d}\mathbb{P} - \mathrm{d}\mathbb{Q}|$, $\chi^2(\mathbb{P}||\mathbb{Q}) = \int \frac{(\mathrm{d}\mathbb{P} - \mathrm{d}\mathbb{Q})^2}{\mathrm{d}\mathbb{Q}}$, and $d_{\mathsf{KL}}(\mathbb{P}||\mathbb{Q}) = \mathbb{E}_{\mathbb{P}} \log \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$, denote the total variation distance, the $\chi^2$-divergence, and the Kullback-Leibler (KL) divergence, respectively. Let $\mathsf{Bern}(p)$ and $\mathsf{Binomial}(n, p)$ denote the Bernoulli and Binomial distributions with parameters $p$ and $n$, respectively. We denote by $\mathsf{Hypergeometric}(n, k, m)$ the Hypergeometric distribution with parameters $(n, k, m)$.

We use standard asymptotic notation. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ if $a_n \leq C b_n$, for some absolute constant $C$ and for all $n$; $a_n = \Omega(b_n)$, if $b_n = O(a_n)$; $a_n = \Theta(b_n)$, if $a_n = O(b_n)$ and $a_n = \Omega(b_n)$,

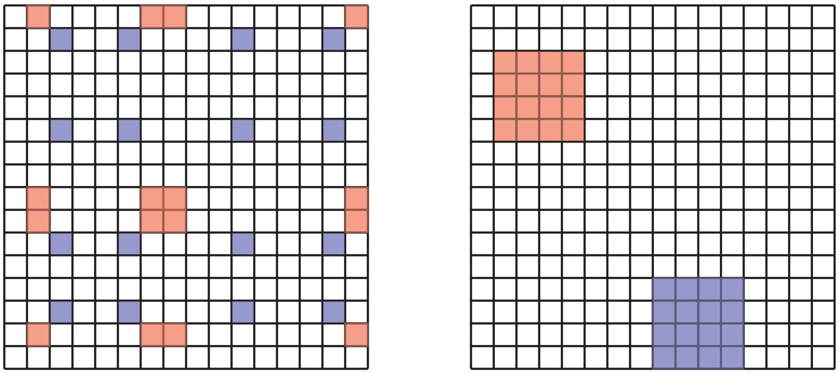


Fig. 1. Illustration of the models considered in this paper: $\mathcal{K}_{k,m,n}$ of Definition 1 (left) and $\mathcal{K}^{\mathsf{con}}_{k,m,n}$ of Definition 2 (right), for $k = 4$, $m = 2$, and $n = 16$.

$a_n = o(b_n)$ or $b_n = \omega(a_n)$, if $a_n/b_n \to 0$, as $n \to \infty$. Finally, for $a, b \in \mathbb{R}$, we let $a \vee b \triangleq \max\{a, b\}$ and $a \wedge b \triangleq \min\{a, b\}$. Throughout the paper, $C$ refers to any constant independent of the parameters of the problem at hand and will be reused for different constants. The notation $\ll$ refers to polynomially less than in $n$, namely, $a_n \ll b_n$ if $\liminf_{n\to\infty} \log_n a_n < \liminf_{n\to\infty} \log_n b_n$, e.g., $n \ll n^2$, but $n \not\ll n \log_2 n$. For $n \in \mathbb{N}$, we let $[n] = \{1, 2, \ldots, n\}$. For a subset $S \subseteq \mathbb{R}$, we let $\mathbb{1}\{S\}$ denote the indicator function of the set $S$.

*c) Paper organization:* The rest of the paper is organized as follows. In Section II, we formulate the submatrix and consecutive submatrix detection and recovery problems. Our main results are presented in Section III; for both the detection and recovery problems we derive asymptotically tight statistical and computational lower and upper bounds. The proofs of some of our main results appear in Section IV; due to page limitation we relegate leftover proofs to an auxiliary file [35]. Finally, our conclusions and outlook appear in Section V.

## II. Problem Formulation

In this section, we present our model and define the detection and recovery problems we investigate, starting with the former. For simplicity of notations, we denote $\mathcal{Q} = \mathcal{N}(0, 1)$ and $\mathcal{P} = \mathcal{N}(\lambda, 1)$, for some $\lambda > 0$, which can be interpreted as the signal-to-noise ratio (SNR) parameter of the underlying model.

### A. The Detection Problem

Let $(m, k, n)$ be three natural numbers, satisfying $m \cdot k \leq n$. We emphasize that the values of $m$, $k$, and $\lambda$, are allowed to be functions of $n$—the dimension of the observation. Let $\mathcal{K}_{k,m,n}$ denote all possible sets that can be represented as a union of $m$ disjoint subsets of $[n]$, each of size $k$; see Fig. 1 for an illustration. Formally,

$$\mathcal{K}_{k,m,n} \triangleq \Bigg\{ \mathsf{K}_{k,m} = \bigcup_{i=1}^{m} \mathsf{S}_i \times \mathsf{T}_i : \ \mathsf{S}_i, \mathsf{T}_i \subset \mathcal{C}_k, \ \forall i \in [m], \\ (\mathsf{S}_i \times \mathsf{T}_i) \cap (\mathsf{S}_j \times \mathsf{T}_j) = \emptyset, \ \forall i \neq j \in [m] \Bigg\}, \quad (1)$$

where $\mathcal{C}_k \triangleq \{\mathsf{S} \subset [n] : \ |\mathsf{S}| = k\}$, namely, it is the set of all subsets of $[n]$ of size $k$. We next formulate two different detection problems that we wish to investigate, starting with the following one, a generalization of the Gaussian planted clique problem (or, bi-clustering, see, e.g., [9]) to multiple hidden submatrices (or, clusters).

*Definition 1 (Submatrix Detection):* Let $(\mathcal{P}, \mathcal{Q})$ be a pair of distributions over a measurable space $(\mathbb{R}, \mathcal{B})$. Let $\mathsf{SD}(n, k, m, \mathcal{P}, \mathcal{Q})$ denote the hypothesis testing problem with observation $\mathsf{X} \in \mathbb{R}^{n\times n}$ and hypotheses

$$\mathcal{H}_0 : \mathsf{X} \sim \mathcal{Q}^{\otimes n\times n} \quad \text{vs.} \quad \mathcal{H}_1 : \mathsf{X} \sim \mathcal{D}(n, k, m, \mathcal{P}, \mathcal{Q}), \quad (2)$$

where $\mathcal{D}(n, k, m, \mathcal{P}, \mathcal{Q})$ is the distribution of matrices $\mathsf{X}$ with entries $\mathsf{X}_{ij} \sim \mathcal{P}$ if $i, j \in \mathsf{K}_{k,m}$ and $\mathsf{X}_{ij} \sim \mathcal{Q}$ otherwise that are conditionally independent given $\mathsf{K}_{k,m}$, which is chosen uniformly at random over all subsets of $\mathcal{K}_{k,m,n}$.

To wit, under $\mathcal{H}_0$ the elements of $\mathsf{X}$ are all distributed i.i.d. according to $\mathcal{Q}$, while under $\mathcal{H}_1$, there are $m$ planted disjoint submatrices $\mathsf{K}_{k,m}$ in $\mathsf{X}$ with entries distributed according to $\mathcal{P}$, and the other entries (outside of $\mathsf{K}_{k,m}$) are distributed according to $\mathcal{Q}$.

Note that the columns and row indices of the planted submatrices in (1) can appear everywhere; in particular, they are not necessarily consecutive. In some applications, however, we would like those submatrices to be defined by a set of consecutive rows and a set of consecutive columns (e.g., when those submatrices model images like in cryo-EM). Accordingly, we consider the following set:

$$\mathcal{K}^{\mathsf{con}}_{k,m,n} \triangleq \Bigg\{ \mathsf{K}_{k,m} = \bigcup_{i=1}^{m} \mathsf{S}_i \times \mathsf{T}_i : \ \mathsf{S}_i, \mathsf{T}_i \subset \mathcal{C}^{\mathsf{con}}_k, \ \forall i \in [m], \\ (\mathsf{S}_i \times \mathsf{T}_i) \cap (\mathsf{S}_j \times \mathsf{T}_j) = \emptyset, \ \forall i \neq j \in [m] \Bigg\}, \quad (3)$$

where $\mathcal{C}^{\mathsf{con}}_k \triangleq \{\mathsf{S} \subset [n] : \ |\mathsf{S}| = k, \ \mathsf{S} \text{ is consecutive}\}$, namely, it is the set of all subsets of $[n]$ of size $k$ with consecutive elements. For example, for $n = 4$, we have $\mathcal{C}^{\mathsf{con}}_3 = \{1, 2, 3\} \cup \{2, 3, 4\}$. The difference between $\mathcal{K}_{k,m,n}$ and $\mathcal{K}^{\mathsf{con}}_{k,m,n}$ is depicted in Fig. 1; it is evident that the submatrices in $\mathcal{K}_{k,m,n}$ can appear everywhere, while those in $\mathcal{K}^{\mathsf{con}}_{k,m,n}$ are consecutive. Consider the following detection problem.

*Definition 2 (Consecutive Submatrix Detection):* Let $(\mathcal{P}, \mathcal{Q})$ be a pair of distributions over a measurable space $(\mathbb{R}, \mathcal{B})$. Let $\mathsf{CSD}(n, k, m, \mathcal{P}, \mathcal{Q})$ denote the hypothesis testing problem with observation $\mathsf{X} \in \mathbb{R}^{n\times n}$ and hypotheses

$$\mathcal{H}_0 : \mathsf{X} \sim \mathcal{Q}^{\otimes n\times n} \quad \text{vs.} \quad \mathcal{H}_1 : \mathsf{X} \sim \widetilde{\mathcal{D}}(n,k,m,\mathcal{P},\mathcal{Q}), \tag{4}$$

where $\widetilde{\mathcal{D}}(n,k,m,\mathcal{P},\mathcal{Q})$ is the distribution of matrices $\mathsf{X}$ with entries $\mathsf{X}_{ij} \sim \mathcal{P}$ if $i,j \in \mathsf{K}_{k,m}$ and $\mathsf{X}_{ij} \sim \mathcal{Q}$ otherwise that are conditionally independent given $\mathsf{K}_{k,m}$, which is chosen uniformly at random over all subsets of $\mathcal{K}^{\mathsf{con}}_{k,m,n}$.

Observing $\mathsf{X}$, a detection algorithm $\mathcal{A}_n$ for the problems above is tasked with outputting a decision in $\{0,1\}$. We define the *risk* of a detection algorithm $\mathcal{A}_n$ as the sum of its Type-I and Type-II errors probabilities, namely,

$$\mathsf{R}(\mathcal{A}_n) = \mathbb{P}_{\mathcal{H}_0}(\mathcal{A}_n(\mathsf{X}) = 1) + \mathbb{P}_{\mathcal{H}_1}(\mathcal{A}_n(\mathsf{X}) = 0), \tag{5}$$

where $\mathbb{P}_{\mathcal{H}_0}$ and $\mathbb{P}_{\mathcal{H}_1}$ denote the probability distributions under the null hypothesis and the alternative hypothesis, respectively. If $\mathsf{R}(\mathcal{A}_n) \to 0$ as $n \to \infty$, then we say that $\mathcal{A}_n$ solves the detection problem. The algorithms we consider here are either unconstrained (and thus might be computationally expensive) or run in polynomial time (computationally efficient). Typically, unconstrained algorithms are considered in order to show that information-theoretic lower bounds are asymptotically tight. An algorithm that runs in polynomial time must run in $\mathsf{poly}(n)$ time, where $n$ is the size of the input. As mentioned in the introduction, our goal is to derive necessary and sufficient conditions for when it is impossible and possible to detect the underlying submatrices, with and without computational constraints, for both the SD and CSD models.

### B. The Recovery Problem

Next, we consider the recovery variant of the problem in Definition 2. Note that the submatrix recovery problem that corresponds to the problem in Definition 1, where the entries of the submatrices are not necessarily consecutive, was investigated in [13]. In the recovery problem, we assume that the data follow the distribution under $\mathcal{H}_1$ in Definition 2, and the inference task is to recover the location of the planted submatrices. This is the analog of the particle picking problem in cryo-EM that was introduced in Section I. Consider the following definition.

*Definition 3 (Consecutive Submatrix Recovery):* Let $(\mathcal{P},\mathcal{Q})$ be a pair of distributions over a measurable space $(\mathbb{R},\mathcal{B})$. Assume that $\mathsf{X} \in \mathbb{R}^{n\times n} \sim \widetilde{\mathcal{D}}(n,k,m,\mathcal{P},\mathcal{Q})$, where $\widetilde{\mathcal{D}}(n,k,m,\mathcal{P},\mathcal{Q})$ is the distribution of matrices $\mathsf{X}$ with entries $\mathsf{X}_{ij} \sim \mathcal{P}$ if $i,j \in \mathsf{K}^\star$ and $\mathsf{X}_{ij} \sim \mathcal{Q}$ otherwise that are conditionally independent given $\mathsf{K}^\star \in \mathcal{K}^{\mathsf{con}}_{k,m,n}$. The goal is to recover the hidden submatrices $\mathsf{K}^\star$, up to a permutation of the submatrices indices, given the matrix $\mathsf{X}$. We let $\mathsf{CSR}(n,k,m,\mathcal{P},\mathcal{Q})$ denote this recovery problem.

Several metrics of reconstruction accuracy are possible, and we will focus on two: *exact* and *correlated* recovery criteria. Our estimation procedures produce a set $\hat{\mathsf{K}} = \hat{\mathsf{K}}(\mathsf{X})$ aimed to estimate at best the underlying true submatrices $\mathsf{K}^\star$. Consider the following definitions.

*Definition 4 (Exact Recovery):* We say that $\hat{\mathsf{K}}$ achieves exact recovery of $\mathsf{K}^\star$, if, as $n \to \infty$, $\sup_{\mathsf{K}^\star \in \mathcal{K}^{\mathsf{con}}_{k,m,n}} \mathbb{P}(\hat{\mathsf{K}} \neq \mathsf{K}^\star) \to 0$.

*Definition 5 (Correlated Recovery):* The overlap of $\mathsf{K}^\star$ and $\hat{\mathsf{K}}$ is defined as the expected size of their intersection, i.e.,

$$\mathsf{overlap}(\mathsf{K}^\star, \hat{\mathsf{K}}) \triangleq \mathbb{E}\langle \mathsf{K}^\star, \hat{\mathsf{K}}\rangle = \sum_{i=1}^{n} \mathbb{P}(i \in \mathsf{K}^\star \cap \hat{\mathsf{K}}). \tag{6}$$

We say that $\hat{\mathsf{K}}$ achieves correlated recovery of $\mathsf{K}^\star$ if there exists a fixed constant $\epsilon > 0$, such that $\lim_{n\to\infty} \sup_{\mathsf{K}^\star \in \mathcal{K}^{\mathsf{con}}_{k,m,n}} \frac{\mathsf{overlap}(\mathsf{K}^\star,\hat{\mathsf{K}})}{mk^2} \geq \epsilon$.

Similarly to the detection problem, also here we will care about both unconstrained and polynomial time algorithms, and we aim to derive necessary and sufficient conditions for when it is impossible and possible to recover the underlying submatrices.

## III. Main Results

In this section, we present our main results for the detection and recovery problems, starting with the former. For both problems, we derive the statistical and computational bounds for the two models we presented in the previous section.

### A. The Detection Problem

*a) Upper bounds:* We start by presenting our upper bounds. To that end, we propose three algorithms and analyze their performance. Define the statistics,

$$\mathsf{T}_{\mathsf{sum}}(\mathsf{X}) \triangleq \sum_{i,j\in[n]} \mathsf{X}_{ij}, \tag{7}$$

$$\mathsf{T}^{\mathsf{SD}}_{\mathsf{scan}}(\mathsf{X}) \triangleq \max_{\mathsf{K}\in\mathcal{K}_{k,1,n}} \sum_{i,j\in\mathsf{K}} \mathsf{X}_{ij}, \tag{8}$$

$$\mathsf{T}^{\mathsf{CSD}}_{\mathsf{scan}}(\mathsf{X}) \triangleq \max_{\mathsf{K}\in\mathcal{K}^{\mathsf{con}}_{k,1,n}} \sum_{i,j\in\mathsf{K}} \mathsf{X}_{ij}. \tag{9}$$

The statistics in (7) amounts to adding up all the elements of $\mathsf{X}$, while (8) and (9) enumerate all $k\times k$ submatrices of $\mathsf{X}$ in $\mathcal{K}_{k,1,n}$ and $\mathcal{K}^{\mathsf{con}}_{k,1,n}$, and take the submatrix with the maximal sum of entries, respectively. Fix $\delta > 0$. Then, our tests are defined as,

$$\mathcal{A}_{\mathsf{sum}}(\mathsf{X}) \triangleq \mathbb{1}\left\{\mathsf{T}_{\mathsf{sum}}(\mathsf{X}) \geq \tau_{\mathsf{sum}}\right\}, \tag{10}$$

$$\mathcal{A}^{\mathsf{SD}}_{\mathsf{scan}}(\mathsf{X}) \triangleq \mathbb{1}\left\{\mathsf{T}^{\mathsf{SD}}_{\mathsf{scan}}(\mathsf{X}) \geq \tau^{\mathsf{SD}}_{\mathsf{scan}}\right\}, \tag{11}$$

$$\mathcal{A}^{\mathsf{CSD}}_{\mathsf{scan}}(\mathsf{X}) \triangleq \mathbb{1}\left\{\mathsf{T}^{\mathsf{CSD}}_{\mathsf{scan}}(\mathsf{X}) \geq \tau^{\mathsf{CSD}}_{\mathsf{scan}}\right\}, \tag{12}$$

where the thresholds are given by $\tau_{\mathsf{sum}} \triangleq \frac{mk^2\lambda}{2}$, $\tau^{\mathsf{SD}}_{\mathsf{scan}} \triangleq \sqrt{(4+\delta)k^2 \log\binom{n}{k}}$, and $\tau^{\mathsf{CSD}}_{\mathsf{scan}} \triangleq \sqrt{(4+\delta)k^2\log n}$, and correspond roughly to the average between the expected values of each of the statistics in (7)–(9) under the null and alternative hypotheses. It should be emphasized that the tests in (10)–(11) were proposed in, e.g., [2], [4], [9], for the special case of a single planted submatrix detection problem ($m = 1$).

A few important remarks are in order. First, note that in the scan test, we search for a single planted matrix rather than $m$ such matrices. Second, the sum test exhibits polynomial computational complexity, of $O(n^2)$ operations, and hence efficient. The scan test in (11), however, exhibits an exponential computational complexity, and thus is inefficient. Indeed, the search space in (11) is of cardinality $|\mathcal{K}_{k,1,n}| = \binom{n}{k}^2$; we will discuss this in detail later on (see, paragraph (c) below). On the other hand, the scan test $\mathcal{A}_{\mathsf{scan}}^{\mathsf{CSD}}$ for the consecutive setting is efficient because $|\mathcal{K}_{k,1,n}^{\mathsf{con}}| \leq n^2$.

The following result provides sufficient conditions under which the risk of each of the above tests is asymptotically small.

*Theorem 1 (Detection Upper Bounds):* Consider the detection problems in Definitions 1 and 2. Then, we have the following bounds:

1) (Efficient SD) There exists an efficient algorithm $\mathcal{A}_{\mathsf{sum}}$ in (10), such that if

$$\lambda = \omega\left(\frac{n}{mk^2}\right), \tag{13}$$

then $\mathsf{R}(\mathcal{A}_{\mathsf{sum}}) \to 0$, as $n \to \infty$, for the problems in Definitions 1 and 2.

2) (Exhaustive SD) There exists an algorithm $\mathcal{A}_{\mathsf{scan}}^{\mathsf{SD}}$ in (11), such that if

$$\lambda = \omega\left(\sqrt{\frac{\log \frac{n}{k}}{k}}\right), \tag{14}$$

then $\mathsf{R}(\mathcal{A}_{\mathsf{scan}}^{\mathsf{SD}}) \to 0$, as $n \to \infty$, for the problem in Definition 1.

3) (Efficient CSD) There exists an efficient algorithm $\mathcal{A}_{\mathsf{scan}}^{\mathsf{CSD}}$ in (12), such that if

$$\lambda = \omega\left(\frac{\sqrt{\log \frac{n}{k}}}{k}\right), \tag{15}$$

then $\mathsf{R}(\mathcal{A}_{\mathsf{scan}}^{\mathsf{CSD}}) \to 0$, as $n \to \infty$, for the problem in Definition 2.

As can be seen from Theorem 1, only the sum test performance barrier exhibits dependency on $m$. The scan test is, for both SD and CSD, inherently independent of $m$. This makes sense because when summing all the elements of $\mathsf{X}$, as $m$ gets larger the mean (the "signal") under the alternative hypothesis gets larger as well. On the other hand, since the scan test searches for a single planted submatrix, the number of planted submatrices does not play a role. One could argue that it might be beneficial to search for the $m$ planted submatrices in the scan test, however, as we show below, this is not needed, and the bounds above are asymptotically tight.

*b) Lower bounds:* To present our lower bounds, we first recall that the optimal testing error probability is determined by the total variation distance between the distributions under the null and the alternative hypotheses as follows (see, e.g., [36, Lemma 2.1]),

$$\min_{\mathcal{A}_n:\mathbb{R}^{n\times n}\to\{0,1\}} \mathbb{P}_{\mathcal{H}_0}(\mathcal{A}_n(\mathsf{X}) = 1) + \mathbb{P}_{\mathcal{H}_1}(\mathcal{A}_n(\mathsf{X}) = 0)$$
$$= 1 - d_{\mathsf{TV}}(\mathbb{P}_{\mathcal{H}_0}, \mathbb{P}_{\mathcal{H}_1}). \tag{16}$$

The following result shows that under certain conditions the total variation between the null and alternative distributions is asymptotically small, and thus, there exists no test which can solve the above detection problems reliably.

*Theorem 2 (Information-Theoretic Lower Bounds):* We have the following results.

1) Consider the detection problem in Definition 1. If,

$$\lambda = o\left(\frac{n}{mk^2} \wedge \frac{1}{\sqrt{k}}\right), \tag{17}$$

then $d_{\mathsf{TV}}(\mathbb{P}_{\mathcal{H}_0}, \mathbb{P}_{\mathcal{H}_1}) = o(1)$.

2) Consider the detection problem in Definition 2. If $\lambda = o(k^{-1})$, then $d_{\mathsf{TV}}(\mathbb{P}_{\mathcal{H}_0}, \mathbb{P}_{\mathcal{H}_1}) = o(1)$.

Theorem 2 above shows that our upper bounds in Theorem 1 are tight up to poly-log factors. Indeed, item 1 in Theorem 2 complements Items 1-2 in Theorem 1, for the SD problem, while item 2 in Theorem 2 complements Item 3 in Theorem 1, for the CSD problem. In the sequel, we illustrate our results using phase diagrams that show the tradeoff between $k$ and $\lambda$ as a function of $n$. We mention here that our results in Theorems 1 and 2 for submatrix detection coincide with [4], [9], where the special case of $m = 1$ was analyzed.

One evident and important observation here is that the statistical limit for the CSD problem is attained using an efficient test. Thus, there is no statistical computational gap in the detection problem in Definition 2, and accordingly, it is either statistically impossible to solve the detection problem or it can be solved in polynomial time. This is not the case for the SD problem. Note that both the efficient sum and the exhaustive scan tests are needed to attain the information-theoretic lower bound (up to poly-log factors). As discussed above, however, here the scan test is not efficient. We next give evidence that, based on the low-degree polynomial conjecture, efficient algorithms that run in polynomial-time do not exist in the regime where the scan test succeeds while the sum test fails.

*c) Computational lower bounds:* Note that the problem in Definition 1 exhibits a gap in terms of what can be achieved by the proposed polynomial-time algorithm and the computationally expensive scan test algorithm. In particular, it can be seen that in the regime where $\frac{1}{\sqrt{k}} \ll \lambda \ll \frac{n}{mk^2}$, while the problem can be solved by an exhaustive search using the scan test, we do not have a polynomial-time algorithm. Next, we give evidence that, in fact, an efficient algorithm does not exist in this region. To that end, we start with a brief introduction to the method of *low-degree polynomials*.

The premise of this method is to take low-degree multivariate polynomials in the entries of the observations as a proxy for efficiently-computable functions. The ideas below were first developed in a sequence of works in the sum-of-squares optimization literature [30], [31], [37], [38].

In the following, we follow the notations and definitions of [31], [39]. Any distribution $\mathbb{P}_{\mathcal{H}_0}$ on $\Omega_n = \mathbb{R}^{n\times n}$ induces an inner product of measurable functions $f, g : \Omega_n \to \mathbb{R}$ given by $\langle f, g\rangle_{\mathcal{H}_0} = \mathbb{E}_{\mathcal{H}_0}[f(\mathsf{X})g(\mathsf{X})]$, and norm $\|f\|_{\mathcal{H}_0} = \langle f, f\rangle_{\mathcal{H}_0}^{1/2}$. We Let $L^2(\mathbb{P}_{\mathcal{H}_0})$ denote the Hilbert space consisting of functions $f$ for which $\|f\|_{\mathcal{H}_0} < \infty$, endowed with the above inner product and norm. In the computationally-unbounded case, the Neyman-Pearson lemma shows that the likelihood ratio test achieves the optimal tradeoff between Type-I and Type-II error probabilities. Furthermore, it is well-known that the same test optimally distinguishes $\mathbb{P}_{\mathcal{H}_0}$ from $\mathbb{P}_{\mathcal{H}_1}$ in the $L^2$ sense. Specifically, denoting by $\mathsf{L}_n \triangleq \mathbb{P}_{\mathcal{H}_1}/\mathbb{P}_{\mathcal{H}_0}$ the likelihood ratio, the second-moment method for contiguity[1] (see, e.g., [39, Lemma 2]) shows that if $\|\mathsf{L}_n\|^2_{\mathcal{H}_0}$ remains bounded as $n \to \infty$, then $\mathbb{P}_{\mathcal{H}_1}$ is contiguous to $\mathbb{P}_{\mathcal{H}_0}$. This implies that $\mathbb{P}_{\mathcal{H}_1}$ and $\mathbb{P}_{\mathcal{H}_0}$ are statistically indistinguishable, i.e., no test can have both Type-I and Type-II error probabilities tending to zero.

We now describe the low-degree method. The idea is to find the low-degree polynomial that best distinguishes $\mathbb{P}_{\mathcal{H}_0}$ from $\mathbb{P}_{\mathcal{H}_1}$ in the $L^2$ sense. To that end, we let $\mathcal{V}_{n,\leq \mathsf{D}} \subset L^2(\mathbb{P}_{\mathcal{H}_0})$ denote the linear subspace of polynomials $\Omega_n \to \mathbb{R}$ of degree at most $\mathsf{D} \in \mathbb{N}$. We further define $\mathcal{P}_{\leq \mathsf{D}} : L^2(\mathbb{P}_{\mathcal{H}_0}) \to \mathcal{V}_{n,\leq \mathsf{D}}$ the orthogonal projection operator. Then, the $\mathsf{D}$-*low-degree likelihood ratio* $\mathsf{L}_n^{\leq \mathsf{D}}$ is the projection of a function $\mathsf{L}_n$ to the span of coordinate-degree-$\mathsf{D}$ functions, where the projection is orthogonal with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$. As discussed above, the likelihood ratio optimally distinguishes $\mathbb{P}_{\mathcal{H}_0}$ from $\mathbb{P}_{\mathcal{H}_1}$ in the $L^2$ sense. The next lemma shows that over the set of low-degree polynomials, the $\mathsf{D}$-low-degree likelihood ratio have exhibit the same property.

*Lemma 1 (Optimally of $\mathsf{L}_n^{\leq \mathsf{D}}$ [30], [38], [39]):* Consider the following optimization problem:

$$\max \mathbb{E}_{\mathcal{H}_1} f(\mathsf{X}) \quad \text{s.t.} \quad \mathbb{E}_{\mathcal{H}_0} f^2(\mathsf{X}) = 1, \; f \in \mathcal{V}_{n,\leq \mathsf{D}}. \tag{18}$$

Then, the unique solution $f^\star$ for (18) is the $\mathsf{D}$-low degree likelihood ratio $f^\star = \mathsf{L}_n^{\leq \mathsf{D}}/\|\mathsf{L}_n^{\leq \mathsf{D}}\|_{\mathcal{H}_0}$, and the value of the optimization problem is $\|\mathsf{L}_n^{\leq \mathsf{D}}\|_{\mathcal{H}_0}$.

As was mentioned above, in the computationally-unbounded regime, an important property of the likelihood ratio is that if $\|\mathsf{L}_n\|_{\mathcal{H}_0}$ is bounded, then $\mathbb{P}_{\mathcal{H}_0}$ and $\mathbb{P}_{\mathcal{H}_1}$ are statistically indistinguishable. The following conjecture states that a computational analog of this property holds, with $\mathsf{L}_n^{\leq \mathsf{D}}$ playing the role of the likelihood ratio. In fact, it also postulates that polynomials of degree $\approx \log n$ are a proxy for polynomial-time algorithms. The conjecture below is based on [30], [31], [38], and [31, Conj. 2.2.4]. We give an informal statement of this conjecture, which appears in [39, Conj. 1]. For a precise statement, we refer the reader to [31, Conj. 2.2.4] and [39, Sec. 4].

*Conjecture 1 (Low-Degree Conjecture, Informal):* Given a sequence of probability measures $\mathbb{P}_{\mathcal{H}_0}$ and $\mathbb{P}_{\mathcal{H}_1}$, if there exists $\epsilon > 0$ and $\mathsf{D} = \mathsf{D}(n) \geq (\log n)^{1+\epsilon}$, such that $\|\mathsf{L}_n^{\leq \mathsf{D}}\|_{\mathcal{H}_0}$ remains bounded as $n \to \infty$, then there is no polynomial-time algorithm that distinguishes $\mathbb{P}_{\mathcal{H}_0}$ and $\mathbb{P}_{\mathcal{H}_1}$.

In the sequel, we will rely on Conjecture 1 to give evidence for the statistical-computational gap observed for the problem in Definition 1 in the regime where $\frac{1}{\sqrt{k}} \ll \lambda \ll \frac{n}{mk^2}$. At this point we would like to mention [31, Hypothesis 2.1.5], which states a more general form of Conjecture 1 in the sense that it postulates that degree-$\mathsf{D}$ polynomials are a proxy for $n^{O(D)}$-time algorithms. Note that if $\|\mathsf{L}_n^{\leq \mathsf{D}}\|_{\mathcal{H}_0} = O(1)$, then we expect detection in time $\mathsf{T}(n) = e^{\mathsf{D}(n)}$ to be impossible.

*Theorem 3 (Computational Lower Bound):* Consider the detection problem in Definition 1. Then, if $\lambda$ is such that $\frac{1}{\sqrt{k}} \ll \lambda \ll \frac{n}{mk^2}$, then $\|\mathsf{L}_n^{\leq \mathsf{D}}\|_{\mathcal{H}_0} \leq O(1)$, for any $\mathsf{D} = \Omega(\log n)$. On the other hand, if $\lambda$ is such that $\lambda \gg \frac{n}{mk^2}$, then $\|\mathsf{L}_n^{\leq \mathsf{D}}\|_{\mathcal{H}_0} \geq \omega(1)$.

Together with Conjecture 1, Theorem 3 implies that if we take degree-$\log n$ polynomials as a proxy for all efficient algorithms, our calculations predict that an $n^{O(\log n)}$ algorithm does not exist when $\frac{1}{\sqrt{k}} \ll \lambda \ll \frac{n}{mk^2}$. This is summarized in the following corollary.

*Corollary 4:* Consider the detection problem in Definition 1, and assume that Conjecture 1 holds. An $n^{O(\log n)}$ algorithm that achieves strong detection does not exist if $\lambda$ is such that $\frac{1}{\sqrt{k}} \ll \lambda \ll \frac{n}{mk^2}$.

These predictions agree precisely with the previously established statistical-computational tradeoffs in the previous subsections. A more explicit formula for the computational barrier which exhibits dependency on $\mathsf{D}$ and $\lambda$ can be deduced from the proof of Theorem 3; to keep the exposition simple we opted to present the refined result above. For the special case of $m = 1$, the same computational lower bound but based on the planted clique conjecture, was proved in [9], [16], [17].

We note that numerical and theoretical evidence for the existence of computational-statistical gaps were observed in other statistical models that are also inspired by cryo-EM, including heterogeneous multi-reference alignment [40], [41] and sparse multi-reference alignment [42].

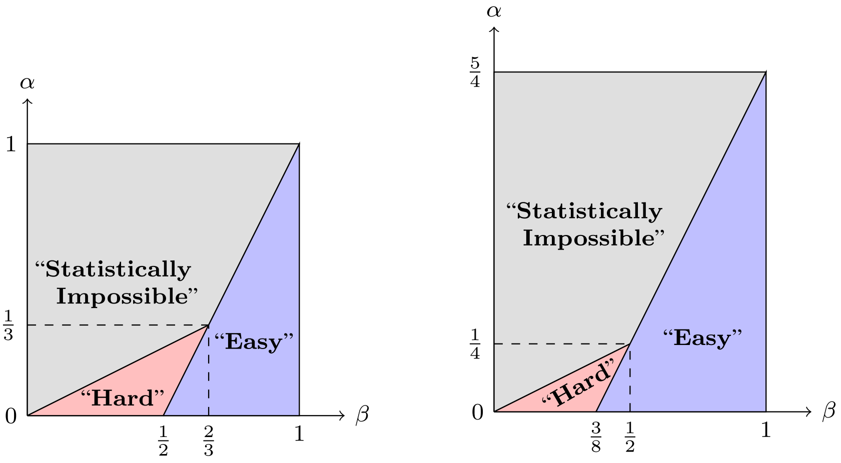
*d) Phase diagrams:* Using Theorems 1–3 we are now in a position to draw the obtained phase diagrams for our detection problems. Specifically, treating $k$ and $\lambda$ as polynomials in $n$, i.e., $k = \Theta(n^\beta)$ and $\lambda = \Theta(n^{-\alpha})$, for some $\alpha \in (0,1)$ and $\beta \in (0,1)$, we obtain the phase diagrams in Fig. 2(a), for a fixed number of submatrices $m = O(1)$. Specifically,

1) *Computationally easy regime (blue region):* there is a polynomial-time algorithm for the detection task when $\alpha < 2\beta - 1$.
2) *Computationally hard regime (red region):* there is an inefficient algorithm for detection when $\alpha < \beta/2$ and $\alpha > 2\beta - 1$, but the problem is computationally hard (no polynomial-time algorithm exists) in the sense that the class of low-degree polynomials fails in this region.
3) *Statistically impossible regime:* detection is statistically impossible when $\alpha > \frac{\beta}{2} \vee (2\beta - 1)$.

When the number of submatrices grows with $n = \omega(1)$, we get different phase diagrams depending on its value. For example, if $m = \Theta(n^{1/4})$, we get Fig. 2(b). Specifically,

1) *Computationally easy regime (blue region):* there is a polynomial-time algorithm for the detection task when $\alpha < 2\beta - \frac{3}{4}$.
2) *Computationally hard regime (red region):* there is an inefficient algorithm for detection when $\alpha < \beta/2$ and $\alpha > 2\beta - \frac{3}{4}$, but the problem is computationally hard (no polynomial-time algorithm exists) in the sense that the class of low-degree polynomials fails in this region.

[1] A sequence $(\mathbb{P}_n)_{n\in\mathbb{N}}$ of probability measures is contiguous to a sequence $(\mathbb{Q}_n)_{n\in\mathbb{N}}$, if whenever $\mathcal{A}_n \in \mathcal{F}_n$ with $\mathbb{Q}_n(\mathcal{A}_n) \to 0$, as $n \to \infty$, then $\mathbb{P}_n(\mathcal{A}_n) \to 0$ as well, over a common sequence of measurable spaces $\{(\Omega_n, \mathcal{F}_n)\}_{n\in\mathbb{N}}$.



(a) $m = O(1)$

(b) $m = \Theta(n^{1/4})$

Fig. 2. Phase diagrams for submatrix detection as a function of $k = \Theta(n^{\beta})$, and $\lambda = \Theta(n^{-\alpha})$, for $m = O(1)$ and $m = \Theta(n^{1/4})$.

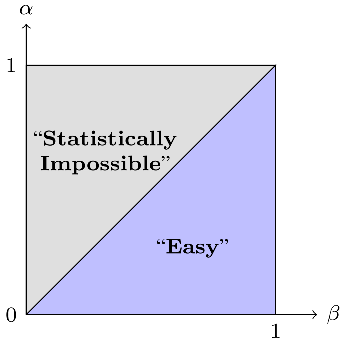


Fig. 3. Phase diagram for consecutive submatrix detection, as a function of $k = \Theta(n^{\beta})$, and $\lambda = \Theta(n^{-\alpha})$, for any $m$.

3) *Statistically impossible regime:* detection is statistically impossible when $\alpha > \frac{\beta}{2} \vee (2\beta - 3/4)$.

Finally, for the consecutive problem, we get the phase diagram in Fig. 3, independently of the value of $m$. Here, there are only two regions where the problem is either statistically impossible or easy to solve.

### *B. The Recovery Problem*

*a) Upper bounds:* We start by presenting our upper bounds for both exact and correlated types of recovery for the consecutive problem in Definition 3. To that end, we propose the following recovery algorithm. It can be shown that the maximum-likelihood (ML) estimator, minimizing the error probability, is given by (see [35, Section 4.4] for a complete derivation),

$$\hat{\mathsf{K}}_{\mathsf{ML}}(\mathsf{X}) = \text{arg max}_{\mathsf{K}\in\mathcal{K}^{\mathsf{con}}_{k,m,n}} \sum_{(i,j)\in\mathsf{K}} \mathsf{X}_{ij}. \tag{19}$$

The computational complexity of the exhaustive search in (19) is of order $n^{2m}$. Thus, for $m = O(1)$, the ML estimator runs in polynomial time, and thus, is efficient. However, if $m = \omega(1)$ then the exhaustive search is not efficient anymore. Nonetheless, the following straightforward modification of (19) provably achieves the same asymptotic performance of the ML estimator above, and at the same time computationally efficient.

Before we present this algorithm, we make a simplifying technical assumption on the possible set of planted submatrices, and then explain how this assumption can be removed. *We assume that each pair of submatrices in the underlying planted submatrices* $\mathsf{K}^{\star}$ *are at least* $k$ *columns and rows far way*. In other words, there are at least $k$ columns and $k$ rows separating any pair of submatrices in $\mathsf{K}^{\star}$. Similar assumptions are frequently taken when analyzing statistical models inspired by cryo-EM, see, for example [43]. We will refer to the above as the *separation assumption*.

Our recovery algorithm works as follows: in the $\ell \in [m]$ step, we find the ML estimate of a single submatrix using,

$$\hat{\mathsf{K}}_{\ell}(\mathsf{X}^{(\ell)}) = \text{arg max}_{\mathsf{K}\in\mathcal{K}^{\mathsf{con}}_{k,1,n}} \sum_{(i,j)\in\mathsf{K}} \mathsf{X}^{(\ell)}_{ij}, \tag{20}$$

where $\mathsf{X}^{(\ell)}$ is defined recursively as follows: $\mathsf{X}^{(1)} \triangleq \mathsf{X}$, and for $\ell \geq 2$,

$$\mathsf{X}^{(\ell)} = \mathsf{X}^{(\ell-1)} \odot \mathsf{E}(\hat{\mathsf{K}}_{\ell-1}), \tag{21}$$

where $\mathsf{E}(\hat{\mathsf{K}}_{\ell-1})$ is an $n \times n$ matrix such that $[\mathsf{E}(\hat{\mathsf{K}}_{\ell-1})]_{ij} = -\infty$, for $(i,j) \in \hat{\mathsf{K}}_{\ell-1}$, and $[\mathsf{E}(\hat{\mathsf{K}}_{\ell-1})]_{ij} = 1$, otherwise. To wit, in each step of the algorithm we "peel" the set of estimated indices (or, estimated submatrices) in previous steps from the search space. This is done by setting the corresponding entries of $\mathsf{X}$ to $-\infty$ so that the sum in (20) will not be maximized by previously chosen sets of indices. We denote by $\hat{\mathsf{K}}_{\mathsf{peel}}(\mathsf{X}) = \{\hat{\mathsf{K}}_{\ell}\}_{\ell=1}^{m}$ the output of the above algorithm.

*Remark 1:* Without the assumption above, the fact that the peeling algorithm succeeds is not trivial. If, for example, the chosen planted matrices are such that they include a pair of adjacent matrices, then it could be the case that at some step of the peeling algorithm, the estimated set of indices corresponds to a certain submatrix of the union of those adjacent matrices. However, one can easily modify the peeling algorithm, drop the assumption above, and obtain the same statistical guarantees

**Algorithm 1:** `Modified Peeling.`

```
1) Initialize $\mathsf{flag} \leftarrow 0$, $\ell \leftarrow 1$, $\mathscr{K} \leftarrow \emptyset$, $\mathsf{A} = \mathbf{0}_{n\times n}$.
2) while $\mathsf{flag} = 0$
   a) $\hat{\mathsf{K}}_\ell(\mathsf{X}) \leftarrow \arg\max_{\mathsf{K}\in\mathcal{K}^{\mathsf{con}}_{k,1,n}\setminus\mathscr{K}} \sum_{(i,j)\in\mathsf{K}} \mathsf{X}_{ij}$.
   b) $\mathsf{A}_{ij} \leftarrow 1$, for $(i,j) \in \hat{\mathsf{K}}_\ell(\mathsf{X})$, and $\mathsf{A}_{ij} \leftarrow 0$, otherwise.
   c) $\mathscr{K} \leftarrow \mathscr{K} \cup \hat{\mathsf{K}}_\ell(\mathsf{X})$.
   d) if $\langle \mathbf{J}, \mathsf{A}\rangle = mk^2$
        $\mathsf{flag} \leftarrow 1$.
   e) else
        $\ell \leftarrow \ell + 1$.
3) Output $\mathsf{A}$.
```

stated below. Indeed, consider the following modification to the peeling routine in Algorithm 1.

The key idea is as follows. In the first step, we find the $k \times k$ submatrix in $\mathsf{X}$ with the maximum sum of entries. We denote this submatrix by $\hat{\mathsf{K}}_1$. This is exactly the same first step of the peeling algorithm. In the second step, we again search for the $k \times k$ submatrix in $\mathsf{X}$ with the maximum sum of entries, but of course, remove $\hat{\mathsf{K}}_1$ from the search space. More generally, in the $\ell$-th step, we again search for the $k \times k$ submatrix in $\mathsf{X}$ with maximum sum of entries, but remove $\mathscr{K} = \cup_{i=1}^{\ell-1}\hat{\mathsf{K}}_i$ from the search space (we will not iterate over these specific *submatrices* anymore, but other submatrices which might intersect with those submatrices are acceptable and might still be in the search space). We terminate this process once $\cup_{i=1}^{\ell}\hat{\mathsf{K}}_i \in \mathcal{K}^{\mathsf{con}}_{k,m,n}$, i.e., the union of the estimated sets of matrices can cast as a proper set of planted submatrices. This can easily be checked by forming the matrix $\mathsf{A}$ in Step 2(b), and checking the conditions in Step 2(d). If the actually planted submatrices are not adjacent, then this will be the case (under the conditions in the theorem below) after $\ell = m$ steps, with high probability. Otherwise, if at least two planted submatrices are adjacent, then while $\ell$ might be larger than $m$ it is bounded by $n^2$, and it is guaranteed that such a union exists. Once we find such a union, it is easy to revert the set of $m$ consecutive $k \times k$ submatrices from $\mathsf{A}$.

We have the following result.

*Theorem 5 (Recovery Upper Bounds):* Consider the recovery problem in Definition 3, and let $\mathsf{C}$ be a universal constant. Then, we have the following set of bounds:

1) (ML Exact Recovery) Consider the ML estimator in (19). If
$$\liminf_{n\to\infty} \frac{\lambda}{\sqrt{\mathsf{C}k^{-1}\log n}} > 1, \tag{22}$$
then exact recovery is possible.

2) (Peeling Exact Recovery) Consider the peeling estimator in (20), and assume that the separation assumption holds. Then, if
$$\liminf_{n\to\infty} \frac{\lambda}{\sqrt{\mathsf{C}k^{-1}\log n}} > 1, \tag{23}$$
then exact recovery is possible.

3) (Peeling Correlated Recovery) Consider the peeling estimator in (20), and assume that the separation assumption holds. If
$$\liminf_{n\to\infty} \frac{\lambda}{\sqrt{\mathsf{C}k^{-2}\log n}} > 1, \tag{24}$$
then correlated recovery is possible.

*b) Lower bounds:* The following result shows that under certain conditions, exact and correlated recoveries are impossible.

*Theorem 6 (Information-Theoretic Recovery Lower Bounds):* Consider the recovery problem in Definition 3. Then:

1) If $\lambda < \mathsf{C}\sqrt{\frac{\log m}{k}}$, exact recovery is impossible, i.e.,
$$\inf_{\hat{\mathsf{K}}} \sup_{\mathsf{K}^\star\in\mathcal{K}^{\mathsf{con}}_{k,m,n}} \mathbb{P}[\hat{\mathsf{K}}(\mathsf{X}) \neq \mathsf{K}^\star] > \frac{1}{2},$$
where the infimum ranges over all measurable functions of the matrix $\mathsf{X}$.

2) If $\lambda = o(k^{-1})$, correlated recovery is impossible, i.e., $\sup_{\mathsf{K}^\star\in\mathcal{K}^{\mathsf{con}}_{k,m,n}} \mathsf{overlap}(\mathsf{K}^\star, \hat{\mathsf{K}}) = o(mk^2)$.

To prove the impossibility of exact recovery we use Fano's inequality. To that end, we construct families of distributions with relatively large cardinality such that the distributions are all relatively close in KL divergence. For the correlated recovery impossibility result we use the I-MMSE formula [44] and some arguments from [45, Subsection 3.1.3]. From Theorem 6, we see that similarly to the detection problem, the consecutive recovery problem is either statistically impossible or easy to solve. The corresponding phase diagram for exact and correlated types of recoveries is given in Fig. 4. Roughly speaking, exact recovery is possible if $\lambda = \omega(k^{-1/2})$ and impossible if $\lambda = o(k^{-1/2})$. Correlated recovery is possible if $\lambda = \omega(k^{-1})$ and impossible if $\lambda = o(k^{-1})$.

A few remarks are in order. First, note that there is a gap between detection and exact recovery; the barrier for $\lambda$ for the former is at $k^{-1}$, while for the latter at $k^{-1/2}$. In the context of cryo-EM, this indicates a gap between the ability to detect the existence of particle images in the data set, and the ability to perform successful particle picking (exact recovery). Recently, new computational methods were devised to elucidate molecular structures without particle picking, thus bypassing the limit of exact recovery, allowing constructing structures in very low SNR environments, e.g., [43], [46], [47]. This in turn opens the door to recovering small molecular structures that induce low SNR [48]. Second, there is no gap between detection and correlated recovery, and these different tasks are asymptotically statistically the same. The same gap exists between correlated and exact recoveries, implying that exact recovery is strictly harder than correlated recovery.

## IV. Proofs

In this section, we provide proofs for Theorems 1–3, and the peeling exact recovery bound in Theorem 5. Due to page limitation, some details in these proofs, as well as the proofs of Theorem 6, and the ML exact and correlated recovery bounds in Theorem 5, are relegated to [35].

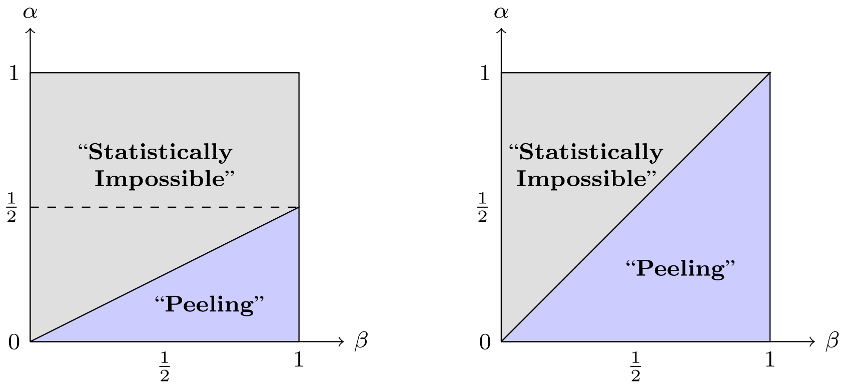


Fig. 4. Phase diagram for consecutive submatrix exact recovery (left) and correlated recovery (right), as a function of $k = \Theta(n^{\beta})$, and $\lambda = \Theta(n^{-\alpha})$, for any $m$.

### A. Proof of Theorem 1

We analyze the performance of the sum test in (10). The analysis of the tests in (11) and (12) rely on similar arguments, and can be found in [35]. Define $\tau \triangleq \frac{mk^2\lambda}{2}$. Let us analyze the corresponding error probability. On the one hand, under $\mathcal{H}_0$, it is clear that $\mathsf{T}_{\mathsf{sum}}(\mathsf{X}) \sim \mathcal{N}(0, n^2)$. Thus,

$$\mathbb{P}_{\mathcal{H}_0}\left(\mathcal{A}_{\mathsf{sum}}(\mathsf{X}) = 1\right) = \mathbb{P}_{\mathcal{H}_0}\left(\mathsf{T}_{\mathsf{sum}}(\mathsf{X}) \geq \tau\right) \leq \frac{1}{2}\exp\left(-\frac{\tau^2}{2n^2}\right). \tag{25}$$

On the other hand, under $\mathcal{H}_1$, $\mathsf{T}_{\mathsf{sum}}(\mathsf{X}) \sim \mathcal{N}(mk^2\lambda, n^2)$. Thus,

$$\mathbb{P}_{\mathcal{H}_1}\left(\mathcal{A}_{\mathsf{sum}}(\mathsf{X}) = 0\right) = \mathbb{P}_{\mathcal{H}_1}\left(\mathsf{T}_{\mathsf{sum}}(\mathsf{X}) \leq \tau\right) \leq \frac{1}{2}\exp\left(-\frac{(\tau - mk^2\lambda)^2}{2n^2}\right). \tag{26}$$

Substituting $\tau = \frac{mk^2\lambda}{2}$, we obtain that

$$\mathsf{R}\left(\mathcal{A}_{\mathsf{sum}}\right) \leq \exp\left(-\frac{m^2\, k^4\lambda^2}{8n^2}\right). \tag{27}$$

Thus, if $\frac{mk^2\lambda}{n} \to \infty$, then $\mathsf{R}(\mathcal{A}_{\mathsf{sum}}) \to 0$, as $n \to \infty$. Note that the analysis above holds true for both detection problems in Definitions 1 and 2.

### B. Proof of Theorem 2

*1) Submatrix Detection:* Recall that the optimal test $\mathcal{A}_n^*$ that minimizes the risk is the likelihood ratio test defined as follows,

$$\mathcal{A}_n^*(\mathsf{X}) \triangleq \mathbb{1}\left\{\mathsf{L}_n(\mathsf{X}) \geq 1\right\}, \tag{28}$$

where $\mathsf{L}_n(\mathsf{X}) \triangleq \frac{\mathbb{P}_{\mathcal{H}_1}(\mathsf{X})}{\mathbb{P}_{\mathcal{H}_0}(\mathsf{X})}$. The optimal risk, denoted by $\mathsf{R}^* = \mathsf{R}(\mathcal{A}_n^*)$, can be lower bounded using the Cauchy–Schwartz inequality as follows,

$$\begin{aligned}\mathsf{R}^* &= 1 - \frac{1}{2}\mathbb{E}_{\mathcal{H}_0}\left|\mathsf{L}_n(\mathsf{X}) - 1\right| \\ &\geq 1 - \frac{1}{2}\sqrt{\mathbb{E}_{\mathcal{H}_0}\left[(\mathsf{L}_n(\mathsf{X}))^2\right] - 1}.\end{aligned} \tag{29}$$

Thus, in order to lower bound the risk, we need to upper bound $\mathbb{E}_{\mathcal{H}_0}[(\mathsf{L}_n(\mathsf{X}))^2]$. Below, we provide a lower bound that holds for any pair of distributions $\mathcal{P}$ and $\mathcal{Q}$.

*Corollary 7:* The following holds:

$$\mathbb{E}_{\mathcal{H}_0}\left[(\mathsf{L}_n(\mathsf{X}))^2\right] = \mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[(1 + \chi^2(\mathcal{P}||\mathcal{Q}))^{|\mathsf{K}\cap\mathsf{K}'|}\right] \tag{30}$$

$$\leq \mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot|\mathsf{K}\cap\mathsf{K}'|}\right], \tag{31}$$

where $\mathsf{K}$ and $\mathsf{K}'$ are two independent copies drawn uniformly at random from $\mathcal{K}_{k,m,n}$ (or, $\bar{\mathcal{K}}_{k,m,n}$), and

$$\chi^2(\mathcal{P}||\mathcal{Q}) \triangleq \mathbb{E}_{X\sim\mathcal{Q}}\left[\frac{\mathcal{P}(X)}{\mathcal{Q}(X)}\right]^2 - 1. \tag{32}$$

*Proof of Corollary 7:* First, note that the likelihood can be written as follows:

$$\mathsf{L}_n(\mathsf{X}) = \frac{\mathbb{P}_{\mathcal{H}_1}(\mathsf{X})}{\mathbb{P}_{\mathcal{H}_0}(\mathsf{X})} = \mathbb{E}_{\mathsf{K}}\left(\prod_{(i,j)\in\mathsf{K}}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\right), \tag{33}$$

where the expectation in (33) is taken w.r.t. $\mathsf{K} \sim \mathsf{Unif}(\mathcal{K}_{k,m,n})$. Now, note that the square of the right-hand side of (33) can be rewritten as:

$$\begin{aligned}&\left[\mathbb{E}_{\mathsf{K}}\left(\prod_{(i,j)\in\mathsf{K}}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\right)\right]^2 \\ &= \mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left(\prod_{(i,j)\in\mathsf{K}}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\prod_{(i,j)\in\mathsf{K}'}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\right).\end{aligned} \tag{34}$$

Therefore,

$$\mathbb{E}_{\mathcal{H}_0}\left[(\mathsf{L}_n(\mathsf{X}))^2\right] = \mathbb{E}_{\mathcal{H}_0}\left[\mathbb{E}_{\mathsf{K}}\left(\prod_{(i,j)\in\mathsf{K}}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\right)\right]^2 \tag{35}$$

$$= \mathbb{E}_{\mathcal{H}_0}\left[\mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left(\prod_{(i,j)\in\mathsf{K}}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\prod_{(i,j)\in\mathsf{K}'}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\right)\right] \tag{36}$$

$$= \mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[\mathbb{E}_{\mathcal{H}_0}\left(\prod_{(i,j)\in\mathsf{K}}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\prod_{(i,j)\in\mathsf{K}'}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\right)\right] \tag{37}$$

$$\begin{aligned}= \mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\Bigg[\mathbb{E}_{\mathcal{H}_0}\Bigg(&\prod_{(i,j)\in\mathsf{K}\cup\mathsf{K}'\setminus\mathsf{K}\cap\mathsf{K}'}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\\ &\prod_{(i,j)\in\mathsf{K}\cap\mathsf{K}'}\left(\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\right)^2\Bigg)\Bigg]\end{aligned} \tag{38}$$

$$\begin{aligned}= \mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\Bigg[&\prod_{(i,j)\in\mathsf{K}\cup\mathsf{K}'\setminus\mathsf{K}\cap\mathsf{K}'}\mathbb{E}_{\mathcal{H}_0}\left[\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\right]\\ &\prod_{(i,j)\in\mathsf{K}\cap\mathsf{K}'}\mathbb{E}_{\mathcal{H}_0}\left[\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\right]^2\Bigg]\end{aligned} \tag{39}$$

$$\overset{(a)}{=} \mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[\left(\mathbb{E}_{\mathcal{H}_0}\left[\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})}\right]^2\right)^{|\mathsf{K}\cap\mathsf{K}'|}\right] \tag{40}$$

$$= \mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[\left(1+\chi^2(\mathcal{P}||\mathcal{Q})\right)^{|\mathsf{K}\cap\mathsf{K}'|}\right] \tag{41}$$

$$\overset{(b)}{\leq} \mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot|\mathsf{K}\cap\mathsf{K}'|}\right], \tag{42}$$

where $(a)$ is because $\mathbb{E}_{\mathcal{Q}}\frac{\mathcal{P}(\mathsf{X}_{ij})}{\mathcal{Q}(\mathsf{X}_{ij})} = 1$, and $(b)$ is because $1+x \leq \exp(x)$, for any $x\in\mathbb{R}$. ■

Based on Corollary 7, it suffices to upper bound $\mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot|\mathsf{K}\cap\mathsf{K}'|}]$. Recall that $\mathsf{K}$ and $\mathsf{K}'$ are decomposed as $\mathsf{K} = \bigcup_{\ell=1}^m \mathsf{S}_\ell\times\mathsf{T}_\ell$ and $\mathsf{K}' = \bigcup_{\ell=1}^m \mathsf{S}'_\ell\times\mathsf{T}'_\ell$. Thus, we note that the intersection of $\mathsf{K}$ and $\mathsf{K}'$ can be rewritten as

$$|\mathsf{K}\cap\mathsf{K}'| = \sum_{\ell_1=1}^m\sum_{\ell_2=1}^m |(\mathsf{S}_{\ell_1}\cap\mathsf{S}'_{\ell_2})\times(\mathsf{T}_{\ell_1}\cap\mathsf{T}'_{\ell_2})| \tag{43}$$

$$= \sum_{\ell_1=1}^m\sum_{\ell_2=1}^m |(\mathsf{S}_{\ell_1}\cap\mathsf{S}'_{\ell_2})|\cdot|(\mathsf{T}_{\ell_1}\cap\mathsf{T}'_{\ell_2})|. \tag{44}$$

For each $\ell_1,\ell_2\in[m]$, define $\mathsf{Z}_{\ell_1,\ell_2}\triangleq|(\mathsf{S}_{\ell_1}\cap\mathsf{S}'_{\ell_2})|$ and $\mathsf{R}_{\ell_1,\ell_2}\triangleq|(\mathsf{T}_{\ell_1}\cap\mathsf{T}'_{\ell_2})|$. Note that the sequence of random variables $\{\mathsf{Z}_{\ell_1,\ell_2}\}_{\ell_1,\ell_2}$ are statistically independent of the sequence $\{\mathsf{R}_{\ell_1,\ell_2}\}_{\ell_1,\ell_2}$. Next, it is easy to show that $\mathsf{Z}_{\ell_1,\ell_2}\sim\mathsf{Hypergeometric}(n,k,k)$ and $\mathsf{R}_{\ell_1,\ell_2}\sim\mathsf{Hypergeometric}(n,k,k)$, for each $\ell_1,\ell_2\in[m]$, for any $\ell_1,\ell_2\in[m]$. Indeed, if we have an urn of $n$ balls among which $k$ balls are red, the random variable $\mathsf{Z}_{\ell_1,\ell_2}$ (and $\mathsf{R}_{\ell_1,\ell_2}$) is exactly the number of red balls if we draw $k$ balls from the urn uniformly at random without replacement, which is the definition of a Hypergeometric random variable. While the random variables $\{\mathsf{Z}_{\ell_1,\ell_2}\}_{\ell_1,\ell_2}$ (and similarly $\{\mathsf{R}_{\ell_1,\ell_2}\}_{\ell_1,\ell_2}$) are not independent, they are negatively associated. Thus,

$$\mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot|\mathsf{K}\cap\mathsf{K}'|}\right]$$

$$\leq \prod_{\ell_1=1}^m\prod_{\ell_2=1}^m\mathbb{E}\left[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot\mathsf{Z}_{\ell_1,\ell_2}\mathsf{R}_{\ell_1,\ell_2}}\right] \tag{45}$$

$$= \left[\mathbb{E}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot\mathsf{Z}_{1,1}\mathsf{R}_{1,1}}\right)\right]^{m^2}. \tag{46}$$

Next, it is well-known that $\mathsf{Z}_{1,1} = \mathsf{Hypergeometric}(n,k,k)$ (and similarly $\mathsf{R}_{1,1} = \mathsf{Hypergeometric}(n,k,k)$) is stochastically dominated by $\mathsf{B}\sim\mathsf{Binomial}(k,k/n) = \sum_{i=1}^k\mathsf{Bern}(k/n)$. Thus,

$$\mathbb{E}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot\mathsf{Z}_{1,1}\mathsf{R}_{1,1}}\right) \leq \mathbb{E}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot\mathsf{B}\mathsf{B}'}\right), \tag{47}$$

where $\mathsf{B}'$ be an independent copy of $\mathsf{B}$. Thus,

$$\mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot|\mathsf{K}\cap\mathsf{K}'|}\right] \leq \left[\mathbb{E}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot\mathsf{B}\mathsf{B}'}\right)\right]^{m^2}. \tag{48}$$

We show that, if $\chi^2(\mathcal{P}||\mathcal{Q})$ satisfies the condition of Theorem 2, the term on the right-hand side of (48) is at most $1+\delta$, for any $\delta>0$. We have

$$\left[\mathbb{E}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot\mathsf{B}\mathsf{B}'}\right)\right]^{m^2}$$

$$= \left[\mathbb{E}\left(1+\frac{k}{n}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\mathsf{B}}-1\right)\right)^k\right]^{m^2}. \tag{49}$$

Next, note that $\mathsf{B}\leq k$, and we assume that $\chi^2(\mathcal{P}||\mathcal{Q})\leq\frac{1}{k}$, for reasons that will become clear later on. Therefore, using the inequality $e^x-1\leq x+x^2$, for $x<1$, the following holds

$$\left[\mathbb{E}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot\mathsf{B}\mathsf{B}'}\right)\right]^{m^2}$$

$$\leq \left[\mathbb{E}\left(1+\frac{k}{n}\left(\chi^2(\mathcal{P}||\mathcal{Q})\mathsf{B}+\chi^4(\mathcal{P}||\mathcal{Q})\mathsf{B}^2\right)\right)^k\right]^{m^2} \tag{50}$$

$$\leq \left[\mathbb{E}\left(1+2\frac{k}{n}\chi^2(\mathcal{P}||\mathcal{Q})\mathsf{B}\right)^k\right]^{m^2} \tag{51}$$

$$\leq \left[\mathbb{E}\left(e^{2\frac{k^2}{n}\chi^2(\mathcal{P}||\mathcal{Q})\mathsf{B}}\right)\right]^{m^2} \tag{52}$$

$$= \left[1+\frac{k}{n}\left(e^{2\frac{k^2}{n}\chi^2(\mathcal{P}||\mathcal{Q})}-1\right)\right]^{km^2}. \tag{53}$$

This is at most $1+\delta$ if

$$\frac{k}{n}\left(e^{2\frac{k^2}{n}\chi^2(\mathcal{P}||\mathcal{Q})}-1\right) \leq (1+\delta)^{\frac{1}{km^2}}-1. \tag{54}$$

Since $(1+\delta)^{\frac{1}{km^2}}-1\geq\log(1+\delta)/(km^2)$, this is implied by

$$\chi^2(\mathcal{P}||\mathcal{Q}) \leq \frac{n}{2\,k^2}\log\left(1+\frac{n\log(1+\delta)}{m^2\,k^2}\right). \tag{55}$$

Putting altogether, we obtained that $\mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot|\mathsf{K}\cap\mathsf{K}'|}]\leq 1+\delta$, if

$$\chi^2(\mathcal{P}||\mathcal{Q}) \leq \min\left\{\frac{1}{k},\frac{n}{2\,k^2}\log\left(1+\frac{n\log(1+\delta)}{m^2\,k^2}\right)\right\} \tag{56}$$

$$= \min\left\{\frac{1}{k},\frac{n^2\log(1+\delta)}{2\,m^2\,k^4}\right\}. \tag{57}$$

Finally, note that in the Gaussian case, $\chi^2(\mathcal{N}(\lambda,1)||\mathcal{N}(0,1)) = \frac{1}{2}[\exp(\lambda^2)-1]$. Thus, for $\lambda = o(1)$, we have $\chi^2(\mathcal{N}(\lambda,1)||\mathcal{N}(0,1)) \to \frac{\lambda^2}{2}$, which concludes the proof.

*2) Consecutive Submatrix Detection:* For the consecutive case, we notice that by using the steps as in the previous subsection, we have

$$\mathbb{E}_{\mathcal{H}_0}\left[\left(\mathsf{L}_n\left(\mathsf{X}\right)\right)^2\right] \leq \mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot|\mathsf{K}\cap\mathsf{K}'|}\right], \tag{58}$$

where $\mathsf{K}$ and $\mathsf{K}'$ are two independent copies drawn uniformly at random from $\mathcal{K}^{\mathsf{con}}_{k,m,n}$. The key distinction from the previous case lies in the distribution of $|\mathsf{K}\cap\mathsf{K}'|$. Recall that $\mathsf{K}$ and $\mathsf{K}'$ are decomposed as $\mathsf{K} = \bigcup_{\ell=1}^m \mathsf{S}_\ell \times \mathsf{T}_\ell$ and $\mathsf{K}' = \bigcup_{\ell=1}^m \mathsf{S}'_\ell \times \mathsf{T}'_\ell$. Thus, we note that the intersection of $\mathsf{K}$ and $\mathsf{K}'$ can be rewritten as

$$|\mathsf{K}\cap\mathsf{K}'| = \sum_{\ell_1=1}^m\sum_{\ell_2=1}^m |(\mathsf{S}_{\ell_1}\cap\mathsf{S}'_{\ell_2})\times(\mathsf{T}_{\ell_1}\cap\mathsf{T}'_{\ell_2})| \tag{59}$$

$$= \sum_{\ell_1=1}^m\sum_{\ell_2=1}^m |(\mathsf{S}_{\ell_1}\cap\mathsf{S}'_{\ell_2})|\cdot|(\mathsf{T}_{\ell_1}\cap\mathsf{T}'_{\ell_2})| \tag{60}$$

$$\triangleq \sum_{\ell_1=1}^m\sum_{\ell_2=1}^m \mathsf{Z}_{\ell_1,\ell_2}. \tag{61}$$

Note that for a given pair $(\ell_1,\ell_2)$, we have

$$\mathbb{P}(|(\mathsf{S}_{\ell_1}\cap\mathsf{S}'_{\ell_2})| = z) = \begin{cases} \frac{n-2k+1}{n}, & \text{for } z = 0 \\ \frac{2}{n}, & \text{for } z = 1,2,\ldots,k-1 \\ \frac{1}{n}, & \text{for } z = k, \end{cases} \tag{62}$$

and the exact same distribution for $|(\mathsf{T}_{\ell_1}\cap\mathsf{T}'_{\ell_2})|$. Thus, we may write $\mathsf{Z}_{\ell_1,\ell_2} \overset{(d)}{=} \mathsf{H}\cdot\mathsf{H}'$, where $\mathsf{H}$ and $\mathsf{H}'$ are statistically independent and follow the distribution given in (62). Thus, using the fact that the random variables $\{\mathsf{Z}_{\ell_1,\ell_2}\}_{\ell_1,\ell_2}$ are negatively associated, we get,

$$\mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot|\mathsf{K}\cap\mathsf{K}'|}\right] \leq \prod_{\ell_1=1}^m\prod_{\ell_2=1}^m \mathbb{E}\left[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot\mathsf{Z}_{\ell_1,\ell_2}}\right]$$

$$= \left[\mathbb{E}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot\mathsf{H}\cdot\mathsf{H}'}\right)\right]^{m^2}. \tag{63}$$

Now,

$$\mathbb{E}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot\mathsf{H}\cdot\mathsf{H}'}\right)$$

$$= \mathbb{E}\left(\frac{n-2k+1}{n} + \frac{2}{n}\sum_{i=1}^{k-1} e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot i\mathsf{H}'} + \frac{e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot k\mathsf{H}'}}{n}\right)$$

$$\leq \mathbb{E}\left(\frac{n-2\,k}{n} + \frac{2\,k}{n}e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot k\mathsf{H}'}\right)$$

$$= \frac{n-2\,k}{n} + \frac{2\,k}{n}\left(\frac{n-2k+1}{n} + \frac{2}{n}\sum_{i=1}^{k-1} e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot ik} + \frac{e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot k^2}}{n}\right)$$

$$\leq \frac{n-2\,k}{n} + \frac{2\,k}{n}\left(\frac{n-2\,k}{n} + \frac{2\,k}{n}e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot k^2}\right)$$

$$= 1 + \frac{4\,k^2}{n^2}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot k^2} - 1\right). \tag{64}$$

Therefore,

$$\mathbb{E}_{\mathsf{K}\perp\!\!\!\perp\mathsf{K}'}\left[e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot|\mathsf{K}\cap\mathsf{K}'|}\right]$$

$$\leq \left[1 + \frac{4\,k^2}{n^2}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})\cdot k^2} - 1\right)\right]^{m^2}. \tag{65}$$

This is at most $1+\delta$ if,

$$\frac{4\,k^2}{n^2}\left(e^{\chi^2(\mathcal{P}||\mathcal{Q})k^2} - 1\right) \leq (1+\delta)^{\frac{1}{m^2}} - 1. \tag{66}$$

Since $(1+\delta)^{\frac{1}{m^2}} - 1 \geq \log(1+\delta)/(m^2)$, this is implied by

$$\chi^2(\mathcal{P}||\mathcal{Q}) \leq \frac{1}{k^2}\log\left(1 + \frac{n^2\log(1+\delta)}{4\,k^2\,m^2}\right). \tag{67}$$

Finally, note that since $km \leq n$, the logarithmic factor in (67) can be lower bounded by $\log(1+\log(1+\delta)/4)$, which concludes the proof.

### C. Proof of Theorem 3

In order to prove Theorem 3, we use the following result [49, Theorem 2.6].

*Lemma 2:* Let $\mathbf{S}$ be an $n$ dimensional random vector drawn from some distribution $\mathcal{D}_n$, and let $\mathbf{Z}$ be an i.i.d. $n$ dimensional random vector with standard normal entries. Consider the detection problem:

$$\mathcal{H}_0 : \mathbf{Y} = \mathbf{Z} \quad \text{vs.} \quad \mathcal{H}_1 : \mathbf{Y} = \mathbf{S} + \mathbf{Z}. \tag{68}$$

Then,

$$\left\|\mathsf{L}_n^{\leq \mathsf{D}}\right\|^2_{\mathcal{H}_0} = \mathbb{E}_{\mathbf{S}\perp\!\!\!\perp\mathbf{S}'}\left[\sum_{d=0}^{\mathsf{D}}\frac{1}{d!}\left\langle\mathbf{S},\mathbf{S}'\right\rangle^d\right], \tag{69}$$

where $\mathbf{S}$ and $\mathbf{S}'$ are drawn from $\mathcal{D}_n$, and $\mathsf{L}_n^{\leq D}$ is the D-low-degree likelihood ratio.

Our SD problem falls under the setting of Lemma 2. Specifically, let $\mathsf{K} \sim \mathsf{Unif}[\mathcal{K}_{k,m,n}]$, and define $\tilde{\mathbf{S}}$ to be an $n\times n$ matrix such that $[\tilde{\mathbf{S}}]_{ij} = \lambda$, if $i,j\in\mathsf{K}$, and $[\tilde{\mathbf{S}}]_{ij} = 0$, otherwise. Also, we define $\mathbf{S}$ as the vectorized version of $\tilde{\mathbf{S}}$. Then, it is clear that our SD problem cast as the detection problem in Lemma 2, and thus,

$$\left\|\mathsf{L}_n^{\leq \mathsf{D}}\right\|^2 = \mathbb{E}_{\mathbf{S}\perp\!\!\!\perp\mathbf{S}'}\left[\sum_{d=0}^{\mathsf{D}}\frac{1}{d!}\left\langle\mathbf{S},\mathbf{S}'\right\rangle^d\right] \tag{70}$$

$$= \sum_{d=0}^{\mathsf{D}}\frac{\lambda^{2\,d}}{d!}\mathbb{E}\left|\mathsf{K}\cap\mathsf{K}'\right|^d, \tag{71}$$

where we have used the fact that $\langle\mathbf{S},\mathbf{S}'\rangle = \mathbf{S}^T\mathbf{S}' = \|\mathbf{S}\odot\mathbf{S}\|_1 = \lambda^2|\mathsf{K}\cap\mathsf{K}'|$, and $\mathsf{K}'$ is an independent copy of $\mathsf{K}$.

Using a certain stochastic dominance argument, we prove in [35] that,

$$\mathbb{E}\,|\mathsf{K}\cap\mathsf{K}'|^d \le \mathsf{B}_d^2 \max\left\{\frac{m^2\,k^4}{n^2}, \left(\frac{m^2\,k^4}{n^2}\right)^d\right\}, \tag{72}$$

where $\mathsf{B}_d$ is the $d$th Bell number. Thus,

$$\left\|\mathsf{L}_n^{\le\mathsf{D}}\right\|^2 \le 1+\sum_{d=1}^{\mathsf{D}}\frac{\lambda^{2\,d}}{d!}\mathsf{B}_d^2\max\left\{\frac{m^2\,k^4}{n^2}, \left(\frac{m^2\,k^4}{n^2}\right)^d\right\} \tag{73}$$

$$\triangleq 1+\sum_{d=1}^{\mathsf{D}}\mathsf{T}_d. \tag{74}$$

If $\frac{m^2\,k^4}{n^2}<1$, then it is clear that for $\sum_{d=1}^{\mathsf{D}}\mathsf{T}_d=O(1)$, it suffices that $\lambda<1$. On the other hand, if $\frac{m^2\,k^4}{n^2}>1$, then consider the ratio between successive terms:

$$\frac{\mathsf{T}_{d+1}}{\mathsf{T}_d}=\frac{\mathsf{B}_{d+1}^2}{(d+1)\mathsf{B}_d^2}\lambda^2\,m^2\frac{k^4}{n^2}. \tag{75}$$

Thus if $\lambda$ is small enough, namely if

$$\frac{mk^2\lambda}{n}\le\frac{\sqrt{d+1}}{\sqrt{2}}\frac{\mathsf{B}_d}{\mathsf{B}_{d+1}}, \tag{76}$$

then $\frac{\mathsf{T}_{d+1}}{\mathsf{T}_d}\le 1/2$, for all $1\le d\le\mathsf{D}$. In this case, by comparing with a geometric sum, we may bound $\|\mathsf{L}_n^{\le\mathsf{D}}\|^2\le O(1)$. To show that the analysis above is tight, we prove in [35] that if $\lambda=\omega(n/(mk^2))$, then $\|\mathsf{L}_n^{\le\mathsf{D}}\|^2=\omega(1)$. This concludes the proof.

### D. Proof of Theorem 5 for Exact Recovery

We analyze the first step of the peeling algorithm (which boils down to the ML estimator for a single planted submatrix), and the strategy to bound each of the other sequential steps is exactly the same. Recall that,

$$\hat{\mathsf{K}}_1(\mathsf{X})=\arg\max\nolimits_{\mathsf{K}\in\mathcal{K}^{\mathsf{con}}_{k,1,n}}\mathcal{S}(\mathsf{K}), \tag{77}$$

where $\mathcal{S}(\mathsf{K})\triangleq\sum_{(i,j)\in\mathsf{K}}\mathsf{X}_{ij}$. where $\mathcal{S}(\mathsf{K})\triangleq\sum_{(i,j)\in\mathsf{K}}\mathsf{X}_{ij}$. We next prove the conditions for which $\hat{\mathsf{K}}_1(\mathsf{X})=\mathsf{K}_\ell^\star$, with high probability, for some $\ell\in[m]$, where $\mathsf{K}^\star=\cup_{\ell=1}^m\mathsf{K}_\ell^\star$ are the $m$ planted submatrices. To prove the theorem it suffices to show that $\mathcal{S}(\mathsf{K})>\max_{\ell\in[m]}\mathcal{S}(\mathsf{K}_\ell^\star)$ is asymptotically small, for all feasible $\mathsf{K}$ with $\mathsf{K}\neq\mathsf{K}_\ell^\star$, for $\ell\in[m]$. Let $\mathsf{D}_\ell(\mathsf{K})\triangleq\mathcal{S}(\mathsf{K}_\ell^\star)-\mathcal{S}(\mathsf{K})$. Note that

$$\mathsf{D}_\ell(\mathsf{K})=\sum_{(i,j)\in\mathsf{K}_\ell^\star}\mathsf{X}_{ij}-\sum_{(i,j)\in\mathsf{K}}\mathsf{X}_{ij} \tag{78}$$

$$=\sum_{(i,j)\in\mathsf{K}_\ell^\star}\mathbb{E}\mathsf{X}_{ij}-\sum_{(i,j)\in\mathsf{K}}\mathbb{E}\mathsf{X}_{ij}+\sum_{(i,j)\in\mathsf{K}_\ell^\star}[\mathsf{X}_{ij}-\mathbb{E}\mathsf{X}_{ij}]$$

$$-\sum_{(i,j)\in\mathsf{K}}[\mathsf{X}_{ij}-\mathbb{E}\mathsf{X}_{ij}] \tag{79}$$

$$=\lambda\cdot(k^2-|\mathsf{K}^\star\cap\mathsf{K}|)+\sum_{(i,j)\in\mathsf{K}_\ell^\star\setminus\mathsf{K}}[\mathsf{X}_{ij}-\lambda]$$

$$-\sum_{(i,j)\in\mathsf{K}\setminus\mathsf{K}_\ell^\star}[\mathsf{X}_{ij}-\mathbb{E}\mathsf{X}_{ij}] \tag{80}$$

$$=\lambda\cdot(k^2-|\mathsf{K}^\star\cap\mathsf{K}|)+\mathsf{W}_1(\mathsf{K})+\mathsf{W}_2(\mathsf{K}), \tag{81}$$

where $\mathsf{W}_1(\mathsf{K})\triangleq\sum_{(i,j)\in\mathsf{K}_\ell^\star\setminus\mathsf{K}}[\mathsf{X}_{ij}-\lambda]$ and $\mathsf{W}_2(\mathsf{K})\triangleq-\sum_{(i,j)\in\mathsf{K}\setminus\mathsf{K}_\ell^\star}[\mathsf{X}_{ij}-\mathbb{E}\mathsf{X}_{ij}]$. Because $|\mathsf{K}|=|\mathsf{K}_\ell^\star|=k^2$, we have $|\mathsf{K}_\ell^\star\setminus\mathsf{K}|=|\mathsf{K}\setminus\mathsf{K}_\ell^\star|=k^2-|\mathsf{K}_\ell^\star\cap\mathsf{K}|$. Thus, both $\mathsf{W}_1(\mathsf{K})$ and $\mathsf{W}_2(\mathsf{K})$ are composed of sum of $k^2-|\mathsf{K}_\ell^\star\cap\mathsf{K}|$ i.i.d. centered Gaussian random variables with unit variance. Accordingly, for $i=1,2$, and each fixed $\mathsf{K}$,

$$\mathbb{P}\left(\mathsf{W}_i(\mathsf{K})\le-\frac{\lambda}{2}(k^2-|\mathsf{K}^\star\cap\mathsf{K}|)\right)$$

$$\le\frac{1}{2}\exp\left[-\frac{\lambda^2}{8}\frac{(k^2-|\mathsf{K}^\star\cap\mathsf{K}|)^2}{k^2-|\mathsf{K}_\ell^\star\cap\mathsf{K}|}\right]. \tag{82}$$

Therefore, by the union bound and (81),

$$\mathbb{P}\left(\mathsf{D}_\ell(\mathsf{K})\le0\right)\le\exp\left[-\frac{\lambda^2}{8}\frac{(k^2-|\mathsf{K}^\star\cap\mathsf{K}|)^2}{k^2-|\mathsf{K}_\ell^\star\cap\mathsf{K}|}\right]. \tag{83}$$

Note that due to the separation assumption, it must be the case that either $|\mathsf{K}^\star\cap\mathsf{K}|=|\mathsf{K}_j^\star\cap\mathsf{K}|\neq0$, for some $j\in[m]$, or $|\mathsf{K}^\star\cap\mathsf{K}|=0$. In the later case, we have

$$\mathbb{P}\left(\mathsf{D}_\ell(\mathsf{K})\le0\right)\le\exp\left[-\frac{\lambda^2\,k^2}{8}\right], \tag{84}$$

while in the former the exists a unique $j\in[m]$, such that,

$$\min_{\ell\in[m]}\mathbb{P}\left(\mathsf{D}_\ell(\mathsf{K})\le0\right)\le\min_{\ell\in[m]}\exp\left[-\frac{\lambda^2}{8}\frac{(k^2-|\mathsf{K}_j^\star\cap\mathsf{K}|)^2}{k^2-|\mathsf{K}_\ell^\star\cap\mathsf{K}|}\right]$$

$$\le\exp\left[-\frac{\lambda^2}{8}(k^2-|\mathsf{K}_j^\star\cap\mathsf{K}|)\right]\le\exp\left[-\frac{\lambda^2\,k}{8}\right], \tag{85}$$

where the third inequity is since $\mathsf{K}_j^\star,\mathsf{K}\in\mathcal{K}^{\mathsf{con}}_{k,1,n}$ and $\mathsf{K}_j^\star\cap\mathsf{K}\neq\emptyset$, we must have that $|\mathsf{K}_j^\star\cap\mathsf{K}|\le k^2-k$.Accordingly, using (83) and the union bound once again, we get

$$\mathbb{P}\left(\hat{\mathsf{K}}_1(\mathsf{X})\neq\mathsf{K}_\ell^\star\text{ for some }\ell\in[m]\right)$$

$$=\mathbb{P}\left[\bigcup_{\mathsf{K}\neq(\mathsf{K}_1^\star,\ldots,\mathsf{K}_m^\star)}\left\{\mathcal{S}(\mathsf{K})>\max_{\ell\in[m]}\mathcal{S}(\mathsf{K}_\ell^\star)\right\}\right] \tag{86}$$

$$=\mathbb{P}\left[\bigcup_{\mathsf{K}\neq(\mathsf{K}_1^\star,\ldots,\mathsf{K}_m^\star)}\{\mathsf{D}_1(\mathsf{K})\le0,\ldots,\mathsf{D}_m(\mathsf{K})\le0\}\right] \tag{87}$$

$$\le\sum_{\mathsf{K}\neq(\mathsf{K}_1^\star,\ldots,\mathsf{K}_m^\star)}\min_{\ell\in[m]}\mathbb{P}\left(\mathsf{D}_\ell(\mathsf{K})\le0\right) \tag{88}$$

$$\le\sum_{\mathsf{K}\neq(\mathsf{K}_1^\star,\ldots,\mathsf{K}_m^\star)}\exp\left[-\frac{\lambda^2\,k}{8}\right]\le n^2e^{-\frac{1}{8}\lambda^2\,k}, \tag{89}$$

where the last inequality is because $|\mathcal{K}^{\mathsf{con}}_{k,1,n}|\le n^2$. Thus, we see that if $\lambda^2>\frac{(24+\epsilon)\log n}{k}$, then $\mathbb{P}(\hat{\mathsf{K}}_1(\mathsf{X})\neq\mathsf{K}_\ell^\star\text{ for some }\ell\in[m])\le n^{-(1+\epsilon/8)}$. Using the same steps above, it is clear that,

$\mathbb{P}(\hat{\mathsf{K}}_\ell(\mathsf{X}) \neq \mathsf{K}^\star_\ell) \leq n^{-(1+\epsilon/8)}$, for any $2 \leq \ell \leq m$, provided that $\lambda^2 > \frac{(24+\epsilon)\log n}{k}$. Thus,

$$\mathbb{P}\left[\hat{\mathsf{K}}_{\mathsf{peel}} \neq \mathsf{K}^\star\right] = \mathbb{P}\left[\bigcup_{\ell=1}^{m} \hat{\mathsf{K}}_\ell \neq \mathsf{K}^\star_\ell\right] \leq \frac{m}{n^{(1+\epsilon/8)}} = n^{-\epsilon/8},$$

which converges to zero as $n \to \infty$.

## V. Conclusions and Future Work

In this paper, we study the computational and statistical boundaries of the submatrix and consecutive submatrix detection and recovery problems. For both models, we derive asymptotically tight lower and upper bounds on the thresholds for detection and recovery. To that end, for each problem, we propose statistically optimal and efficient algorithms for detection and recovery and analyze their performance. Our statistical lower bounds are based on classical techniques from information theory. Finally, we use the framework of low-degree polynomials to provide evidence for the statistical-computational gap in the submatrix detection problem.

There are several exciting directions for future work. First, it would be interesting to generalize our results to any pair of distributions $\mathcal{P}$ and $\mathcal{Q}$. While our information-theoretic lower bounds hold for general distributions, it is left to construct and analyze algorithms for this case, as well as to derive computational lower bounds. In our paper, we assume that the elements inside the planted submatrices are i.i.d., however, it is of practical interest to generalize this assumption and consider the case of dependent entries, e.g., Gaussians with a general covariance matrix. For example, this is the typical statistical model of cryo-EM data [25]. Finally, it will be interesting to prove a computational lower bound for the submatrix recovery problem using the recent framework of low-degree polynomials for recovery [50], and well as providing other forms of evidence to the statistical computational gaps for the submatrix detection problem with a growing number of planted submatrices, e.g., using average-case reductions [16].

## References

[1] A. A. Shabalin et al., "Finding large average submatrices in high dimensional data," *Ann. Appl. Statist.*, vol. 3, no. 3, pp. 985–1012, 2009.

[2] M. Kolar, S. Balakrishnan, A. Rinaldo, and A. Singh, "Minimax localization of structural information in large noisy matrices," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2011, pp. 909–917.

[3] S. Balakrishnan, M. Kolar, A. Rinaldo, A. Singh, and L. Wasserman, "Statistical and computational tradeoffs in biclustering," in *Proc. NIPS Workshop Comput. Trade-Offs Stat. Learn.*, 2011, pp. 1–4.

[4] C. Butucea and Y. I. Ingster, "Detection of a sparse submatrix of a high-dimensional noisy matrix," *Bernoulli*, vol. 19, no. 5B, pp. 2652–2688, 2013.

[5] E. Arias-Castro and N. Verzelen, "Community detection in dense random networks," *Ann. Statist.*, vol. 42, no. 3, pp. 940–969, 2014.

[6] B. Hajek, Y. Wu, and J. Xu, "Computational lower bounds for community detection on random graphs," in *Proc. 28th Conf. Learn. Theory*, 2015, pp. 899–928.

[7] A. Montanari, D. Reichman, and O. Zeitouni, "On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 217–225.

[8] N. Verzelen and E. Arias-Castro, "Community detection in sparse random networks," *Ann. Appl. Probability*, vol. 25, no. 6, pp. 3465–3510, 2015.

[9] Z. Ma and Y. Wu, "Computational barriers in minimax submatrix detection," *Ann. Statist.*, vol. 43, no. 3, pp. 1089–1116, 2015.

[10] X. Sun and A. Nobel, "On the maximal size of large-average and ANOVA-fit submatrices in a Gaussian random matrix," *Bernoulli*, vol. 19, pp. 275–294, Feb. 2013.

[11] E. Arias-Castro, E. Candés, and A. Durand, "Detection of an anomalous cluster in a network," *Ann. Statist.*, vol. 39, pp. 278–304, Jan. 2010.

[12] S. Bhamidi, P. Dey, and A. Nobel, "Energy landscape for large average submatrix detection problems in Gaussian random matrices," *Probability Theory Related Fields*, vol. 168, pp. 919–983, Aug. 2017.

[13] Y. Chen and J. Xu, "Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices," *J. Mach. Learn. Res.*, vol. 17, no. 27, pp. 1–57, 2016.

[14] T. Cai, T. Liang, and A. Rakhlin, "Computational and statistical boundaries for submatrix localization in a large noisy matrix," *Ann. Statist.*, vol. 45, no. 4, pp. 1403–1430, Aug. 2017.

[15] W. Huleihel, "Inferring hidden structures in random graphs," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 8, pp. 855–867, 2022.

[16] M. Brennan, G. Bresler, and W. Huleihel, "Reducibility and computational lower bounds for problems with planted sparse structure," in *Proc. 31st Conf. Learn. Theory*, 2018, pp. 48–166.

[17] M. Brennan, G. Bresler, and W. Huleihel, "Universality of computational lower bounds for submatrix detection," in *Proc. 32nd Conf. Learn. Theory*, 2019, pp. 417–468.

[18] A. Montanari, "Finding one community in a sparse graph," *J. Stat. Phys.*, vol. 161, no. 2, pp. 273–299, 2015.

[19] U. O. Candogan and V. Chandrasekaran, "Finding planted subgraphs with few eigenvalues using the Schur–Horn relaxation," *SIAM J. Optim.*, vol. 28, no. 1, pp. 735–759, 2018.

[20] B. Hajek, Y. Wu, and J. Xu, "Achieving exact cluster recovery threshold via semidefinite programming," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2788–2797, May 2016.

[21] B. Hajek, Y. Wu, and J. Xu, "Information limits for recovering a hidden community," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 4729–4745, Aug. 2017.

[22] X.-C. Bai, G. McMullan, and S. H. Scheres, "How cryo-EM is revolutionizing structural biology," *Trends Biochem. Sci.*, vol. 40, no. 1, pp. 49–57, 2015.

[23] D. Lyumkis, "Challenges and opportunities in cryo-EM single-particle analysis," *J. Biol. Chem.*, vol. 294, no. 13, pp. 5181–5197, 2019.

[24] A. Singer, "Mathematics for cryo-electron microscopy," in *Proc. Int. Congr. Mathematicians*, 2018, pp. 3995–4014.

[25] T. Bendory, A. Bartesaghi, and A. Singer, "Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities," *IEEE Signal Process. Mag.*, vol. 37, no. 2, pp. 58–76, Mar. 2020.

[26] F. Wang et al., "DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM," *J. Struct. Biol.*, vol. 195, no. 3, pp. 325–336, 2016.

[27] A. Heimowitz, J. Andén, and A. Singer, "APPLE picker: Automatic particle picking, a low-effort cryo-EM framework," *J. Struct. Biol.*, vol. 204, no. 2, pp. 215–227, 2018.

[28] T. Bepler et al., "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs," *Nature Methods*, vol. 16, no. 11, pp. 1153–1160, 2019.

[29] A. Eldar, B. Landa, and Y. Shkolnisky, "KLT picker: Particle picking using data-driven optimal templates," *J. Struct. Biol.*, vol. 210, no. 2, 2020, Art. no. 107473.

[30] S. B. Hopkins and D. Steurer, "Efficient Bayesian estimation from few samples: Community detection and related problems," in *Proc. IEEE 58th Annu. Symp. Found. Comput. Sci.*, 2017, pp. 379–390.

[31] S. Hopkins, "Statistical inference and the sum of squares method," Ph.D. dissertation, Cornell Univ., Ithaca, NY, USA, 2018.

[32] A. S. Bandeira, D. Kunisky, and A. S. Wein, "Computational hardness of certifying bounds on constrained PCA problems," in *Proc. 11th Innov. Theor. Comput. Sci. Conf.*, 2020, pp. 78:1–78:29.

[33] Y. Cherapanamjeri, S. B. Hopkins, T. Kathuria, P. Raghavendra, and N. Tripuraneni, "Algorithms for heavy-tailed statistics: Regression, covariance estimation, and beyond," in *Proc. 52nd Annu. ACM SIGACT Symp. Theory Comput.*, 2020, pp. 601–609.

[34] T. Bendory, N. Boumal, W. Leeb, E. Levin, and A. Singer, "Toward single particle reconstruction without particle picking: Breaking the detection limit," *SIAM J. Imag. Sci.*, vol. 16, no. 2, pp. 886–910, 2023.

[35] M. Dadon, W. Huleihel, and T. Bendory, "Detection and recovery of hidden submatrices," 2023, *arXiv:2306.06643*.

[36] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, 1st ed. Berlin, Germany: Springer, 2008.

[37] B. Barak, S. B. Hopkins, J. Kelner, P. Kothari, A. Moitra, and A. Potechin, “A nearly tight sum-of-squares lower bound for the planted clique problem,” in *Proc. IEEE 57th Annu. Symp. Found. Comput. Sci.*, 2016, pp. 428–437.

[38] S. B. Hopkins, P. K. Kothari, A. Potechin, P. Raghavendra, T. Schramm, and D. Steurer, “The power of sum-of-squares for detecting hidden structures,” in *Proc. IEEE 58th Found. Comput. Sci.*, 2017, pp. 720–731.

[39] D. Kunisky, A. S. Wein, and A. S. Bandeira, “Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio,” in *Mathematical Analysis, its Applications and Computation*. Berlin, Germany: Springer, 2022, pp. 1–50.

[40] N. Boumal, T. Bendory, R. R. Lederman, and A. Singer, “Heterogeneous multireference alignment: A single pass approach,” in *Proc. IEEE 52nd Annu. Conf. Inf. Sci. Syst.*, 2018, pp. 1–6.

[41] A. S. Wein, “Statistical estimation in the presence of group actions,” Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2018.

[42] T. Bendory, O. Mickelin, and A. Singer, “Sparse multi-reference alignment: Sample complexity and computational hardness,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2022, pp. 8977–8981.

[43] T. Bendory, N. Boumal, W. Leeb, E. Levin, and A. Singer, “Toward single particle reconstruction without particle picking: Breaking the detection limit,” 2018, *arXiv:1810.00226*.

[44] D. Guo, S. Shamai, and S. Verdu, “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[45] Y. Wu and J. Xu, “Statistical problems with planted structures: Information-theoretical and computational limits,” in *Information-Theoretic Methods in Data Science*, M. R. D. Rodrigues and Y. C. Eldar, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2020.

[46] S. Kreymer and T. Bendory, “Two-dimensional multi-target detection: An autocorrelation analysis approach,” *IEEE Trans. Signal Process.*, vol. 70, pp. 835–849, 2022.

[47] S. Kreymer, A. Singer, and T. Bendory, “A stochastic approximate expectation-maximization for structure determination directly from cryo-EM micrographs,” 2023, *arXiv:2303.02157*.

[48] R. Henderson, “The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules,” *Quart. Rev. Biophys.*, vol. 28, no. 2, pp. 171–193, 1995.

[49] A. S. Bandeira, A. Perry, and A. S. Wein, “Notes on computational-to-statistical gaps: Predictions using statistical physics,” *Portugaliae Mathematica*, vol. 75, no. 2, pp. 159–186, 2018.

[50] T. Schramm and S. A. Wein, “Computational barriers to estimation from low-degree polynomials,” *Ann. Statist.*, vol. 50, pp. 1833–1858, Sep. 2022.