The Annals of Statistics
2022, Vol. 50, No. 1, 584–613
https://doi.org/10.1214/21-AOS2123

TENSOR CLUSTERING WITH PLANTED STRUCTURES: STATISTICAL OPTIMALITY AND COMPUTATIONAL LIMITS

BY YUETIAN LUO[1,*] AND ANRU R. ZHANG[1,2,†]

[1]*Department of Statistics, University of Wisconsin-Madison,* [*]*yluo86@wisc.edu*
[2]*Department of Biostatistics & Bioinformatics, Duke University,* [†]*anru.zhang@duke.edu*

This paper studies the statistical and computational limits of high-order clustering with planted structures. We focus on two clustering models, constant high-order clustering (CHC) and rank-one higher-order clustering (ROHC), and study the methods and theory for testing whether a cluster exists (detection) and identifying the support of cluster (recovery).

Specifically, we identify the sharp boundaries of signal-to-noise ratio for which CHC and ROHC detection/recovery are statistically possible. We also develop the tight computational thresholds: when the signal-to-noise ratio is below these thresholds, we prove that polynomial-time algorithms cannot solve these problems under the computational hardness conjectures of hypergraphic planted clique (HPC) detection and hypergraphic planted dense subgraph (HPDS) recovery. We also propose polynomial-time tensor algorithms that achieve reliable detection and recovery when the signal-to-noise ratio is above these thresholds. Both sparsity and tensor structures yield the computational barriers in high-order tensor clustering. The interplay between them results in significant differences between high-order tensor clustering and matrix clustering in literature in aspects of statistical and computational phase transition diagrams, algorithmic approaches, hardness conjecture, and proof techniques. To our best knowledge, we are the first to give a thorough characterization of the statistical and computational trade-off for such a double computational-barrier problem. Finally, we provide evidence for the computational hardness conjectures of HPC detection (via low-degree polynomial and Metropolis methods) and HPDS recovery (via low-degree polynomial method).

1. Introduction. The high-dimensional tensor data have been increasingly prevalent in many domains, such as genetics, social sciences, engineering. In a wide range of applications, unsupervised analysis, in particular the high-order clustering, can be applied to discover the hidden modules in these high-dimensional tensor data. For example, in microbiome studies, microbiome samples are often measured across multiple body sites from multiple subjects (Faust et al. (2012), Flores et al. (2014)), resulting in the three-way tensors with subjects, body sites, and bacteria taxa as three modes. It has been reported that multiple microbial taxa can coexist within or across multiple body sites and subjects can form different subpopulations (Faust et al. (2012)). Similar data structures can also be found in multi-tissue multi-individual gene expression data (Wang, Fischer and Song (2019)). Mathematically, these patterns correspond to high-order clusters, that is, the underlying multi-way block structures in the data tensor. We also refer readers to the recent survey (Henriques and Madeira (2019)) on high-order clustering in applications.

In the literature, a number of methods have been proposed for triclustering or high-order clustering of tensor data, such as divide and conquer (Li and Tuck (2009)), seed growth

Received August 2020; revised May 2021.
MSC2020 subject classifications. Primary 62H15; secondary 62C20.
Key words and phrases. Average-case complexity, high-order clustering, hypergraphic planted clique, hypergraphic planted dense subgraph, statistical-computational phase transition.

(Sim, Aung and Gopalkrishnan (2010)), stochastic approach (Amar et al. (2015)), exhaustive approaches (Jiang et al. (2004)), pattern-based approach (Ji, Tan and Tung (2006)), etc. However, the theoretical guarantees for those existing procedures are not well established to our best knowledge.

This paper aims to fill the void of theory in high-order clustering. Suppose we observe an $n_1 \times \cdots \times n_d$-dimensional order-$d$ tensor $\boldsymbol{\mathcal{Y}}$ that satisfies

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{Z}}, \tag{1}$$

where $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ is the underlying signal with planted structure and $\boldsymbol{\mathcal{Z}}$ is the noise that has i.i.d. standard normal distributed entries. Our goal is to detect or recover the "planted structure" of the signal $\boldsymbol{\mathcal{X}}$. The specific problems in this paper are listed below.

1.1. *Problem formulations.* First, we consider the signal tensor $\boldsymbol{\mathcal{X}}$ that contains a constant planted structure:

$$\begin{aligned} &\boldsymbol{\mathcal{X}} \in \mathscr{X}_{\text{CHC}}(\mathbf{k}, \mathbf{n}, \lambda), \\ &\mathscr{X}_{\text{CHC}}(\mathbf{k}, \mathbf{n}, \lambda) = \{\lambda' \mathbf{1}_{I_1} \circ \cdots \circ \mathbf{1}_{I_d} : I_i \subseteq [n_i], |I_i| = k_i, \lambda' \geq \lambda\}. \end{aligned} \tag{2}$$

Here, "$\circ$" denotes the vector outer product, $\mathbf{1}_{I_i}$ is the n_i-dimensional indicator vector such that $(\mathbf{1}_{I_i})_j = 1$ if $j \in I_i$ and $(\mathbf{1}_{I_i})_j = 0$ if $j \notin I_i$; λ represents the signal strength. We collectively denote $\mathbf{k} = (k_1, \dots, k_d)$ and $\mathbf{n} = (n_1, \dots, n_d)$ for convenience. The support of the planted structure of $\boldsymbol{\mathcal{X}}$ is denoted as $\mathcal{S}(\boldsymbol{\mathcal{X}}) := (I_1, \dots, I_d)$. We refer to this model (1)(2) as the *constant high-order clustering (CHC)*. The constant planted clustering model in tensor or matrix biclustering (BC) data has been considered in a number of recent literature (see, e.g., Brennan, Bresler and Huleihel (2018), Brennan, Bresler and Huleihel (2019), Butucea and Ingster (2013), Butucea, Ingster and Suslina (2015), Cai, Liang and Rakhlin (2017), Chen and Xu (2016), Chi, Allen and Baraniuk (2017), Kolar et al. (2011), Sun and Nobel (2013), Xia and Zhou (2019)).

We also consider a more general setting that $\boldsymbol{\mathcal{X}}$ contains a rank-one planted structure:

$$\boldsymbol{\mathcal{X}} \in \mathscr{X}_{\text{ROHC}}(\mathbf{k}, \mathbf{n}, \mu), \quad \mathscr{X}_{\text{ROHC}}(\mathbf{k}, \mathbf{n}, \mu) = \{\mu' \mathbf{v}_1 \circ \cdots \circ \mathbf{v}_d : \mathbf{v}_i \in \mathcal{V}_{n_i,k_i}, \mu' \geq \mu\}, \tag{3}$$

where

$$\mathcal{V}_{n,k} := \{\mathbf{v} \in \mathbb{S}^{n-1} : \|\mathbf{v}\|_0 \leq k \text{ and } k^{-1/2} \leq |\mathbf{v}_i| \leq Ck^{-1/2} \text{ for } i \in S(\mathbf{v})\}, \quad C > 1$$

is the set of all k-sparse unit vectors with near-uniform magnitude. Here $S(\mathbf{v})$ denotes the support of the vector $\mathbf{v}$ and its formal definition is given in Section 2. Throughout the paper, we refer to the model in (1)(3) as the *rank-one high-order clustering (ROHC)*. Especially if $d = 2$, that is, in the matrix case, this model (rank-one submatrix (ROS)) was considered in Brennan, Bresler and Huleihel (2018), Busygin, Prokopyev and Pardalos (2008), Madeira and Oliveira (2004), Sun and Nobel (2013). For both models, we hope to answer the following questions on detection ($\mathcal{P}_D$) and recovery ($\mathcal{P}_R$):

$\mathcal{P}_D$ *When we can **detect** if any high-order cluster exists and when such conclusion cannot be made.* To be specific, consider the following hypothesis tests:

$$\begin{aligned} &\text{CHC}_D(\mathbf{n}, \mathbf{k}, \lambda): \quad H_0: \boldsymbol{\mathcal{X}} = \mathbf{0} \quad \text{v.s.} \quad H_1: \boldsymbol{\mathcal{X}} \in \mathscr{X}_{\text{CHC}}(\mathbf{k}, \mathbf{n}, \lambda), \\ &\text{ROHC}_D(\mathbf{n}, \mathbf{k}, \mu): \quad H_0: \boldsymbol{\mathcal{X}} = \mathbf{0} \quad \text{v.s.} \quad H_1: \boldsymbol{\mathcal{X}} \in \mathscr{X}_{\text{ROHC}}(\mathbf{k}, \mathbf{n}, \mu), \end{aligned} \tag{4}$$

we ask when there is a sequence of algorithms that can achieve reliable detection, that is, both type-I and II errors tend to zero.

$\mathcal{P}_R$ *How to* ***recover*** *the support of the cluster when it exists.* Specifically, we assume H_1 holds and aim to develop an algorithm that recovers the support $S(\boldsymbol{\mathcal{X}})$ based on the observation of $\boldsymbol{\mathcal{Y}}$. Denote the CHC and ROHC recovery problems considered in this paper as $\mathrm{CHC}_R(\mathbf{n}, \mathbf{k}, \lambda)$ and $\mathrm{ROHC}_R(\mathbf{n}, \mathbf{k}, \mu)$, respectively. We would like to know when there exists a sequence of algorithms that can achieve reliable recovery, that is, the probability of correctly recovering $S(\boldsymbol{\mathcal{X}})$ tends to one.

We study the performance of both *unconstrained-time algorithms* and *polynomial-time algorithms* for both detection $\mathcal{P}_D$ and recovery $\mathcal{P}_R$. The class of unconstrained algorithms includes all procedures with unlimited computational resources, while an algorithm that runs in polynomial-time has access to $\mathrm{poly}(n)$ independent random bits and must finish in $\mathrm{poly}(n)$ time, where n is the size of input. For convenience of exposition, we assume the explicit expressions can be exactly computed and $N(0, 1)$ random variable can be sampled in $O(1)$ time.

1.2. *Main results.* In this paper, we give a comprehensive characterization of the statistical and computational limits of the detection and recovery for both CHC and ROHC models. Denote $n := \max_i n_i$, $k := \max_i k_i$, and assume d is fixed. For technical convenience, our discussions are based on two asymptotic regimes:

$$\text{(A1)} \qquad \forall i \in [d], \quad n_i \to \infty, \qquad k_i \to \infty \quad \text{and} \quad k_i/n_i \to 0;$$

$$\text{(A2)} \qquad \text{or} \qquad \begin{aligned} &\text{for fixed } 0 \le \alpha \le 1, \beta \in \mathbb{R}, n \to \infty, \quad n_1 = \cdots = n_d = \tilde{\Theta}(n), \\ &k = k_1 = \cdots = k_d = \tilde{\Theta}(n^{\alpha}), \qquad \lambda = \tilde{\Theta}(n^{-\beta}), \qquad \mu/\sqrt{k^d} = \tilde{\Theta}(n^{-\beta}). \end{aligned}$$

In (A2), α and β represent the sparsity level and the signal strength of the cluster, respectively. The cluster becomes sparser as α decreases and the signal becomes stronger as β decreases. A rescaling of μ in (A2) is to make the magnitude of normalized entries in cluster of ROHC to be approximately one, which enables a valid comparison between the computational hardness of CHC and ROHC.

The following informal statements summarize the main results of this paper.

THEOREM 1 (Informal: Phase transitions in CHC). *Define*

$$\begin{aligned} &\beta^s_{\mathrm{CHC}_D} := (d\alpha - d/2) \vee (d-1)\alpha/2, \qquad \beta^s_{\mathrm{CHC}_R} := (d-1)\alpha/2, \\ &\beta^c_{\mathrm{CHC}_D} := (d\alpha - d/2) \vee 0, \qquad \beta^c_{\mathrm{CHC}_R} := \big((d-1)\alpha - (d-1)/2\big) \vee 0. \end{aligned} \tag{5}$$

Under the asymptotic regime (A2), *the statistical and computational limits of* $\mathrm{CHC}_D(\mathbf{k}, \mathbf{n}, \lambda)$ *and* $\mathrm{CHC}_R(\mathbf{k}, \mathbf{n}, \lambda)$ *exhibit the following phase transitions*:

- *CHC Detection*:
  - (i) $\beta > \beta^s_{\mathrm{CHC}_D}$: *reliable detection is information-theoretically impossible.*
  - (ii) $\beta^c_{\mathrm{CHC}_D} < \beta < \beta^s_{\mathrm{CHC}_D}$: *the computational inefficient test* $\psi^s_{\mathrm{CHC}_D}$ *in Section* 4.1 *succeeds, but polynomial-time reliable detection is impossible based on the hypergraphic planted clique* (*HPC*) *conjecture* (*Conjecture* 1).
  - (iii) $\beta < \beta^c_{\mathrm{CHC}_D}$: *the polynomial-time test* $\psi^c_{\mathrm{CHC}_D}$ *in Section* 4.2 *based on combination of sum and max statistics succeeds.*
- *CHC Recovery*:
  - (i) $\beta > \beta^s_{\mathrm{CHC}_R}$: *reliable recovery is information-theoretically impossible.*
  - (ii) $\beta^c_{\mathrm{CHC}_D} < \beta < \beta^s_{\mathrm{CHC}_R}$: *the exhaustive search* (*Algorithm* 1) *succeeds, but polynomial-time reliable recovery is impossible based on HPC conjecture* (*Conjecture* 1) *and hypergraphic planted dense subgraph* (*HPDS*) *recovery conjecture* (*Conjecture* 2).

TABLE 1

Phase transition and algorithms for detection and recovery in CHC and ROHC under the asymptotic regime (A2). Here, easy, hard, and impossible mean polynomial-time solvable, unconstrained-time solvable but polynomial-time unsolvable, and unconstrained-time unsolvable, respectively

	CHC_D	CHC_R	ROHC_D & ROHC_R
Impossible	$\lambda^2 \ll \frac{n^d}{k^{2d}} \wedge \frac{1}{k^{d-1}}$	$\lambda^2 \ll \frac{1}{k^{d-1}}$	$\frac{\mu^2}{k^d} \ll \frac{1}{k^{d-1}}$
Hard	$\frac{n^d}{k^{2d}} \wedge \frac{1}{k^{d-1}} \lesssim \lambda^2 \ll \frac{n^d}{k^{2d}} \wedge 1$	$\frac{1}{k^{d-1}} \lesssim \lambda^2 \ll \frac{n^{d-1}}{k^{2(d-1)}} \wedge 1$	$\frac{1}{k^{d-1}} \lesssim \frac{\mu^2}{k^d} \ll \frac{n^{d/2}}{k^d} \wedge 1$
Algorithms	$\psi^s_{\mathrm{CHC}_D}$	Alg 1	$\psi^s_{\mathrm{ROHC}_D}$ & Alg 2
Easy	$\lambda^2 \gtrsim \frac{n^d}{k^{2d}} \wedge 1$	$\lambda^2 \gtrsim \frac{n^{d-1}}{k^{2(d-1)}} \wedge 1$	$\frac{\mu^2}{k^d} \gtrsim \frac{n^{d/2}}{k^d} \wedge 1$
Algorithms	$\psi^c_{\mathrm{CHC}_D}$	Algs 3 and 5	$\psi^c_{\mathrm{ROHC}_D}$ & Algs 3 and 4

(iii) $\beta < \beta^c_{\mathrm{CHC}_D}$: *the combination of polynomial-time Algorithms 3 and 5 succeeds.*

THEOREM 2 (Informal: Phase transitions in ROHC). *Define*

$$\begin{aligned} \beta^s_{\mathrm{ROHC}} = \beta^s_{\mathrm{ROHC}_D} = \beta^s_{\mathrm{ROHC}_R} &:= (d-1)\alpha/2, \\ \beta^c_{\mathrm{ROHC}} = \beta^c_{\mathrm{ROHC}_D} = \beta^c_{\mathrm{ROHC}_R} &:= (\alpha d/2 - d/4) \vee 0. \end{aligned} \tag{6}$$

Under the asymptotic regime (A2), the statistical and computational limits of $\mathrm{ROHC}_D(\mathbf{k}, \mathbf{n}, \mu)$ *and* $\mathrm{ROHC}_R(\mathbf{k}, \mathbf{n}, \mu)$ *exhibit the following phase transitions*:

(i) $\beta > \beta^s_{\mathrm{ROHC}}$: *reliable detection and recovery are information-theoretically impossible.*

(ii) $\beta^c_{\mathrm{ROHC}} < \beta < \beta^s_{\mathrm{ROHC}}$: *the computational inefficient test* $\psi^s_{\mathrm{ROHC}_D}$ *in Section 4.1 succeeds in detection and the search Algorithm 2 succeeds in recovery, but polynomial-time reliable detection and recovery are impossible based on the HPC conjecture (Conjecture 1).*

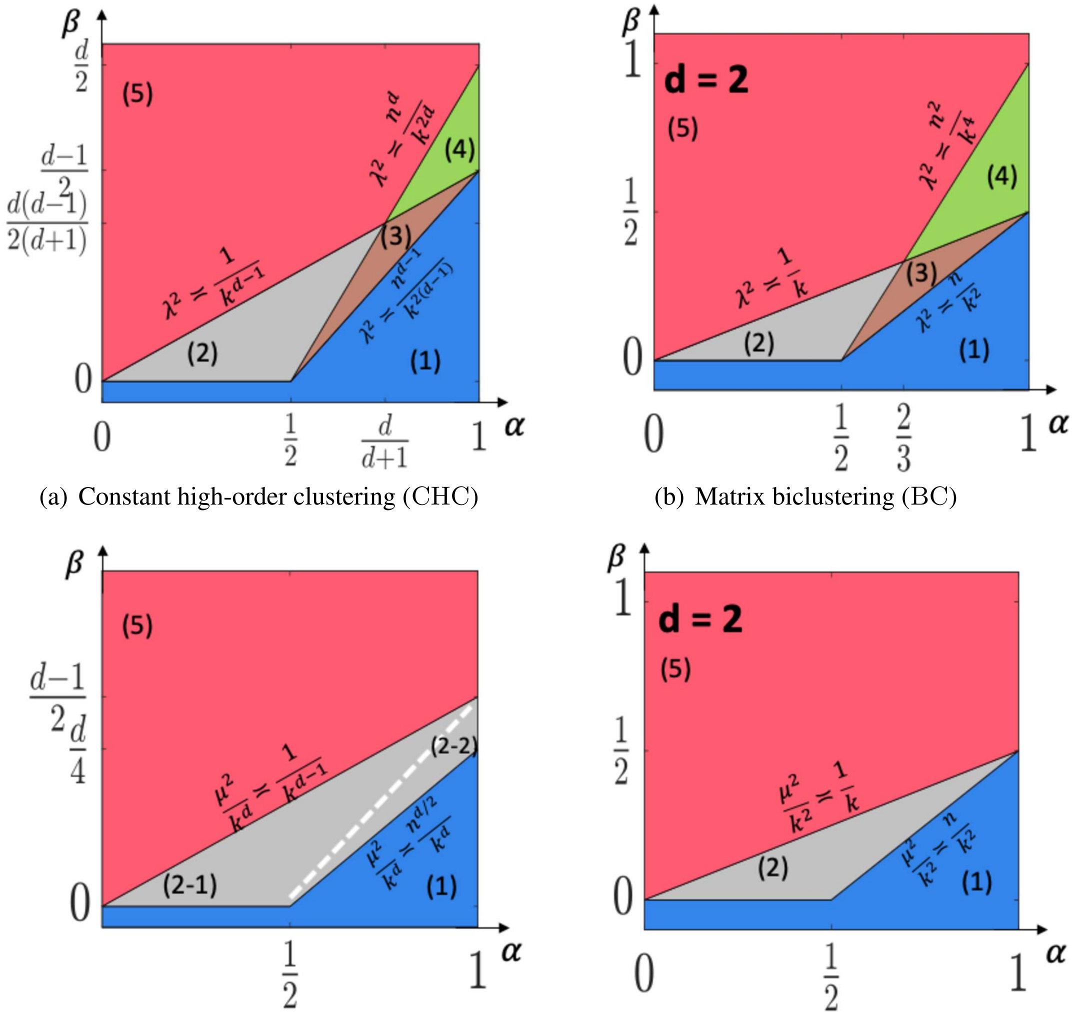
(iii) $\beta < \beta^c_{\mathrm{ROHC}}$: *the polynomial-time test* $\psi^c_{\mathrm{ROHC}_D}$ *in Section 4.2 succeeds in detection and the combination of polynomial-time Algorithms 3 and 4 succeeds in recovery.*

In Table 1, we summarize the statistical and computational limits in Theorems 1 and 2 in terms of the original parameters k, n, λ, μ and provide the corresponding algorithms that achieve these limits.

We also illustrate the phase transition diagrams for both CHC, ROHC ($d \geq 3$) in Figure 1, Panels (a) and (c). When $d = 2$, the phase transition diagrams in Panels (a) and (c) of Figure 1 reduce to constant biclustering (BC) diagram (Brennan, Bresler and Huleihel (2018), Cai, Liang and Rakhlin (2017), Chen and Xu (2016), Ma and Wu (2015)) and rank-one submatrix (ROS) diagram (Brennan, Bresler and Huleihel (2018)) in Panels (b) and (d) of Figure 1.

1.3. *Comparison with matrix clustering and our contributions.* The high-order ($d \geq 3$) clustering problems show many distinct aspects from their matrix counterparts ($d = 2$). We summarize the differences and highlight our contributions in the aspects of *phase transition diagrams*, *algorithms*, *hardness conjecture*, and *proof techniques* below.

(*Phase transition diagrams*) We can see the order-d ($d \geq 3$) tensor clustering has an additional regime: (2-2) in Figure 1 Panel (c). Specifically if $d = 2$, CHC_R, ROHC_R become BC_R, ROS_R that share the same computational limit and there is no gap between the statistical limit and computational efficiency for $\alpha = 1$ in ROS_R (see Panels (b) and (d), Figure 1). If $d \geq 3$, we need a strictly stronger signal-to-noise ratio to solve ROHC_R than CHC_R and there is always a gap between the statistical optimality and computational efficiency for ROHC_R.



(a) Constant high-order clustering (CHC) (b) Matrix biclustering (BC)

(c) Rank-one high-order clustering (ROHC) (d) Rank-one submatrix clustering (ROS)

FIG. 1. *Statistical and computational phase transition diagrams for constant high-order and rank-one high- -order* ($d \geq 3$) *clustering models* (*CHC and ROHC*) (*left two panels*) *and constant biclustering and rank-one submatrix* ($d = 2$) *clustering models* (*BC and ROS*) (*right two panels*) *under asymptotic regime* (A2). *Meaning of each region*: (1) *all problems detection and recovery both easy*; (2),(2-1),(2-2) *all problems detection hard and recovery hard*; (3) *CHC and BC detection easy and recovery hard*; (4) *CHC and BC detection easy and recovery impossible*; (5) *all problems detection and recovery impossible.*

This difference roots in two level computation barriers, sparsity and tensor structure, in high-order ($d \geq 3$) clustering. To our best knowledge, we are the first to characterize such double computational barriers.

(Algorithms) In addition, we develop new algorithms for high-order clustering. For CHC_R and ROHC_R, we introduce polynomial-time algorithms *Power-iteration* (Algorithm 4), *Aggregated-SVD* (Algorithm 5), both of which can be viewed as high-order analogues of the matrix spectral clustering. Also, see Section 1.4 for a comparison with the methods in the literature. We compare these algorithms and the *exhaustive search* (Algorithms 1 and 2) under the asymptotic regime (A2) in Figure 2. Compared to matrix clustering recovery diagram, that is, Figure 1(d), a new Regime (2) appears in the high-order ($d \geq 3$) clustering diagram. Different from the matrix clustering, where the polynomial-time spectral method reaches the computational limits for both BC_R and ROS_R when $\frac{1}{2} \leq \alpha \leq 1$, the optimal polynomial-time

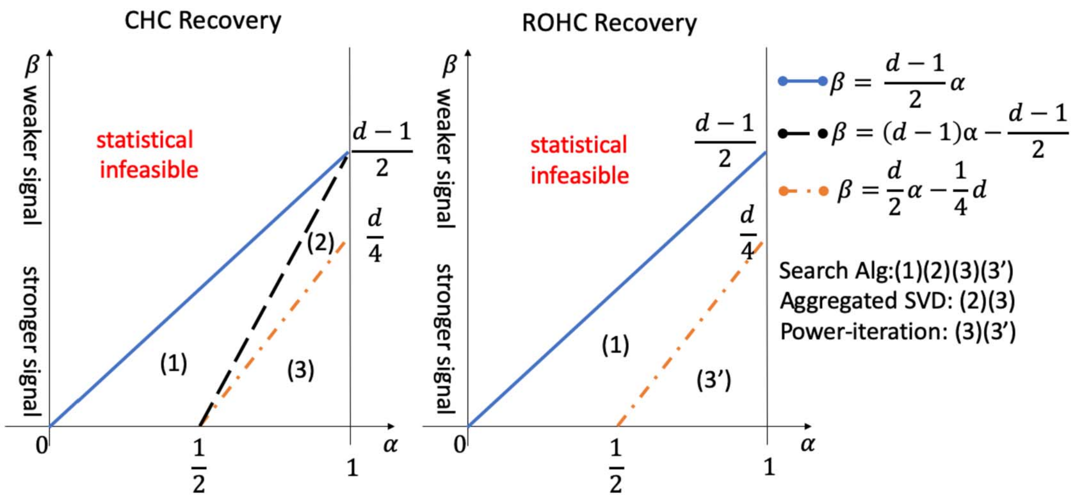


FIG. 2. CHC_R *and* ROHC_R *diagrams for Exhaustive search, Aggregated-SVD and Power-iteration algorithms under asymptotic regime* (A2). *In the right bottom corner, we provide the feasible signal-to-noise ratio regimes for each algorithm.*

algorithms for CHC_R and ROHC_R are distinct: Power-iteration is optimal for ROHC_R but is suboptimal for CHC_R; the Aggregated-SVD is optimal for CHC_R but does not apply for ROHC_R. This difference stems from the unique tensor algebraic structure in CHC.

(Hardness conjecture) We adopt the average-case reduction approach to establish the computational lower bounds. It would be ideal to do average-case reduction from the commonly raised conjectures, such as the planted clique (PC) detection or Boolean satisfiability (SAT), so that all of the hardness results of these well-studied conjectures can be inherited to the target problem. However, this route is complicated by the multiway structure in the high-order clustering. Instead, we apply a new average-case reduction scheme from hypergraphic planted clique (HPC) and the hypergraphic planted dense subgraph (HPDS) since HPC and HPDS have a more natural tensor structure that enables a more straightforward average-case reduction. Despite the widely studied planted clique (PC) and planted dense subgraph (PDS) in literature, the HPC and HPDS are far less understood and so are their computational hardness conjectures. The relationship between the computational hardness of PC and HPC remains an open problem (Luo and Zhang (2020a)). This paper is among the first to explore the average computational complexities of HPC and HPDS. To provide evidence for the computation hardness conjecture, we show a class of powerful algorithms, including the polynomial-time low-degree polynomials and Metropolis algorithms, are not able to solve HPC detection unless the clique size is sufficiently large. Also, we show low-degree polynomial method only succeeds in HPDS recovery in a restricted parameter regime. These results on HPC and HPDS may be of independent interests in analyzing average-case computational complexity, given the steadily increasing popularity on tensor data analysis recently and the commonly observed statistical-computational gaps therein (Barak and Moitra (2016), Dudeja and Hsu (2021), Hopkins, Shi and Steurer (2015), Lesieur et al. (2017), Perry, Wein and Bandeira (2020), Richard and Montanari (2014), Wein, El Alaoui and Moore ([2019] ©2019), Zhang and Xia (2018)).

(Proof techniques) The theoretical analysis in high-order clustering incorporates sparsity, low-rankness, and tensor algebra simultaneously, which is significantly more challenging than its counterpart in matrix clustering. Specifically, to prove the statistical lower bound of ROHC_D, we introduce the new Lemma 6, which gives an upper bound for the moment generating function of any power of a symmetric random walk on $\mathbb{Z}$ stopped after a hypergeometric distributed number of steps. This lemma is proved by utilizing Hoeffding's inequality

and the tail bound integration, which is different from the literature and can be of independent interest. To prove the statistical lower bound of CHC_D, we introduce a new technique that "sequentially decompose event" to bound the second moment of the truncated likelihood ratio (see Lemma 3). To prove the computational lower bounds, we introduce new average-case reduction schemes from HPC and HPDS, including a new reduction technique of *tensor reflection cloning* (Algorithm 7). This technique spreads the signal in the planted high-order cluster along each mode evenly, maintains the independence of entries in the tensor, and only mildly reduces the signal magnitude.

1.4. *Related literature.* This work is related to a wide range of literature on biclustering, tensor decomposition, tensor SVD, and theory of computation. When the order of the observation $d = 2$, the problem (1) reduces to the matrix clustering (Ames and Vavasis (2011), Busygin, Prokopyev and Pardalos (2008), Butucea, Ingster and Suslina (2015), Chi, Allen and Baraniuk (2017), Mankad and Michailidis (2014), Tanay, Sharan and Shamir (2002)). The statistical and computational limits of matrix clustering have been extensively studied in the literature (Balakrishnan et al. (2011), Brennan, Bresler and Huleihel (2018), Brennan, Bresler and Huleihel (2019), Butucea and Ingster (2013), Cai, Liang and Rakhlin (2017), Chen and Xu (2016), Kolar et al. (2011), Ma and Wu (2015), Schramm and Wein (2020)). As discussed in Section 1.3, the high-order ($d \geq 3$) tensor clustering exhibits significant differences from the matrix problems in various aspects.

Another related topic is on tensor decomposition and best low-rank tensor approximation. Although the best low-rank matrix approximation can be efficiently solved by the matrix singular value decomposition (Eckart–Young–Mirsky theorem), the best low-rank tensor approximation is NP-hard to calculate in general (Hillar and Lim (2013)). Various polynomial-time algorithms, which can be seen as the polynomial-time relaxations of the best low-rank tensor approximation, have been proposed in the literature, including the Newton method (Zhang and Golub (2001)), alternating minimization (Richard and Montanari (2014), Zhang and Golub (2001)), high-order singular value decomposition (De Lathauwer, De Moor and Vandewalle (2000a)), high-order orthogonal iteration (De Lathauwer, De Moor and Vandewalle (2000b)), k-means power iteration (Anandkumar, Ge and Janzamin (2014), Sun et al. (2017)), sparse high-order singular value decomposition (Zhang and Han (2019)), regularized gradient descent (Han, Willett and Zhang (2020)), etc. The readers are referred to surveys Cichocki et al. (2015), Kolda and Bader (2009). Departing from most of these previous results, the high-order clustering considered this paper involves both sparsity and low-rankness structures, which requires new methods and theoretical analysis as discussed in Section 1.3.

Our work is also related to a line of literature on average-case computational hardness and the statistical and computational trade-offs. The average-case reduction approach has been commonly used to show computational lower bounds for many recent high-dimensional problems, such as testing k-wise independence (Alon et al. (2007)), biclustering (Cai, Liang and Rakhlin (2017), Ma and Wu (2015), Cai and Wu (2020)), community detection (Hajek, Wu and Xu (2015)), RIP certification (Wang, Berthet and Plan (2016), Koiran and Zouzias (2014)), matrix completion (Chen (2015)), sparse PCA (Berthet and Rigollet (2013a), Berthet and Rigollet (2013b), Brennan and Bresler (2019a), Brennan, Bresler and Huleihel (2018), Gao, Ma and Zhou (2017), Wang, Berthet and Samworth (2016)), universal submatrix detection (Brennan, Bresler and Huleihel (2019)), sparse mixture and robust estimation (Brennan and Bresler (2019b)), a financial model with asymmetry information (Arora et al. (2011)), finding dense common subgraphs (Charikar, Naamad and Wu (2018)), graph logistic regression (Berthet and Baldin (2020)), online local learning (Awasthi et al. (2015)). See also a web of average-case reduction to a number of problems in Brennan and Bresler (2020), Brennan, Bresler and Huleihel (2018) and a recent survey (Wu and Xu (2021)). The average-case reduction is delicate, requiring that a distribution over instances in a conjecturally hard problem be mapped precisely to the target distribution. For this reason, many recent literature turn to show computational hardness results under the restricted models of computation, such as sum of squares (Barak et al. (2019), Hopkins et al. (2017), Ma and Wigderson (2015)), statistical query (Diakonikolas, Kane and Stewart (2017), Diakonikolas, Kong and Stewart (2019), Fan et al. (2018), Feldman, Perkins and Vempala (2018), Feldman et al. (2017), Kannan and Vempala (2017), Wang, Gu and Liu (2015)), class of circuit (Rossman (2008), Rossman (2014)), convex relaxation (Chandrasekaran and Jordan (2013)), local algorithms (Gamarnik and Sudan (2014)), meta-algorithms based on low-degree polynomials (Hopkins and Steurer (2017), Kunisky, Wein and Bandeira (2019)) and others. As discussed in Section 1.3, this paper is among the first to investigate the hypergraphic planted clique (HPC) and hypergraphic planted dense subgraph (HPDS) problems and their computational hardness. We perform new average-case reduction scheme from these conjectures and develop the computational lower bounds for CHC and ROHC.

1.5. *Organization.* The rest of this article is organized as follows. After a brief introduction of notation and preliminaries in Section 2, the statistical limits of high-order cluster recovery and detection are given in Sections 3 and 4, respectively. In Section 5, we establish the computational limits of high-order clustering, along with the hypergraphic planted clique (HPC) and hypergraphic planted dense subgraph (HPDS) models, computational hardness conjectures, and evidence. Discussion and future work are given in Section 6. The technical proofs are collected in Supplementary Material Luo and Zhang (2022).

2. Notation and definitions. The following notation will be used throughout this article. For any nonnegative integer n, let $[n] = \{1, \dots, n\}$. The lowercase letters (e.g., a, b), lowercase boldface letters (e.g., $\mathbf{u}$, $\mathbf{v}$), uppercase boldface letters (e.g., $\mathbf{A}$, $\mathbf{U}$), and boldface calligraphic letters (e.g., $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{X}}$) are used to denote scalars, vectors, matrices, and order-3-or-higher tensors respectively. For any two series of numbers, say $\{a_n\}$ and $\{b_n\}$, denote $a \asymp b$ if there exist uniform constants $c, C > 0$ such that $ca_n \le b_n \le Ca_n, \forall n$; and $a = \Omega(b)$ if there exists uniform constant $c > 0$ such that $a_n \ge cb_n, \forall n$. The notation $a = \tilde{\Theta}(b)$ and $a \gg b$ mean $\lim_{n\to\infty} a_n/n = \lim_{n\to\infty} b_n/n$ and $\lim_{n\to\infty}\log(a_n/n) > \lim_{n\to\infty}\log(b_n/n)$, respectively. $a \lesssim b$ means $a \le b$ up to polylogarithmic factors in n. We use bracket subscripts to denote subvectors, submatrices, and subtensors. For example, $\mathbf{v}_{[2:r]}$ is the vector with the 2nd to rth entries of $\mathbf{v}$; $\mathbf{D}_{[(r+1):n_1,:]}$ contains the $(r+1)$th to the n_1th rows of $\mathbf{D}$; $\boldsymbol{\mathcal{A}}_{[1:s_1,1:s_2,1:s_3]}$ is the s_1-by-s_2-by-s_3 subtensor of $\boldsymbol{\mathcal{A}}$ with index set $\{(i_1, i_2, i_3) : 1 \le i_1 \le s_1, 1 \le i_2 \le s_2, 1 \le i_3 \le s_3\}$. For any vector $\mathbf{v} \in \mathbb{R}^{n_1}$, define its ℓ_2 norm as $\|\mathbf{v}\|_2 = (\sum_i |\mathbf{v}_i|^2)^{1/2}$ and $\|\mathbf{v}\|_0$ is defined to be the number of nonzero entries in $\mathbf{v}$. Given vectors $\{\mathbf{v}_i\}_{i=1}^d \in \mathbb{R}^{n_i}$, the outer product $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times\cdots\times n_d} = \mathbf{v}_1 \circ \cdots \circ \mathbf{v}_d$ is defined such that $\boldsymbol{\mathcal{A}}_{[i_1,\dots,i_d]} = (\mathbf{v}_1)_{i_1} \cdots (\mathbf{v}_d)_{i_d}$. For any event A, let $\mathbb{P}(A)$ be the probability that A occurs.

For any order-d tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times\cdots\times n_d}$. The matricization $\mathcal{M}(\cdot)$ is the operation that unfolds or flattens the order-d tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ into the matrix $\mathcal{M}_z(\boldsymbol{\mathcal{A}}) \in \mathbb{R}^{n_z \times \prod_{j\neq z} n_j}$ for $z = 1, \dots, d$. Specifically, the mode-z matricization of $\boldsymbol{\mathcal{A}}$ is formally defined as

$$\boldsymbol{\mathcal{A}}_{[i_1,\dots,i_d]} = \big(\mathcal{M}_z(\boldsymbol{\mathcal{A}})\big)_{[i_z,j]}, \qquad j = 1 + \sum_{\substack{l=1\\ l\neq z}}^{d} \left\{(i_l - 1) \prod_{\substack{m=1\\ m\neq z}}^{l-1} n_m\right\}$$

for any $1 \le i_l \le n_l$, $l = 1, \dots, d$. Also see Kolda and Bader ((2009), Section 2.4) for more discussions on tensor matricizations. The mode-z product of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ with a matrix $\mathbf{U} \in \mathbb{R}^{k_z\times n_z}$ is denoted by $\boldsymbol{\mathcal{A}} \times_z \mathbf{U}$ and is of size $n_1 \times \cdots \times n_{z-1} \times k_z \times n_{z+1} \times \cdots \times n_d$, such that

$$(\boldsymbol{\mathcal{A}} \times_z \mathbf{U})_{[i_1,\dots,i_{z-1},j,i_{z+1},\dots,i_d]} = \sum_{i_z=1}^{n_z} \boldsymbol{\mathcal{A}}_{[i_1,i_2,\dots,i_d]} \mathbf{U}_{[j,i_z]}.$$

For any two distinct $k_1, k_2 \in [d] (k_1 < k_2)$ and $j_1 \in [n_{k_1}]$ and $j_2 \in [n_{k_2}]$, we denote

$$\boldsymbol{\mathcal{A}}^{(k_1,k_2)}_{j_1,j_2} = \boldsymbol{\mathcal{A}}_{[:,\ldots,:,\underbrace{j_1}_{k_1\text{th index}},:,\ldots,:,\underbrace{j_2}_{k_2\text{th index}},:,\ldots,:]} \in \mathbb{R}^{n_1\times\cdots\times n_{k_1-1}\times n_{k_1+1}\times\cdots\times n_{k_2-1}\times n_{k_2+1}\times\cdots\times n_d}$$

as a subtensor of $\boldsymbol{\mathcal{A}}$. The support of an order-d tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ is denoted by $S(\boldsymbol{\mathcal{X}}) := (I_1, \ldots, I_d)$ where $I_j \in \mathbb{R}^{n_j}$ and $(I_j)_i$ equals to zero when $\mathcal{M}_j(\boldsymbol{\mathcal{X}})_{[i,:]}$ is a zero vector and equals to one otherwise. In particular, when the tensor order is 1, we simply have the support of a vector $\mathbf{v}$ is $S(\mathbf{v}) = \{j : \mathbf{v}_j \neq 0\}$.

Given a distribution Q, let $Q^{\otimes n}$ be the distribution of $(X_1, \ldots, X_n)$ if $\{X_i\}_{i=1}^n$ are i.i.d. copies of Q. Similarly, let $Q^{\otimes m\times n}$ and $Q^{\otimes (n^{\otimes d})}$ denote the distribution on $\mathbb{R}^{m\times n}$ and $\mathbb{R}^{n^{\otimes d}}$ with i.i.d. entries distributed as Q. Here $n^{\otimes d} := n \times n \times \cdots \times n$ denotes the order-d Cartesian product. In addition, we use C, C_1, C_2, c and other variations to represent the large and small constants, whose actual values may vary from line to line.

Next, we formally define the statistical and computational risks to quantify the fundamental limits of high-order clustering. First, we define the risk of testing procedure $\phi_D(\boldsymbol{\mathcal{Y}}) \in \{0, 1\}$ as the sum of Type-I and Type-II errors for detection problems CHC_D and ROHC_D:

$$\mathcal{E}_{\mathcal{P}_D}(\phi_D) = \mathbb{P}_0\big(\phi_D(\boldsymbol{\mathcal{Y}}) = 1\big) + \sup_{\substack{\boldsymbol{\mathcal{X}}\in\mathscr{X}_{\mathrm{CHC}}(\mathbf{k},\mathbf{n},\lambda)\\ (\text{or } \boldsymbol{\mathcal{X}}\in\mathscr{X}_{\mathrm{ROHC}}(\mathbf{k},\mathbf{n},\mu))}} \mathbb{P}_{\boldsymbol{\mathcal{X}}}\big(\phi_D(\boldsymbol{\mathcal{Y}}) = 0\big),$$

where $\mathbb{P}_0$ is the probability under H_0 and $\mathbb{P}_{\boldsymbol{\mathcal{X}}}$ is the probability under H_1 with the signal tensor $\boldsymbol{\mathcal{X}}$. We say $\{\phi_D\}_n$ reliably detect in $\mathcal{P}_D$ if $\lim_{n\to\infty} \mathcal{E}_{\mathcal{P}_D}(\phi_D) = 0$. Second, for recovery problem CHC_R and ROHC_R, define the recovery error for any recovery algorithm $\phi_R(\boldsymbol{\mathcal{Y}}) \in \{(I_1, \ldots, I_d) : I_i \subseteq \{1, \ldots, n_i\}\}$ as

$$\mathcal{E}_{\mathcal{P}_R}(\phi_R) = \sup_{\boldsymbol{\mathcal{X}}\in\mathscr{X}_{\mathrm{CHC}}(\mathbf{k},\mathbf{n},\lambda)} \mathbb{P}_{\boldsymbol{\mathcal{X}}}\big(\phi_R(\boldsymbol{\mathcal{Y}}) \neq S(\boldsymbol{\mathcal{X}})\big) \quad \text{or} \quad \sup_{\boldsymbol{\mathcal{X}}\in\mathscr{X}_{\mathrm{ROHC}}(\mathbf{k},\mathbf{n},\mu)} \mathbb{P}_{\boldsymbol{\mathcal{X}}}\big(\phi_R(\boldsymbol{\mathcal{Y}}) \neq S(\boldsymbol{\mathcal{X}})\big).$$

We say $\{\phi_R\}_n$ reliably recover in $\mathcal{P}_R$ if $\lim_{n\to\infty} \mathcal{E}_{\mathcal{P}_R}(\phi_R) = 0$. Third, denote AllAlg^D, AllAlg^R, $\mathrm{PolyAlg}^D$, $\mathrm{PolyAlg}^R$ as the collections of unconstrained-time algorithms and polynomial-time algorithms for detection and recover problems, respectively. Then we can define four different statistical and computational risks as follows:

$$\mathcal{E}^s_{\mathcal{P}_D} := \inf_{\phi_D\in\mathrm{AllAlg}^D} \mathcal{E}_{\mathcal{P}_D}(\phi_D), \qquad \mathcal{E}^c_{\mathcal{P}_D} := \inf_{\phi_D\in\mathrm{PolyAlg}^D} \mathcal{E}_{\mathcal{P}_D}(\phi_D),$$

$$\mathcal{E}^s_{\mathcal{P}_R} := \inf_{\phi_R\in\mathrm{AllAlg}^R} \mathcal{E}_{\mathcal{P}_R}(\phi_R), \qquad \mathcal{E}^c_{\mathcal{P}_R} := \inf_{\phi_R\in\mathrm{PolyAlg}^R} \mathcal{E}_{\mathcal{P}_R}(\phi_R).$$

3. High-order cluster recovery: Statistical limits and polynomial-time algorithms. This section studies the statistical limits of high-order cluster recovery. We first present the statistical lower bounds of λ and μ that guarantee reliable recovery, then we give unconstrained-time algorithms that achieves these lower bounds. We also propose computationally efficient algorithms, Thresholding Algorithm, Power-iteration, and Aggregated-SVD, with theoretical guarantees.

3.1. CHC_R *and* ROHC_R: *Statistical limits.* Recall (5) and (6), we first present the statistical lower bounds for reliable recovery of CHC_R and ROHC_R.

THEOREM 3 (Statistical lower bounds for CHC_R and ROHC_R). *Consider* $\mathrm{CHC}_R(\mathbf{k}, \mathbf{n}, \lambda)$ *and* $\mathrm{ROHC}_R(\mathbf{k}, \mathbf{n}, \mu)$. *Let* $0 < \eta < \frac{1}{8}$ *be fixed. Under the asymptotic regime* (A1), *if*

$$\lambda \le \max\left(\left\{\sqrt{\frac{\eta\log(n_i - k_i)}{\prod_{z=1,z\neq i}^d k_z}}\right\}_{i=1}^d\right) \quad \left(or\ \frac{\mu}{\sqrt{\prod_{i=1}^d k_i}} \le \max\left(\left\{\sqrt{\frac{\eta\log(n_i - k_i)}{\prod_{z=1,z\neq i}^d k_z}}\right\}_{i=1}^d\right)\right),$$

Algorithm 1 CHC_R Search

1: **Input**: $\mathcal{Y} \in \mathbb{R}^{n_1\times\cdots\times n_d}$, sparsity level $\mathbf{k} = (k_1, \ldots, k_d)$.
2: **Output**:

$$(\hat{I}_1, \ldots, \hat{I}_d) = \arg \max_{\substack{I_i \subseteq [n_i], |I_i| = k_i \\ i=1,\ldots,d}} \sum_{i_1 \in I_1} \cdots \sum_{i_d \in I_d} \mathcal{Y}_{[i_1,\ldots,i_d]}.$$

we have

$$\mathcal{E}^s_{\text{CHC}_R}(\text{ or } \mathcal{E}^s_{\text{ROHC}_R}) \geq \frac{\sqrt{M}}{1+\sqrt{M}}\left(1 - 2\eta - \frac{2\eta}{\log M}\right) \to 1 - 2\eta,$$

where $M = \max(\{n_i - k_i\}_{i=1}^d)$. *Moreover, under the asymptotic regime* (A2), *if* $\beta > \beta^s_{\text{CHC}_R}$ (*or* $\beta^s_{\text{ROHC}_R}$), *we have* $\mathcal{E}^s_{\text{CHC}_R}(\text{ or } \mathcal{E}^s_{\text{ROHC}_R}) \to 1 - 2\eta$.

We further propose the CHC_R Search (Algorithm 1) and ROHC_R Search (Algorithm 2) with the following theoretical guarantees. These algorithms exhaustively search all possible cluster positions and find one that best matches the data. In particular, Algorithm 1 is exactly the maximum likelihood estimator. It is note worthy in Algorithm 2, we generate $\mathcal{Z}_1$ with i.i.d. standard Gaussian entries and construct $\mathcal{A} = \frac{\mathcal{Y}+\mathcal{Z}_1}{\sqrt{2}}$ and $\mathcal{B} = \frac{\mathcal{Y}-\mathcal{Z}_1}{\sqrt{2}}$. In that case, $\mathcal{A}$ and $\mathcal{B}$ becomes two independent sample tensors, which facilitate the theoretical analysis. Such a scheme is mainly for technical convenience and not necessary in practice.

THEOREM 4 (Guarantee of CHC_R search). *Consider* $\text{CHC}_R(\mathbf{k}, \mathbf{n}, \lambda)$ *under the asymptotic regime* (A1). *There exists* $C_0 > 0$ *such that when* $\lambda \geq C_0\sqrt{\frac{\sum_{i=1}^d \log(n_i - k_i)}{\min_{1\leq i\leq d}\{\prod_{z=1,z\neq i}^d k_z\}}}$, *Algorithm* 1 *identifies the true support of* $\mathcal{X}$ *with probability at least* $1 - C\sum_{i=1}^d (n_i - k_i)^{-c}$ *for some* $c, C > 0$. *Moreover, under the asymptotic regime* (A2), *Algorithm* 1 *achieves the reliable recovery of* CHC_R *when* $\beta < \beta^s_{\text{CHC}_R}$.

Algorithm 2 ROHC_R Search

1: **Input:** $\mathcal{Y} \in \mathbb{R}^{n_1\times\cdots\times n_d}$, sparsity upper bound $\mathbf{k}$.
2: Sample $\mathcal{Z}_1 \sim N(0, 1)^{\otimes n_1\times\cdots\times n_d}$ and construct $\mathcal{A} = \frac{\mathcal{Y}+\mathcal{Z}_1}{\sqrt{2}}$ and $\mathcal{B} = \frac{\mathcal{Y}-\mathcal{Z}_1}{\sqrt{2}}$.
3: For each $\{\bar{k}_i\}_{i=1}^d$ satisfying $\bar{k}_i \in [1, k_i]$ $(1 \leq i \leq d)$ do:

(a) Compute

$$(\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_d) = \arg \max_{(\mathbf{u}_1,\ldots,\mathbf{u}_d)\in S^{n_1}_{\bar{k}_1}\times\cdots\times S^{n_d}_{\bar{k}_d}} \mathcal{A} \times_1 \mathbf{u}_1^\top \times \cdots \times_d \mathbf{u}_d^\top.$$

Here, S^n_t is the set of vectors $\mathbf{u} \in \{-1, 1, 0\}^n$ with exactly t nonzero entries.

(b) For each tuple $(\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_d)$ computed from Step (a), mark it if it satisfies

$$\left\{ j : \left(\mathcal{B} \times_1 \hat{\mathbf{u}}_1^\top \times \cdots \times_{i-1} \hat{\mathbf{u}}_{i-1}^\top \times_{i+1} \hat{\mathbf{u}}_{i+1}^\top \times \cdots \times_d \hat{\mathbf{u}}_d^\top\right)_j (\hat{\mathbf{u}}_i)_j \geq \frac{1}{2\sqrt{2}} \frac{\mu}{\sqrt{\prod_{i=1}^d k_i}} \prod_{z\neq i} \bar{k}_z \right\}$$

is exactly the support of $\hat{\mathbf{u}}_i$, $S(\hat{\mathbf{u}}_i)$ for all $1 \leq i \leq d$.

4: Among all marked tuples $(\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_d)$, we find the one, say $(\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_d)$, that maximizes $\sum_{i=1}^d |S(\hat{\mathbf{u}}_i)|$.
5: **Output:** $\hat{I}_i = S(\tilde{\mathbf{u}}_i)(1 \leq i \leq d)$.

Algorithm 3 CHC_R and ROHC_R Thresholding Algorithm

1: **Input**: $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$.
2: **Output**:

$$(\hat{I}_1, \dots, \hat{I}_d) = \{(i_1, \dots, i_d) : |\boldsymbol{\mathcal{Y}}_{[i_1,\dots,i_d]}| \geq \sqrt{2(d+1)\log n}\}.$$

THEOREM 5 (Guarantee of ROHC_R search). *Consider* $\mathrm{ROHC}_R(\mathbf{k}, \mathbf{n}, \mu)$ *under the asymptotic regime* (A1). *There is an absolute constant* $C_0 > 0$ *such that if* $\mu \geq C_0\sqrt{k \log n}$, *then Algorithm* 2 *identifies the true support of* $\boldsymbol{\mathcal{X}}$ *with probability at least* $1 - C\sum_{i=1}^{d}(n_i - k_i)^{-1}$ *for some constant* $C > 0$. *Moreover, under the asymptotic regime* (A2), *Algorithm* 2 *achieves the reliable recovery of* ROHC_R *when* $\beta < \beta^s_{\mathrm{ROHC}_R}$.

Combining Theorems 3, 4, and 5, we can see if $k_1 \asymp k_2 \asymp \cdots \asymp k_d$, Algorithms 1, 2 achieve the minimax statistical lower bounds for CHC_R, ROHC_R. On the other hand, Algorithms 1 and 2 are based on computationally inefficient exhaustive search. Next, we introduce the polynomial-time algorithms.

3.2. CHC_R *and* ROHC_R: *Polynomial-time algorithms.* The polynomial-time algorithms for solving CHC_R and ROHC_R rely on the sparsity level k_i $(1 \leq i \leq d)$. First, when $k \lesssim \sqrt{n}$ (sparse regime), we propose Thresholding Algorithm (Algorithm 3) that selects the high-order cluster based on the largest entry in absolute value from each tensor slice. The theoretical guarantee of this algorithm is given in Theorem 6.

THEOREM 6 (Guarantee of thresholding algorithm for CHC_R and ROHC_R). *Consider* $\mathrm{CHC}_R(\mathbf{k}, \mathbf{n}, \lambda)$ *and* $\mathrm{ROHC}_R(\mathbf{k}, \mathbf{n}, \mu)$. *If* $\lambda \geq 2\sqrt{2(d+1)\log n}$ (*or* $\mu/\sqrt{\prod_{i=1}^{d} k_i} \geq 2\sqrt{2(d+1)\log n}$), *Algorithm* 3 *exactly recovers the true support of* $\boldsymbol{\mathcal{X}}$ *with probability at least* $1 - O(n^{-1})$. *Moreover, under the asymptotic regime* (A2), *Algorithm* 3 *achieves the reliable recovery of* CHC_R *and* ROHC_R *when* $\beta < 0$.

Second, when $k \gtrsim \sqrt{n}$ (dense regime), we consider the Power-iteration given in Algorithm 4, which is a modification of the tensor PCA methods in the literature (Anandkumar, Ge and Janzamin (2014), Richard and Montanari (2014), Zhang and Xia (2018)) and can be seen as an tensor analogue of the matrix spectral clustering method.

We also propose another polynomial-time algorithm, *Aggregated SVD*, in Algorithm 5 for the dense regime of CHC_R. As its name suggests, the central idea is to first transform the tensor $\boldsymbol{\mathcal{Y}}$ into a matrix by taking average, then apply matrix SVD. Aggregated-SVD is in a similar vein of the hypergraph adjacency matrix construction in the hypergraph community recovery literature (Ghoshdastidar and Dukkipati (2017), Kim, Bandeira and Goemans (2017)).

We give guarantees of Power-iteration and Aggregated-SVD for high-order cluster recovery. In particular, Aggregated SVD achieves strictly better performance than Power-iteration in CHC_R, but does not apply for ROHC_R.

THEOREM 7 (Guarantee of power-iteration for CHC_R and ROHC_R). *Consider* $\mathrm{CHC}_R(\mathbf{k}, \mathbf{n}, \lambda)$ *and* $\mathrm{ROHC}_R(\mathbf{k}, \mathbf{n}, \mu)$. *Assume* $n_i \geq c_0 n$ $(1 \leq i \leq d)$ *for constant* $c_0 > 0$ *where* $n := \max_i n_i$. *Under the asymptotic regime* (A1), *there exists a uniform constant* $C_0 > 0$ *such that if* $\lambda\sqrt{\prod_{i=1}^{d} k_i} \geq C_0 n^{\frac{d}{4}}$ (*or* $\mu \geq C_0 n^{\frac{d}{4}}$)

$$\text{and} \quad t_{\max} \geq C \log\left(\frac{n}{\lambda\sqrt{\prod_{i=1}^{d} k_i}}\right) \vee C \quad (\text{or } t_{\max} \geq C(\log(n/\mu) \vee 1)),$$

Algorithm 4 Power-iteration for CHC_R and ROHC_R

1: **Input:** $\mathcal{Y} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, maximum number of iterations $t_{\max}$.

2: Sample $\mathcal{Z}_1 \sim N(0,1)^{\otimes n_1 \times \cdots \times n_d}$ and construct $\mathcal{A} = (\mathcal{Y} + \mathcal{Z}_1)/\sqrt{2}$ and $\mathcal{B} = (\mathcal{Y} - \mathcal{Z}_1)/\sqrt{2}$.

3: (Initiation) Set $t = 0$. For $i = 1 : d$, compute the top left singular vector of $\mathcal{M}_i(\mathcal{A})$ and denote it as $\hat{\mathbf{u}}_i^{(0)}$.

4: For $t = 1, \ldots, t_{\max}$, do

(a) For $i = 1$ to d, update

$$\hat{\mathbf{u}}_i^{(t)} = \mathrm{NORM}\big(\mathcal{A} \times_1 (\hat{\mathbf{u}}_1^{(t)})^\top \times \cdots \times_{i-1} (\hat{\mathbf{u}}_{i-1}^{(t)})^\top \times_{i+1} (\hat{\mathbf{u}}_{i+1}^{(t-1)})^\top \times \cdots \times_d (\hat{\mathbf{u}}_d^{(t-1)})^\top\big).$$

Here, $\mathrm{NORM}(\mathbf{v}) = \mathbf{v}/\|\mathbf{v}\|_2$ is the normalization of vector $\mathbf{v}$.

5: Let $(\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_d) := (\hat{\mathbf{u}}_1^{(t_{\max})}, \ldots, \hat{\mathbf{u}}_d^{(t_{\max})})$. For $i = 1, \ldots, d$, calculate

$$\mathbf{v}_i := \mathcal{B} \times_1 \hat{\mathbf{u}}_1^\top \times \cdots \times_{i-1} \hat{\mathbf{u}}_{i-1}^\top \times_{i+1} \hat{\mathbf{u}}_{i+1}^\top \times \cdots \times_d \hat{\mathbf{u}}_d^\top \in \mathbb{R}^{n_i}. \tag{7}$$

- If the problem is CHC_R, the component values of $\mathbf{v}_i$ form two clusters. Sort $\{(\mathbf{v}_i)_j\}_{j=1}^{n_i}$ and cut the sequence at the largest gap between the consecutive values. Let the index subsets of two parts be $\hat{I}_i$ and $[n_i] \setminus \hat{I}_i$. Output: $\hat{I}_i$
- If the problem is ROHC_R, the component values of $\mathbf{v}_i$ form three clusters. Sort the sequence $\{(\mathbf{v}_i)_j\}_{j=1}^{n_i}$, cut at the two largest gaps between the consecutive values, and form three parts. Among the three parts, pick the two smaller-sized ones, and let the index subsets of these two parts be $\hat{I}_i^1, \hat{I}_i^2$. Output: $\hat{I}_i = \hat{I}_i^1 \cup \hat{I}_i^2$

6: **Output:** $\{\hat{I}_i\}_{i=1}^d$.

Algorithm 4 *identifies the true support of* $\mathcal{X}$ *with probability at least* $1 - \sum_{i=1}^d n_i^{-c} - C\exp(-cn)$ *for constants* $c, C > 0$. *Moreover, under the asymptotic regime* (A2), *Algorithm* 4 *achieves the reliable recovery of* CHC_R *and* ROHC_R *when* $\beta < (\alpha - 1/2)d/2$.

Algorithm 5 Aggregated-SVD for CHC_R

1: **Input:** $\mathcal{Y} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$.

2: **For** $i = 1, 2, \ldots, d$, do:

(a) Find $i^* = \arg\min_{j \neq i} n_j$ and calculate $\mathbf{Y}^{(i,i^*)} \in \mathbb{R}^{n_i \times n_{i^*}}$ where $\mathbf{Y}^{(i,i^*)}_{[k_1,k_2]} := \mathrm{SUM}(\mathcal{Y}^{(i,i^*)}_{k_1,k_2})/\sqrt{\prod_{j=1, j\neq i,i^*}^d n_j}$ for $1 \le k_1 \le n_i$, $1 \le k_2 \le n_{i^*}$. Here $\mathrm{SUM}(\mathcal{A}) := \sum_{i_1} \cdots \sum_{i_d} \mathcal{A}_{[i_1,\ldots,i_d]}$ and $\mathcal{Y}^{(i,i^*)}_{k_1,k_2}$ is the subtensor of $\mathcal{Y}$ defined in Section 2.

(b) Sample $\mathbf{Z}_1 \sim N(0,1)^{\otimes n_i \times n_{i^*}}$ and form $\mathbf{A}^{(i,i^*)} = (\mathbf{Y}^{(i,i^*)} + \mathbf{Z}_1)/\sqrt{2}$ and $\mathbf{B}^{(i,i^*)} = (\mathbf{Y}^{(i,i^*)} - \mathbf{Z}_1)/\sqrt{2}$. Compute the top right singular vector of $\mathbf{A}^{(i,i^*)}$, denote it as $\mathbf{v}$.

(c) To compute $\hat{I}_i$, calculate $(\mathbf{B}^{(i,i^*)}_{[j,:]} \cdot \mathbf{v})$ for $1 \le j \le n_j$. These values form two data driven clusters and a cut at the largest gap at the ordered value of $\{\mathbf{B}^{(i,i^*)}_{[j,:]} \cdot \mathbf{v}\}_{j=1}^{n_i}$ returns the set $\hat{I}_i$ and $[n_i] \setminus \hat{I}_i$.

3: **Output:** $\{\hat{I}_i\}_{i=1}^d$.

THEOREM 8 (Guarantee of aggregated-SVD for CHC_R). *Consider* $\mathrm{CHC}_R(\mathbf{k}, \mathbf{n}, \lambda)$ *and Algorithm 5. There exists a uniform constant* $C_0 > 0$ *such that if*

$$\lambda \geq C_0 \frac{k\sqrt{\prod_{i=1}^d n_i}}{\sqrt{n_{\min}}\prod_{i=1}^d k_i}\left(1 + \sqrt{\frac{k\log n}{n_{\min}}}\right), \tag{8}$$

the support recovery algorithm based on Aggregated-SVD identifies the true support of $\boldsymbol{\mathcal{X}}$ *with probability at least* $1 - \sum_{i=1}^d n_i^{-c} - C\exp(-cn_{\min})$. *Here*, $n_{\min} = \min(n_1, \dots, n_d)$. *Moreover, under the asymptotic regime* (A2), *Aggregated-SVD achieves reliable recovery of* CHC_R *when* $\beta < (\alpha - 1/2)(d-1)$.

Combining Theorems 6–8, we can see the reliable recovery of CHC_R and ROHC_R is polynomial-time possible if $\beta < \beta^c_{\mathrm{CHC}_R} := (\alpha - 1/2)(d-1) \vee 0$ and $\beta < \beta^c_{\mathrm{ROHC}_R} := (\alpha - 1/2)d/2 \vee 0$. Since $\beta^c_{\mathrm{CHC}_R} < \beta^s_{\mathrm{CHC}_R}$ and $\beta^c_{\mathrm{ROHC}_R} < \beta^s_{\mathrm{ROHC}_R}$, the proposed polynomial-time algorithms (Algorithms 3, 4 and 5) require a strictly stronger signal-to-noise ratio than the proposed unconstrained-time ones (Algorithms 1 and 2) which leaves a significant gap between statistical optimality and computational efficiency to be discussed in Section 5.

4. High-order cluster detection: Statistical limits and polynomial-time algorithms. In this section, we investigate the statistical limits of both CHC_D and ROHC_D. For each model, we first present the statistical lower bounds of signal strength that guarantees reliable detection, then we propose the algorithms, though being computationally intense, that provably achieve the statistical lower bounds. Finally, we introduce the computationally efficient algorithms and provide the theoretical guarantees under the stronger signal-to-noise ratio.

4.1. CHC_D *and* ROHC_D: *Statistical limits.* Recall (5) and (6), Theorems 9 and 10 below give the statistical lower bounds that guarantee reliable detection for CHC_D and ROHC_D, respectively.

THEOREM 9 (Statistical lower bound of CHC_D). *Consider* $\mathrm{CHC}_D(\mathbf{k}, \mathbf{n}, \lambda)$ *under the asymptotic regime* (A1) *and assume*

$$\frac{\log(n_j/k_j)}{k_i} \to 0, \qquad \frac{\log\log(n_i/k_i)}{\log(n_j/k_j)} \to 0, \quad \text{and} \quad k_i \log\frac{n_i}{k_i} \asymp k_j \log\frac{n_j}{k_j} \tag{9}$$

for all $i, j \in [d]$, $i \neq j$. *Then if*

$$\frac{\lambda \prod_{i=1}^d k_i}{\sqrt{\prod_{i=1}^d n_i}} \to 0 \quad \text{and} \quad \limsup_{n\to\infty} \frac{\lambda(\prod_{i=1}^d k_i)^{\frac{1}{2}}}{\sqrt{2(\sum_{i=1}^d k_i \log(n_i/k_i))}} < 1, \tag{10}$$

we have $\mathcal{E}^s_{\mathrm{CHC}_D} \to 1$. *Moreover, under* (A2), *if* $\beta > \beta^s_{\mathrm{CHC}_D}$, $\mathcal{E}^s_{\mathrm{CHC}_D} \to 1$.

THEOREM 10 (Statistical lower bound of ROHC_D). *Consider* $\mathrm{ROHC}_D(\mathbf{k}, \mathbf{n}, \mu)$. *Under the asymptotic regime* (A1), *if* $\frac{\mu}{\sqrt{k\log(en/k)}} \to 0$, *then* $\mathcal{E}^s_{\mathrm{ROHC}_D} \to 1$. *Under the asymptotic regime* (A2), *if* $\beta > \beta^s_{\mathrm{ROHC}_D}$, $\mathcal{E}^s_{\mathrm{ROHC}_D} \to 1$.

Next, we present the hypothesis tests $\psi^s_{\mathrm{CHC}_D}$ and $\psi^s_{\mathrm{ROHC}_D}$ that achieve reliable detection on the statistical limits in Theorems 9 and 10. For CHC_D, define $\psi^s_{\mathrm{CHC}_D} := \psi_{\mathrm{sum}} \vee \psi_{\mathrm{scan}}$. Here, ψ_{sum} and ψ_{scan} are respectively the sum and scan tests:

$$\psi_{\mathrm{sum}} = \mathbf{1}\left(\sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} \boldsymbol{\mathcal{Y}}_{[i_1,\dots,i_d]}/\sqrt{n_1\cdots n_d} > W\right) \tag{11}$$

for some to-be-specified $W > 0$ and

$$\psi_{\text{scan}} = \mathbf{1}\Big(T_{\text{scan}} > \sqrt{2\log(G_{\mathbf{k}}^{\mathbf{n}})}\Big), \qquad T_{\text{scan}} = \max_{C \in \mathcal{S}_{\mathbf{k},\mathbf{n}}} \frac{\sum_{(i_1,\ldots,i_d)\in C} \mathcal{Y}_{[i_1,\ldots,i_d]}}{\sqrt{k_1 \cdots k_d}}, \tag{12}$$

where $G_{\mathbf{k}}^{\mathbf{n}} = \binom{n_1}{k_1}\binom{n_2}{k_2}\cdots\binom{n_d}{k_d}$ and $\mathcal{S}_{\mathbf{k},\mathbf{n}}$ represents the set of all possible supports of planted signal:

$$\begin{aligned} \mathcal{S}_{\mathbf{k},\mathbf{n}} = \big\{&(I_1 \times I_2 \times \cdots \times I_d) : I_1 \subseteq [n_1], \\ &I_2 \subseteq [n_2], \ldots, I_d \subseteq [n_d] \text{ and } |I_i| = k_i, 1 \le i \le d\big\}. \end{aligned} \tag{13}$$

The following Theorem 11 provides the statistical guarantee for $\psi^s_{\text{CHC}_D}$.

THEOREM 11 (Guarantee for $\psi^s_{\text{CHC}_D}$). *Consider* $\text{CHC}_D(\mathbf{k}, \mathbf{n}, \lambda)$. *Under the asymptotic regime* (A1), *when*

$$\frac{\lambda \prod_{i=1}^d k_i}{\sqrt{\prod_{i=1}^d n_i}} \to \infty, \qquad W \to \infty, \qquad W \le c\lambda \frac{\prod_{i=1}^d k_i}{\sqrt{\prod_{i=1}^d n_i}} \quad (0 < c < 1), \tag{14}$$

or when

$$\liminf_{n\to\infty} \frac{\lambda(\prod_{i=1}^d k_i)^{\frac{1}{2}}}{\sqrt{2(\sum_{i=1}^d k_i \log(\frac{n_i}{k_i}))}} > 1, \tag{15}$$

we have $\mathcal{E}_{\text{CHC}_D}(\psi^s_{\text{CHC}_D}) \to 0$. *Under the asymptotic regime* (A2), $\psi^s_{\text{CHC}_D}$ *succeeds in reliable detection when* $\beta < \beta^s_{\text{CHC}_D}$.

The test for ROHC_D is built upon the ROHC Search (Algorithm 2 in Section 3) designed for ROHC_R. To be specific, generate $\mathcal{Z}_1$ with i.i.d. standard Gaussian entries and calculate $\mathcal{A} = \frac{\mathcal{Y}+\mathcal{Z}_1}{\sqrt{2}}$ and $\mathcal{B} = \frac{\mathcal{Y}-\mathcal{Z}_1}{\sqrt{2}}$. Then $\mathcal{A}$ and $\mathcal{B}$ becomes two independent samples. Apply Algorithm 2 on $\mathcal{A}$ and let $(\mathbf{u}_1, \ldots, \mathbf{u}_d)$ be the output of Step 4 of Algorithm 2. Define the test statistic as

$$\psi^s_{\text{ROHC}_D} = \mathbf{1}(\mathcal{B} \times_1 \mathbf{u}_1^\top/\sqrt{k_1} \times \cdots \times_d \mathbf{u}_d^\top/\sqrt{k_d} \ge C\sqrt{k}),$$

where $C > 0$ is a fixed constant. We have the following theoretical guarantee for $\psi^s_{\text{ROHC}_D}$.

THEOREM 12 (Guarantee for $\psi^s_{\text{ROHC}_D}$). *Consider* $\text{ROHC}_D(\mathbf{k}, \mathbf{n}, \mu)$ *under the asymptotic regime* (A1). *There exists some constant* $C > 0$ *such that when* $\mu \ge C\sqrt{k \log n}$, $\mathcal{E}_{\text{ROHC}_D}(\psi^s_{\text{ROHC}_D}) \to 0$. *Moreover, under the asymptotic regime* (A2), $\psi^s_{\text{ROHC}_D}$ *succeeds in reliable detection when* $\beta < \beta^s_{\text{ROHC}_D}$.

Combining Theorems 9 and 11, we have shown $\psi^s_{\text{CHC}_D}$ achieves sharply minimax lower bound of λ for reliable detection of CHC_D. From Theorems 10 and 12, we see ψ^s_{ROHC} achieves the minimax optimal rate of μ for reliable detection of ROHC_D. However, both $\psi^s_{\text{CHC}_D}$ and $\psi^s_{\text{ROHC}_D}$ are computationally inefficient.

REMARK 1. The proposed $\psi^s_{\text{CHC}_D}$ and $\psi^s_{ROHC_D}$ share similar spirits with the matrix clustering algorithms in the literature (Brennan, Bresler and Huleihel (2018), Butucea and Ingster (2013)), though the tensor structure here causes extra layer of difficulty. Particularly when $d = 2$, the lower and upper bounds results in Theorem 9–12 match the ones in Brennan, Bresler and Huleihel (2018), Butucea and Ingster (2013), although the proof for high-order clustering is much more complicated.

4.2. CHC_D *and* ROHC_D: *Polynomial-time algorithms.* Next, we introduce polynomial-time algorithms for high-order cluster detection. For CHC_D, define $\psi^c_{\mathrm{CHC}_D} := \psi_{\mathrm{sum}} \vee \psi_{\max}$, where ψ_{sum} is defined in (11) and $\psi_{\max}$ is defined below based on max statistic,

$$\psi_{\max} = \mathbf{1}\left(\max_{\substack{1\le i_j\le n_j\\ j=1,\dots,d}} \mathcal{Y}_{[i_1,\dots,i_d]} > \sqrt{2\sum_{i=1}^d \log n_i}\right). \tag{16}$$

THEOREM 13 (Theoretical guarantee for $\psi^c_{\mathrm{CHC}_D}$). *Consider* $\mathrm{CHC}_D(\mathbf{k},\mathbf{n},\lambda)$. *Under the asymptotic regime* (A1), *if condition* (14) *holds or*

$$\liminf_{n\to\infty} \frac{\lambda}{\sqrt{2\sum_{i=1}^d \log n_i}} > 1, \tag{17}$$

holds, then $\mathcal{E}_{\mathrm{CHC}_D}(\psi^c_{\mathrm{CHC}_D}) \to 0$. *Moreover, under the asymptotic regime* (A2), $\psi^c_{\mathrm{CHC}_D}$ *succeeds in reliable detection when* $\beta < \beta^c_{\mathrm{CHC}_D}$.

We also propose a polynomial-time algorithm for ROHC_D based on a high-order analogue of the largest matrix singular value in tensor. Following the procedure of $\psi^s_{\mathrm{ROHC}_D}$, we construct $\mathcal{A}$ and $\mathcal{B}$ as two independent copies. Apply Algorithm 4 in Section 3 on $\mathcal{A}$ and let $(\mathbf{u}_1,\dots,\mathbf{u}_d)$ to be the output of Step 4 of Algorithm 4. We define

$$\psi^c_{\mathrm{ROHC}_D} = \psi_{\mathrm{sing}} \vee \psi_{\max}, \qquad \psi_{\mathrm{sing}} = \mathbf{1}\big(\mathcal{B} \times_1 \mathbf{u}_1^\top \times \cdots \times_d \mathbf{u}_d^\top \ge C\sqrt{k}\big), \tag{18}$$

where $\psi_{\max}$ is defined in (16) and $C > 0$ is a fixed constant.

THEOREM 14 (Theoretical guarantee for $\psi^c_{\mathrm{ROHC}_D}$). *Consider* $\mathrm{ROHC}_D(\mathbf{k},\mathbf{n},\mu)$ *under the asymptotic regime* (A1). *There exists a constant* $C > 0$ *such that when*

$$\mu \ge C n^{\frac{d}{4}} \quad \text{or} \quad \liminf_{n\to\infty} \frac{\mu}{\sqrt{2(\prod_{i=1}^d k_i)(\sum_{i=1}^d \log n_i)}} > 1, \tag{19}$$

we have $\mathcal{E}_{\mathrm{ROHC}_D}(\psi^c_{\mathrm{ROHC}_D}) \to 0$. *Moreover, under the asymptotic regime* (A2), $\psi^c_{\mathrm{ROHC}_D}$ *succeeds in reliable detection when* $\beta < \beta^c_{\mathrm{ROHC}_D}$.

Since $\beta^c_{\mathrm{CHC}_D} < \beta^s_{\mathrm{CHC}_D}$ and $\beta^c_{\mathrm{ROHC}_D} < \beta^s_{\mathrm{ROHC}_D}$, the proposed polynomial-time algorithms $\psi^c_{\mathrm{CHC}_D}$ and $\psi^c_{\mathrm{ROHC}_D}$ require a strictly stronger signal-noise-ratio than the unrestricted-time algorithms.

5. Computational lower bounds. To provide the computational lower bounds for high-order clustering, it suffices to focus on the asymptotic regime (A2) as it also implies the computational lower bounds in the general parameterization regime (A1). We first consider the detection of CHC. Theorem 15 below and Theorem 13 in Section 4.2 together yield a tight computational lower bound for CHC_D.

THEOREM 15 (Computational lower bound of CHC_D). *Consider* $\mathrm{CHC}_D(\mathbf{k},\mathbf{n},\lambda)$ *under the asymptotic regime* (A2). *If* $\beta > \beta^c_{\mathrm{CHC}_D}$, *then* $\liminf_{n\to\infty} \mathcal{E}^c_{\mathrm{CHC}_D} \ge 1/2$ *under the HPC detection Conjecture* 1.

Next, Theorems 6, 8, and Theorem 16 below together give a tight computational lower bound for CHC_R.

THEOREM 16 (Computational lower bound of CHC_R). *Consider* $\mathrm{CHC}_R(\mathbf{k}, \mathbf{n}, \lambda)$ *under the asymptotic regime* (A2). *If* $\alpha \geq 1/2$ *and* $\beta > (d-1)\alpha - (d-1)/2$, *then* $\liminf_{n\to\infty} \mathcal{E}^c_{\mathrm{CHC}_R} \geq 1/2$ *under the HPDS recovery conjecture* (*Conjecture* 2). *If* $0 < \alpha < 1/2$, $\beta > 0$, *then* $\liminf_{n\to\infty} \mathcal{E}^c_{\mathrm{CHC}_R} \geq 1/2$ *under the HPC detection conjecture* (*Conjecture* 1). *Combined together, we have if* $\beta > \beta^c_{\mathrm{CHC}_R}$, *then* $\liminf_{n\to\infty} \mathcal{E}^c_{\mathrm{CHC}_R} \geq 1/2$ *under Conjectures* 1 *and* 2.

Then, we consider rank-one high-order cluster detection and recovery. By Lemma 10 in Luo and Zhang (2022) Section B, we can show that the computational lower bound of ROHC_R is implied by ROHC_D. We specifically have the following theorem.

THEOREM 17 (Computational lower bounds of ROHC_D and ROHC_R). *Consider* $\mathrm{ROHC}_D(\mathbf{k}, \mathbf{n}, \mu)$ *and* $\mathrm{ROHC}_R(\mathbf{k}, \mathbf{n}, \mu)$ *under the asymptotic regime* (A2) *and the HPC detection Conjecture* 1. *If* $\beta > \beta^c_{\mathrm{ROHC}_D}$, *then* $\liminf_{n\to\infty} \mathcal{E}^c_{\mathrm{ROHC}_D} \geq 1/2$, $\liminf_{n\to\infty} \mathcal{E}^c_{\mathrm{ROHC}_R} \geq 1/2$.

Combining Theorems 6, 7, 17, and 14 (provided in Section 4.2), we have obtained the tight computational lower bounds for ROHC_D and ROHC_R. Furthermore, since ROHC is a special case of sparse tensor PCA/SVD studied in literature (Sun et al. (2017), Zhang and Han (2019)), Theorem 17 also provides a computational lower bound for the signal-to-noise ratio requirement for sparse tensor PCA/SVD.

REMARK 2. The computational lower bounds in Theorems 15, 16 and 17 are for asymmetric tensor clustering under the CHC and ROHC models. To establish the computational lower bounds for a symmetric version of the CHC or ROHC models that both the planted signal and the noise tensors are symmetric, a new proof scheme is required as the same sparsity across all modes must be ensured while constructing instance tensors in performing the average-case reduction.

Theorems 15–17 above are based on the HPC and HPDS conjectures. Next, we will elaborate the HPC, HPDS conjectures in Sections 5.1, 5.2, and discuss the evidence in Section 5.3. Then in Section 5.4, we provide the high level ideas on the average-case reduction from HPC and HPDS to high-order clustering, and prove these computational lower bounds.

5.1. *Hypergraphic planted clique detection.* A d-hypergraph can be seen as an order-d extension of regular graph. In a d-hypergraph $G = (V(G), E(G))$, each hyper-edge $e \in E$ includes an unordered group of d vertices in V. Define $\mathcal{G}_d(N, 1/2)$ as Erdős–Rényi random d-hypergraph with N vertices, where each hyper-edge $(i_1, \dots, i_d)$ is independently included in E with probability $\frac{1}{2}$. Given a d-hypergraph $G = (V(G), E(G))$, define its adjacency tensor $\boldsymbol{\mathcal{A}} := \boldsymbol{\mathcal{A}}(G) \in (\{0,1\}^N)^{\otimes d}$ as

$$\boldsymbol{\mathcal{A}}_{[i_1,\dots,i_d]} = \begin{cases} 1, & \text{if } (i_1,\dots,i_d) \in E; \\ 0, & \text{otherwise.} \end{cases}$$

We define $\mathcal{G}_d(N, \frac{1}{2}, \kappa)$ as the hypergraphic planted clique (HPC) model with clique size κ. To generate $G \sim \mathcal{G}_d(N, \frac{1}{2}, \kappa)$, we sample a random hypergraph from $\mathcal{G}_d(N, \frac{1}{2})$, pick κ vertices uniformly at random from $[N]$, denote them as K, and connect all hyper-edges e if all vertices of e are in K. The focus of this section is on the *hypergraphic planted clique detection (HPC)* problem:

$$H_0^G : G \sim \mathcal{G}_d(N, 1/2) \quad \text{v.s.} \quad H_1^G : G \sim \mathcal{G}_d(N, 1/2, \kappa). \tag{20}$$

Given the hypergraph G and its adjacency tensor $\boldsymbol{\mathcal{A}}$, the risk of test ϕ for (20) is defined as the sum of Type-I and II errors $\mathcal{E}_{\mathrm{HPC}_D}(\phi) = \mathbb{P}_{H_0^G}(\phi(\boldsymbol{\mathcal{A}}) = 1) + \mathbb{P}_{H_1^G}(\phi(\boldsymbol{\mathcal{A}}) = 0)$. Our aim is to find out the consistent test $\phi = \{\phi_N\}$ such that $\lim_{N\to\infty} \mathcal{E}_{\mathrm{HPC}_D}(\phi_N) = 0$.

When $d = 2$, HPC detection (20) reduces to the planted clique (PC) detection studied in literature. It is helpful to have a quick review of existing results for PC before addressing HPC. Since the size of the largest clique in Erdős–Rényi graph $G \sim \mathcal{G}_2(N, \frac{1}{2})$ converges to $2\log_2 N$ asymptotically, reliable PC detection can be achieved by exhaustive search whenever $\kappa \geq (2+\epsilon)\log_2 N$ for any $\epsilon > 0$ (Bollobás and Erdős (1976)). When $\kappa = \Omega(\sqrt{N})$, many computational-efficient algorithms, including the spectral method, approximate message passing, semidefinite programming, nuclear norm minimization, and combinatorial approaches (Alon, Krivelevich and Sudakov (1998), Ames and Vavasis (2011), Dekel, Gurel-Gurevich and Peres (2014), Deshpande and Montanari (2015a), Feige and Krauthgamer (2000), McSherry (2001), Ron and Feige (2010), Chen and Xu (2016)), have been developed for PC detection. Despite enormous previous efforts, no polynomial-time algorithm has been found for reliable detection of PC when $\kappa = o(N^{1/2})$ and it has been widely conjectured that no polynomial-time algorithm can achieve so. The hardness conjecture of PC detection was strengthened by several pieces of evidence, including the failure of Metropolis process methods (Jerrum (1992)), low-degree polynomial methods (Hopkins (2018), Brennan and Bresler (2019b)), statistical query model (Feldman et al. (2017)), Sum-of-Squares (Barak et al. (2019), Deshpande and Montanari (2015b), Meka, Potechin and Wigderson (2015)), landscape of optimization (Gamarnik and Zadik (2019)), etc.

When moving to HPC detection (20) with $d \geq 3$, the computational hardness remains little studied. Bollobás and Erdős (1976) proved that $\frac{K_N^d}{(d!\log_2(N))^{1/(d-1)}} \overset{a.s.}{\to} 1$ if K_N^d is the largest clique in $G \sim \mathcal{G}_d(N, \frac{1}{2})$. So HPC detection problem (20) is statistical possible by exhaustive search when $\kappa \geq ((d! + \epsilon)\log_2(N))^{1/(d-1)}$ for any $\epsilon > 0$. However, Zhang and Xia (2018) observed that the spectral algorithm solves HPC detection if $\kappa = \Omega(\sqrt{N})$ but fails when $\kappa = N^{\frac{1}{2}-\epsilon}$ for any $\epsilon > 0$. We present the following hardness conjecture for HPC detection.

CONJECTURE 1 (HPC detection conjecture). Consider the HPC detection problem (20) and suppose $d \geq 2$ is a fixed integer. If

$$\limsup_{N\to\infty} \log\kappa / \log\sqrt{N} \leq 1 - \tau \quad \text{for any } \tau > 0, \tag{21}$$

for any polynomial-time test sequence $\{\phi\}_N : \boldsymbol{\mathcal{A}} \to \{0, 1\}$, $\liminf_{N\to\infty} \mathcal{E}_{\mathrm{HPC}_D}(\phi(\boldsymbol{\mathcal{A}})) \geq \frac{1}{2}$.

REMARK 3 (Choice of type-I, II error lower bound). We set the lower bound for the sum of Type-I, II errors to be $1/2$ in the HPC Detection Conjecture above (i.e., $\{\phi\}_N : \boldsymbol{\mathcal{A}} \to \{0, 1\}$, $\liminf_{N\to\infty} \mathcal{E}_{\mathrm{HPC}_D}(\phi(\boldsymbol{\mathcal{A}})) \geq 1/2$). In the literature, there is no universal choice of this constant. For example, Berthet and Rigollet (2013a) considers PC detection conjecture with the sum of type I and type II errors to be some constant close to 1; Ma and Wu (2015) uses the PC detection conjecture with the error constant $2/3$; Brennan and Bresler (2019a), Brennan and Bresler (2020), Brennan, Bresler and Huleihel (2018), Hajek, Wu and Xu (2015) choose this constant to be 1.

In Section 5.3, we provide two pieces of evidence for HPC detection conjecture: a general class of Monte Carlo Markov chain process methods (Jerrum (1992)) and a general class of low-degree polynomial tests (Hopkins (2018), Hopkins and Steurer (2017), Kunisky, Wein and Bandeira (2019), Brennan and Bresler (2019b)) fail to solve HPC detection under the asymptotic condition (21). Also, see a recent note Luo and Zhang (2020a) for several open questions on HPC detection conjecture, in particular, whether HPC detection is equivalently hard as PC detection.

5.2. *Hypergraphic planted dense subgraph.* We consider the *hypergraphic planted dense subgraph (HPDS)*, a hypergraph model with denser connections within a community and sparser connections outside, in this section. Let $\mathcal{G}_d$ be a d-hypergraph. To generate a HPDS $G = (V(G), E(G)) \sim \mathcal{G}_d(N, \kappa, q_1, q_2)$ with $q_1 > q_2$, we first select a size-κ subset K from $[N]$ uniformly at random, then for each hyper-edge $e = (i_1, \dots, i_d)$,

$$\mathbb{P}\big(e \in E(G)\big) = \begin{cases} q_1, & i_1, \dots, i_d \in K, \\ q_2, & \text{otherwise.} \end{cases}$$

The aim of HPDS detection is to test

$$H_0 : G \sim \mathcal{G}_d(N, q_2) \text{versus} H_1 : G \sim \mathcal{G}_d(N, \kappa, q_1, q_2); \tag{22}$$

the aim of HPDS recovery is to locate the planted support K given $G \sim \mathcal{G}_d(N, \kappa, q_1, q_2)$.

When $d = 2$, HPDS reduces to the planted dense subgraph (PDS) considered in literature. Various statistical limits of PDS have been studied (Arias-Castro and Verzelen (2014), Brennan, Bresler and Huleihel (2018), Chen and Xu (2016), Feldman et al. (2017), Hajek, Wu and Xu (2015), Verzelen and Arias-Castro (2015)) and generalizations of PDS recovery has also been considered in Candogan and Chandrasekaran (2018), Hajek, Wu and Xu (2016), Montanari (2015). In Brennan, Bresler and Huleihel (2018), Hajek, Wu and Xu (2015), a reduction from PC has shown the statistical and computational phase transition for PDS detection problem for all $q_1 > q_2$ with $q_1 - q_2 = O(q_2)$ where $q_2 = \tilde{\Theta}(N^{-\beta})$. For PDS recovery problem, Brennan, Bresler and Huleihel (2018), Chen and Xu (2016), Hajek, Wu and Xu (2015) observed that PDS appears to have a detection-recovery gap in the regime when $\kappa \gg \sqrt{N}$.

When moving to HPDS detection, if $q_1 = \omega(q_2)$, the computational barrier for this problem is conjectured to be the log-density threshold $\kappa = \tilde{\Theta}(N^{\log_{q_2} q_1})$ when $\kappa \ll \sqrt{N}$ (Chlamtáč, Dinitz and Krauthgamer (2012), Chlamtáč, Dinitz and Makarychev (2017)). Recently, Chlamtác and Manurangsi (2018) showed that $\tilde{\Omega}(\log N)$ rounds of the Sherali-Adams hierarchy cannot solve the HPDS detection problem below the log-density threshold in the regime $q_1 = \omega(q_2)$. The HPDS recovery, to the best of our knowledge, remains unstudied in the literature.

In the following Proposition 1, we show that a variant of Aggregated-SVD (presented in Algorithm 6) requires a restricted condition on κ, q_1, q_2, N for reliable recovery in HPDS in the regime $\kappa \gg \sqrt{N}$.

PROPOSITION 1. *Suppose $G \sim \mathcal{G}_d(N, \kappa, q_1, q_2)$ with $q_1 > q_2$. Let $\boldsymbol{\mathcal{A}}$ be the adjacency tensor of G. When* $\liminf_{N\to\infty} \log_N \kappa \geq 1/2$, *and*

$$\limsup_{N\to\infty} \log_N \left(\frac{\kappa^{d-1}(q_1 - q_2)}{\sqrt{q_2(1 - q_2)}} \right) \geq \frac{d}{2} - \frac{1}{2}, \tag{23}$$

Algorithm 6 Support Recovery of HPDS via Aggregated-SVD

1: **Input:** $\boldsymbol{\mathcal{A}}$.
2: Let $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}_{[1:\lfloor \frac{N}{d} \rfloor, \lfloor \frac{N}{d} \rfloor+1:2\lfloor \frac{N}{d} \rfloor, \dots, (d-1)\lfloor \frac{N}{d} \rfloor+1:N]]}$.
3: Let $\widetilde{\boldsymbol{\mathcal{A}}}_{[i_1,\dots,i_d]} = \frac{\breve{\boldsymbol{\mathcal{A}}}_{[i_1,\dots,i_d]} - q_2}{\sqrt{q_2(1-q_2)}}$ for all $1 \leq i_1 \leq \lfloor \frac{N}{d} \rfloor, \dots, (d-1)\lfloor \frac{N}{d} \rfloor + 1 \leq i_d \leq N$. Then apply Algorithm 5 with input $\widetilde{\boldsymbol{\mathcal{A}}}$ and denote the estimated support for each mode of $\widetilde{\boldsymbol{\mathcal{A}}}$ as $\hat{K}_i$.
4: Compute $\hat{K} = \bigcup_{i=1}^{d} \hat{K}_i$.
5: **Output:** $\hat{K}$.

Algorithm 6 recovers the support of the planted dense subgraph with probability at least $1 - d(N/d)^{-c} - C\exp(-cN/d)$ *for some* $c, C > 0$.

On the other hand, the theoretical analysis in Proposition 1 breaks down when condition (23) does not hold. We conjecture that the signal-to-noise ratio requirement in (23) is essential for HPDS recovery and propose the following computational hardness conjecture.

CONJECTURE 2 (HPDS recovery conjecture). Suppose $G \sim \mathcal{G}_d(N, \kappa, q_1, q_2)$ with $1 - \Omega(1) > q_1 > q_2$. Denote its adjacency tensor as $\boldsymbol{\mathcal{A}}$. If

$$\liminf_{N\to\infty} \log_N \kappa \geq \frac{1}{2} \quad \text{and} \quad \limsup_{N\to\infty} \log_N \left(\frac{\kappa^{d-1}(q_1 - q_2)}{\sqrt{q_2(1-q_2)}} \right) < \frac{d}{2} - \frac{1}{2}, \tag{24}$$

then for any randomized polynomial-time algorithm $\{\phi\}_N$, $\liminf_{N\to\infty} \mathcal{E}_{\mathrm{HPDS}_R}(\phi(\boldsymbol{\mathcal{A}})) \geq \frac{1}{2}$.

In the proof of Proposition 1, we provide evidence for Conjecture 2 by showing a variant of Aggregated-SVD fails to solve HPDS recovery under the PC detection conjecture. A stronger piece of evidence for Conjecture 2 via low-degree polynomial method is given in Section 5.3.3.

5.3. *Evidence for HPC detection conjecture.* In this section, we provide two pieces of evidence for HPC Conjecture 1 via Monte Carlo Markov chain process and low-degree polynomial test and one piece of evidence for HPDS recovery Conjecture 2 via low-degree polynomial method.

5.3.1. *Evidence of HPC conjecture 1 via Metropolis process.* We first show a general class of Metropolis processes are not able to detect or recover the large planted clique in hypergraph. Motivated by Alon et al. (2007), in Lemma 1 we first prove that if it is computationally hard to recover a planted clique in HPC, it is also computationally hard to detect.

LEMMA 1. *Assume* $\kappa > \Omega(\log N)$. *Consider the* $\mathrm{HPC}_D(N, 1/2, \kappa)$ *problem: test*

$$H_0 : G \sim \mathcal{G}_d(N, 1/2) \quad \textit{versus} \quad H_1 : G \sim \mathcal{G}_d(N, 1/2, \kappa)$$

and $\mathrm{HPC}_R(N, 1/2, \kappa)$ *problem: recover the exact support of the planted clique if* H_1 *holds. If there is no polynomial time recovery algorithm can output the right clique of* $\mathrm{HPC}_R(N, 1/2, \kappa)$ *with success probability at least* $1 - 1/N$, *then there is no polynomial time detection algorithm can output the right hypothesis for* $\mathrm{HPC}_D(N, 1/2, \kappa/3)$ *with probability* $1 - 1/(4N^d)$.

By Lemma 1, to show the computational hardness of HPC detection, we only need to show the HPC recovery.

Motivated by the seminal work of Jerrum (1992), we consider the following simulated annealing method for planted clique recovery in hypergraph. Given a hypergraph $G = (V, E) \sim \mathcal{G}_d(N, 1/2, \kappa)$ on the vertex set $V = \{0, \dots, N-1\}$ and a real number $\theta \geq 1$, we consider a Metropolis process on the state space of the collection $\Gamma \subseteq 2^V$ of all cliques in G, that is, all subsets of V which induces the complete subgraph in G. A transition from state K to state K' is allowed if $|K \oplus K'| \leq 1$. (Here, $K \oplus K' = \{i : i \in K, i \notin K'\} \cup \{i : i \in K', i \notin K\}$ is the set symmetric difference.)

For all distinct states $K, K' \in \Gamma$, the transition probability from K to K' is

$$P(K, K') = \begin{cases} \frac{1}{N\theta}, & \text{if } K \oplus K' = 1, K \supset K'; \\ \frac{1}{N}, & \text{if } K \oplus K' = 1, K \subset K'; \\ 0, & \text{if } |K \oplus K'| \geq 2. \end{cases} \tag{25}$$

The loop probability $P(K, K) = 1 - \sum_{K' \neq K} P(K, K')$ are defined by complementation. The transition probability can be interpreted by the following random process. Suppose the current state is K. Pick a vertex v uniformly at random from V.

1. If $v \notin K$ and $K \cup \{v\}$ is a clique, then let $K' = K \cup \{v\}$;
2. If $v \notin K$ and $K \cup \{v\}$ is not a clique, then let $K' = K$;
3. If $v \in K$, with probability $\frac{1}{\theta}$, set $K' = K \setminus \{v\}$, else set $K' = K$.

When $\theta > 1$, the Metropolis process defined above is aperiodic and then has a unique statitionary distribution. Let $\pi : \Gamma \to [0, 1]$ be defined as

$$\pi(K) = \frac{\theta^{|K|}}{\sum_{K \in \Gamma} \theta^{|K|}}.$$

We can check that π satisfies the following detailed balance property:

$$\theta^{|K|} P(K, K') = \theta^{|K'|} P(K', K), \quad \text{for all } K, K' \in \Gamma. \tag{26}$$

This means π is indeed the stationary distribution of this Markov chain. The following theorem shows that it takes superpolynomial time to locate a clique in G of size $\Omega((\log_2 N)^{1/(d-1)})$ by described Metropolis process.

THEOREM 18 (Hardness of finding large clique in $\mathcal{G}_d(N, 1/2, N^\beta)$, $0 < \beta < \frac{1}{2}$). *Suppose $\epsilon > 0$ and $0 < \beta < \frac{1}{2}$. For almost every $G \in \mathcal{G}_d(N, 1/2, N^\beta)$ and every $\theta > 1$, there exists an initial state from which the expected time for the Metropolis process to reach a clique of size at least m exceeds $N^{\Omega((\log_2 N)^{1/(d-1)})}$. Here,*

$$m = 2\left\lceil \left(\left(1 + \frac{2}{3}\epsilon\right)\frac{d!}{2}\log_2 N\right)^{\frac{1}{d-1}} \right\rceil - \left\lceil \left(\left(1 + \frac{2}{3}\epsilon\right)(d-1)!\log_2 N\right)^{\frac{1}{d-1}} \right\rceil \asymp_d (\log_2 N)^{\frac{1}{d-1}}.$$

5.3.2. *Evidence of HPC conjecture 1 via low-degree polynomial test.* We also consider the low-degree polynomial tests to establish the computational hardness for hypergraphic planted clique detection. The idea of using low-degree polynomial to predict the statistical and computational gap is recently developed in a line of papers (Barak et al. (2019), Hopkins (2018), Hopkins and Steurer (2017), Hopkins et al. (2017)). Many state-of-the-art algorithms, such as spectral algorithm, approximate message passing (Donoho, Maleki and Montanari (2009)) can be represented as low-degree polynomial functions as the input, where "low" means logarithmic in the dimension. In comparison to sum-of-squares (SOS) computational lower bounds, the low-degree method is simpler to carry out and appears to always yields the same results for natural average-case problems, such as the planted clique detection (Hopkins (2018), Barak et al. (2019)), community detection in stochastic block model (Hopkins and Steurer (2017), Hopkins (2018)), the spiked tensor model (Hopkins (2018), Hopkins et al. (2017), Kunisky, Wein and Bandeira (2019)), the spiked Wishart model (Bandeira, Kunisky and Wein (2020)), sparse PCA (Ding et al. (2019)), spiked Wigner model (Kunisky, Wein and Bandeira (2019)), sparse clustering (Löffler, Wein and Bandeira (2020)), certifying RIP (Ding et al. (2020)) and a variant of planted clique and planted dense subgraph models

(Brennan and Bresler (2019b)). It is gradually believed that the low-degree polynomial method is able to capture the essence of what makes SOS succeed or fail (Hopkins (2018), Hopkins and Steurer (2017), Hopkins et al. (2017), Kunisky, Wein and Bandeira (2019), Raghavendra, Schramm and Steurer (2018)). Therefore, we apply this method to give the evidence for the computational hardness of HPC detection (20). Specifically, we have the following Theorem 19 for low degree polynomial tests in HPC.

THEOREM 19 (Failure of low-degree polynomial tests for HPC). *Consider the HPC detection problem* (20) *for* $\kappa = N^{\beta} (0 < \beta < \frac{1}{2})$. *Suppose* $\boldsymbol{\mathcal{A}}$ *is the adjacency tensor of* G *and* $f(\boldsymbol{\mathcal{A}})$ *is a polynomial test such that* $\mathbb{E}_{H_0^G} f(\boldsymbol{\mathcal{A}}) = 0$, $\mathbb{E}_{H_0^G}(f^2(\boldsymbol{\mathcal{A}})) = 1$, *and the degree of* f *is at most* D *with* $D \leq C \log N$ *for constant* $C > 0$. *Then we have* $\mathbb{E}_{H_1^G} f(\boldsymbol{\mathcal{A}}) = O(1)$.

It has been widely conjectured in the literature that for a broad class of hypothesis testing problems: H_0 versus H_1, there is a test with runtime $n^{\tilde{O}(D)}$ and Type I + II error tending to zero if and only if there is a successful D-simple statistic, that is, a polynomial f of degree at most D, such that $\mathbb{E}_{H_0} f(X) = 0$, $\mathbb{E}_{H_0}(f^2(X)) = 1$, and $\mathbb{E}_{H_1} f(X) \to \infty$ (Brennan and Bresler (2019b), Ding et al. (2019), Hopkins (2018), Kunisky, Wein and Bandeira (2019)). Thus, Theorem 19 provides the firm evidence that there is no polynomial-time test algorithm that can reliably distinguish between $\mathcal{G}_d(N, 1/2)$ and $\mathcal{G}_d(N, 1/2, N^{\beta})$ for $0 < \beta < 1/2$.

5.3.3. *Evidence of HPDS recovery Conjecture* 2 *via low-degree polynomial method.* Compared to the hardness evidence for the hypothesis testing problems, it is much less explored in the literature to establish hardness evidence for the *estimation or recovery problems*. Recently, Schramm and Wein (2020) provides the first sharp computational lower bounds for recovery in biclustering and planted dense subgraph via the low-degree polynomial method and resolve the "detection-recovery gap" open problem mentioned in Brennan, Bresler and Huleihel (2018), Chen and Xu (2016), Hajek, Wu and Xu (2015), Ma and Wu (2015). In this work, we leverage the results in Schramm and Wein (2020) and provide the firm evidence for HPDS recovery Conjecture 2 via the low-degree polynomial method.

Recall the HPDS recovery problem in Section 5.2. Let $G \sim \mathcal{G}_d(N, \kappa, q_1, q_2)$ with $q_1 > q_2$ and planted subset K. Denote $v_1 \in \{0, 1\}$ as the membership of vertex 1 such that $v_1 = 1$ if the first vertex is in K and $v_1 = 0$ otherwise. The following theorem shows that it is impossible to estimate v_1 well in the conjectured hard regime via low-degree polynomials, which implies the computational difficulty of recovering K in general.

THEOREM 20 (Failure of low-degree polynomials for HPDS recovery). *Suppose* $G \sim \mathcal{G}_d(N, \kappa, q_1, q_2)$ *with* $q_1 > q_2$ *and* $\boldsymbol{\mathcal{A}}$ *is the adjacency tensor of* G. *For any* $0 < r < 1$ *and* $D \geq 1$, *if*

$$\frac{q_1 - q_2}{\sqrt{q_2(1 - q_1)}} \leq \frac{\sqrt{r}}{D + 1} \min\Big(\big(D(d - 1) + 1\big)^{-d/2}, \frac{N^{(d-1)/2}}{\sqrt{Dd(d - 1)}\kappa^{d-1}}\Big), \tag{27}$$

then for any $f : \boldsymbol{\mathcal{A}} \to \mathbb{R}$ *with degree at most* D, *we have* $\mathbb{E}(f(\boldsymbol{\mathcal{A}}) - v_1)^2 \geq \frac{\kappa}{N} - (\frac{\kappa}{N})^2(1 + \frac{r}{(1-r)^2})$.

In particular, suppose $q_2 < q_1 < 1 - \Omega(1)$. *Consider the asymptotic regime of Conjecture* 2 *that*

$$\liminf_{N\to\infty} \log_N \kappa \geq \frac{1}{2} \quad \textit{and} \quad \limsup_{N\to\infty} \log_N \bigg(\frac{\kappa^{d-1}(q_1 - q_2)}{\sqrt{q_2(1 - q_2)}}\bigg) < \frac{d}{2} - \frac{1}{2}.$$

Let f_0 be the trivial constant estimator of v_1: $f_0(\boldsymbol{\mathcal{A}}) = \kappa/N$. Then for any polynomial f : $\boldsymbol{\mathcal{A}} \to \mathbb{R}$ with degree at most D with $D \leq \text{polylog}(N)$, we have

$$\liminf_{N\to\infty} \frac{\mathbb{E}(f(\boldsymbol{\mathcal{A}}) - v_1)^2}{\mathbb{E}(f_0(\boldsymbol{\mathcal{A}}) - v_1)^2} \geq 1.$$

Theorem 20 shows that under the conjectured hard regime of HPDS (24) and $q_2 < q_1 < 1 - \Omega(1)$, the mean square error of any f with degree equal or less than polylog(N) is no better than the trivial estimator f_0. This gives strong evidence for the HPDS recovery Conjecture 2.

5.4. *Proofs of computational lower bounds.* Now, we are in position to prove the computational lower bounds. Before the detailed analysis, we first outline the high-level idea.

Consider a hypothesis testing problem B: H_0 versus H_1. To establish a computational lower bound for B, we can construct a randomized polynomial-time reduction φ from the conjecturally hard problem A to B such that the total variation distance between $\varphi(A)$ and B converges to zero under both H_0 and H_1. If such a φ can be found, whenever there exists a polynomial-time algorithm ϕ for solving B, we can also solve A using $\phi \circ \varphi$ in polynomial-time. Since A is conjecturally hard, we can conclude that B must also be polynomial-time hard by the contradiction argument. To establish the computational lower bound for a recovery problem, we can either follow the same idea above or establish a reduction from recovery to an established detection lower bound. A key challenge of average-case reduction is often how to construct an appropriate randomized polynomial-time map φ.

We summarize the procedure of constructing randomized polynomial-time maps for the high-order clustering computational lower bounds as follows.

- Input: Hypergraph G and its adjacency tensor $\boldsymbol{\mathcal{A}}$
- *Step* 1: Apply the *rejection kernel* technique, which was proposed by Ma and Wu (2015) and formalized by Brennan, Bresler and Huleihel (2018), to simultaneously map Bern(p) distribution to $N(\xi, 1)$ and Bern(q) distribution to $N(0, 1)$ approximately.
- *Step* 2: Simultaneously change the magnitude and sparsity of the planted signal guided by the target problem. In this step, we develop several new techniques and apply several ones in the literature. In CHC$_D$ (Algorithm 2), we use the average-trick idea in Ma and Wu (2015); in CHC$_R$ (Algorithm 3), we use the invariant property of Gaussian to handle the multiway-symmetricity of hypergraph; to achieve a sharper scaling of signal strength and sparsity in ROHC$_D$, ROHC$_R$ (Algorithm 8), the *tensor reflection cloning*, a generalization of reflection cloning (Brennan, Bresler and Huleihel (2018)), is introduced that spreads the signal in the planted high-order cluster along each mode evenly, maintains the independence of entries in the tensor, and only mildly reduces the signal magnitude.
- *Step* 3: Randomly permute indices of different modes to transform the symmetric planted signal tensor to an asymmetric one (Lemmas 14 and 16 in Luo and Zhang (2022)) that maps to the high-order clustering problem.

Then, we give a detailed proof of Theorem 17, that is, computational lower bounds for ROHC$_D$ and ROHC$_R$. The proofs for the computational limits of CHC$_D$ and CHC$_R$ are similar and postponed to the Supplementary Material (Luo and Zhang (2022)).

We first introduce the rejection kernel scheme given in Algorithm 9 in Luo and Zhang (2022) Section C, which simultaneously maps Bern(p) to distribution f_X and Bern(q) to distribution g_X approximately. In our high-order clustering problem, f_X and g_X are $N(\xi, 1)$ and $N(0, 1)$, that is, the distribution of the entries inside and outside the planted cluster, respectively. Here, ξ is to be specified later. Denote RK$(p \to f_X, q \to g_X, T)$ as the rejection kernel map, where T is the number of iterations in the rejection kernel algorithm.

Algorithm 7 Tensor Reflecting Cloning

1: **Input:** Tensor $\mathcal{W}_0 \in \mathbb{R}^{n^{\otimes d}}$ (n is an even number), number of iterations ℓ.
2: Initialize $\mathcal{W} = \mathcal{W}_0$.
3: For $i = 1, \dots, \ell$, do:

1. Generate a permutation σ of $[n]$ uniformly at random.
2. Calculate

$$\mathcal{W}' = \mathcal{W}^{\sigma^{\otimes d}} \times_1 \frac{\mathbf{A}+\mathbf{B}}{\sqrt{2}} \times \cdots \times_d \frac{\mathbf{A}+\mathbf{B}}{\sqrt{2}},$$

where $\mathcal{W}^{\sigma^{\otimes d}}$ means permuting each mode indices of $\mathcal{W}$ by σ and $\mathbf{B}$ is a $n \times n$ matrix with ones on its anti-diagonal and zeros elsewhere and $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_{\frac{n}{2}} & 0 \\ 0 & -\mathbf{I}_{\frac{n}{2}} \end{bmatrix}, \tag{28}$$

where $\mathbf{I}_{n/2}$ is a $n/2 \times n/2$ identity matrix.

3. Set $\mathcal{W} = \mathcal{W}'$.

4: **Output:** $\mathcal{W}$.

We then propose a new tensor reflection cloning technique in Algorithm 7. Note that the input tensor $\mathcal{W}_0$ to Algorithm 7 often has independent entries and a sparse planted cluster, we multiply $\mathcal{W}^{\sigma^{\otimes d}}$, a random permutation of $\mathcal{W}_0$, by $\frac{\mathbf{A}+\mathbf{B}}{\sqrt{2}}$ in each mode to "spread" the signal of the planted cluster along all modes while keep the entries independent. We prove some properties related to tensor reflection cloning in Lemma 16 of Luo and Zhang (2022) Section C.

We construct the randomized polynomial-time reduction from HPC to ROHC in Algorithm 8. The next lemma shows that the randomized polynomial-time mapping we construct in Algorithm 8 maps HPC to ROHC asymptotically.

LEMMA 2. *Suppose that n is even and sufficiently large. Let $\xi = \frac{\log 2}{2\sqrt{2(d+1)\log n + 2\log 2}}$. Then the randomized polynomial-time map $\varphi : \mathcal{G}_d(n) \to \mathbb{R}^{n^{\otimes d}}$ in Algorithm 8 satisfies if $G \sim \mathcal{G}_d(n, \frac{1}{2})$, it holds that*

$$\mathrm{TV}\big(\mathcal{L}(\varphi(G)), N(0,1)^{\otimes (n^{\otimes d})}\big) = O(1/n),$$

and if $G \sim \mathcal{G}_d(n, \frac{1}{2}, \kappa)$, there is a prior π on unit vectors in $\mathcal{V}_{n,2^\ell \kappa}$ such that

$$\mathrm{TV}\left(\mathcal{L}(\varphi(G)), \int \mathcal{L}\left(\frac{\xi \kappa^{\frac{d}{2}}}{\sqrt{d!}} \mathbf{u}_1 \circ \cdots \circ \mathbf{u}_d + N(0,1)^{\otimes (n^{\otimes d})}\right) d\pi(\mathbf{u}_1, \dots, \mathbf{u}_d)\right) = O(1/\sqrt{\log n}).$$

Here TV *denotes the total variation distance and $\mathcal{L}(X)$ denotes the distribution of random variable X.*

Lemma 2 specifically implies that if $k = 2^\ell \kappa$, $\mu = \frac{\xi \kappa^{\frac{d}{2}}}{\sqrt{d!}}$ with $\xi = \frac{\log 2}{2\sqrt{2(d+1)\log n + 2\log 2}}$, the reduction map $\varphi(G)$ we constructed from Algorithm 8 satisfies $\mathrm{TV}(\varphi(\mathrm{HPC}_D(n, \frac{1}{2}, \kappa)), \mathrm{ROHC}_D(\mathbf{n}, \mathbf{k}, \mu)) \to 0$ under both H_0 and H_1.

Next, we prove the computational lower bound of ROHC_D under the asymptotic regime (A2) by a contradiction argument.

Algorithm 8 Randomized Polynomial-time Reduction from HPC to ROHC

1: **Input:** Hypergraph $G \sim \mathcal{G}_d(n)$, number of iterations ℓ.
2: Let $\mathrm{RK}_G = \mathrm{RK}(1 \to N(\xi, 1), \frac{1}{2} \to N(0, 1), T)$ where $T = \lceil 2(d+1)\log_2 n \rceil$ and $\xi = \frac{\log 2}{2\sqrt{2(d+1)\log n + 2\log 2}}$ and compute the symmetric tensor $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{n^{\otimes d}}$ with $\boldsymbol{\mathcal{W}}_{[i_1,\dots,i_d]} = \mathrm{RK}_G(\mathbf{1}((i_1,\dots,i_d) \in E(G)))$. Let the diagonal entries of $\boldsymbol{\mathcal{W}}_{[i,\dots,i]}$ to be i.i.d. $N(0, 1)$.
3: Generate $(d!-1)$ i.i.d. symmetric random tensor $\boldsymbol{\mathcal{B}}^{(1)}, \dots, \boldsymbol{\mathcal{B}}^{(d!-1)}$ in the following way: their diagonal values are 0 and nondiagonal values are i.i.d. $N(0, 1)$. Given any non-diagonal index $\mathbf{i} = (i_1, \dots, i_d)$ $(i_1 \le i_2 \le \cdots \le i_d)$, suppose it has $D(D \le d!)$ unique permutations and denote them as $\mathbf{i}_{(0)} := \mathbf{i}, \mathbf{i}_{(1)}, \dots, \mathbf{i}_{(D-1)}$, then we transform $\boldsymbol{\mathcal{W}}$ in the following way:

$$\begin{pmatrix} \boldsymbol{\mathcal{W}}_{\mathbf{i}_{(0)}} \\ \boldsymbol{\mathcal{W}}_{\mathbf{i}_{(1)}} \\ \vdots \\ \boldsymbol{\mathcal{W}}_{\mathbf{i}_{(D-1)}} \end{pmatrix} = \left(\frac{\mathbf{1}}{\sqrt{d!}}, \left(\frac{\mathbf{1}}{\sqrt{d!}}\right)_{\perp}\right)_{[1:D,:]} \times \begin{pmatrix} \boldsymbol{\mathcal{W}}_{[i_1,\dots,i_d]} \\ \boldsymbol{\mathcal{B}}^{(1)}_{[i_1,\dots,i_d]} \\ \vdots \\ \boldsymbol{\mathcal{B}}^{(d!-1)}_{[i_1,\dots,i_d]} \end{pmatrix}.$$

Here $\frac{\mathbf{1}}{\sqrt{d!}}$ is a $\mathbb{R}^{d!}$ vector with all entries to be $\frac{1}{\sqrt{d!}}$ and $(\frac{\mathbf{1}}{\sqrt{d!}})_{\perp} \in \mathbb{R}^{d! \times (d!-1)}$ is an orthogonal complement of $\frac{\mathbf{1}}{\sqrt{d!}}$.
4: Generate independent permutations $\sigma_1, \dots, \sigma_{d-1}$ of $[n]$ uniformly at random and let $\boldsymbol{\mathcal{W}} = \boldsymbol{\mathcal{W}}^{\mathrm{id}, \sigma_1, \dots, \sigma_{d-1}}$.
5: Apply Tensor Reflecting Cloning to $\boldsymbol{\mathcal{W}}$ with ℓ iterations.
6: **Output:** $\boldsymbol{\mathcal{W}}$.

- If $\alpha \ge \frac{1}{2}$ (α is defined in (A2)), that is, in the dense cluster case, let $\ell = \lceil \frac{2}{d}\beta \log_2 n \rceil$ and φ be this mapping from Algorithm 8. Suppose $\kappa = \lceil n^{\gamma} \rceil$ in $\mathrm{HPC}_D(n, \frac{1}{2}, \kappa)$, then after mapping, the sparsity and signal strength in (A2) of $\mathrm{ROHC}(\mathbf{n}, \mathbf{k}, \mu)$ model satisfies

$$\lim_{n\to\infty} \frac{\log(\mu/k^{\frac{d}{2}})^{-1}}{\log n} = \frac{\frac{d}{2}(\frac{2}{d}\beta + \gamma)\log n - \frac{d}{2}\gamma \log n}{\log n} = \beta,$$

$$\lim_{n\to\infty} \frac{\log k}{\log n} = \frac{2}{d}\beta + \gamma =: \alpha.$$

  If $\beta > (\alpha - \frac{1}{2})\frac{d}{2}$, there exists a sequence of polynomial-time tests $\{\phi_n\}$ such that $\liminf_{n\to\infty} \mathcal{E}_{\mathrm{ROHC}_D}(\phi_n) < \frac{1}{2}$. Then by Lemmas 2 and 11 in Luo and Zhang (2022) Section C, we have $\liminf_{n\to\infty} \mathcal{E}_{\mathrm{HPC}_D}(\phi_n \circ \varphi) < \frac{1}{2}$, that is, $\phi_n \circ \varphi$ has asymptotic risk less than to $\frac{1}{2}$ in HPC detection. On the other hand, the size of the planted clique in HPC satisfies $\lim_{n\to\infty} \frac{\log \kappa}{\log n} = \gamma = \alpha - \frac{2}{d}\beta < \alpha - (\alpha - \frac{1}{2}) = \frac{1}{2}$. The combination of these two facts contradicts HPC detection Conjecture 1, so we conclude there are no polynomial-time tests $\{\phi_n\}$ that make $\liminf_{n\to\infty} \mathcal{E}_{\mathrm{ROHC}_D}(\phi_n) < \frac{1}{2}$ if $\beta > (\alpha - \frac{1}{2})\frac{d}{2}$.
- If $0 < \alpha < \frac{1}{2}$, that is, in the sparse cluster case, since $\mathrm{CHC}_D(\mathbf{k}, \mathbf{n}, \lambda)$ is a special case of $\mathrm{ROHC}_D(\mathbf{k}, \mathbf{n}, \mu)$ with $\mu = \lambda k^{d/2}$, the computational lower bound in CHC_D in Theorem 15 implies that if $\beta > 0$, then $\liminf_{n\to\infty} \mathcal{E}_{\mathrm{ROHC}_D}(\phi_n) \ge \frac{1}{2}$ based on HPC Conjecture 1.

In summary, we conclude if $\beta > (\alpha - \frac{1}{2})\frac{d}{2} \vee 0 := \beta^c_{\mathrm{ROHC}_D}$, any sequence of polynomial-time tests has asymptotic risk at least $1/2$ for $\mathrm{ROHC}_D(\mathbf{n}, \mathbf{k}, \mu)$. This has finished the proof of computational lower bound for ROHC_D.

Next, we show the computational lower bound for ROHC_R. Suppose there is a sequence of polynomial-time recovery algorithm $\{\phi_R\}_n$ such that $\liminf_{n\to\infty} \mathcal{E}_{\mathrm{ROHC}_R}(\phi_R) < \frac{1}{2}$ when $\beta >$

$(\alpha-\frac{1}{2})\frac{d}{2}\vee 0$. In this regime, it is easy to verify $\mu \geq Ck^{\frac{d}{4}}$ for some $C>0$ in $\mathrm{ROHC}_D(\mathbf{n},\mathbf{k},\mu)$. By Lemma 10 in Luo and Zhang (2022) Section B, we know there is a sequence polynomial-time detection algorithms $\{\phi_D\}_n$ such that $\liminf_{n\to\infty}\mathcal{E}_{\mathrm{ROHC}_D}(\phi_D)<\frac{1}{2}$, which contradicts the computational lower bound established in the first part. This has finished the proof of the computational lower bound for ROHC_R.

6. Discussion and future work. In this paper, we study the statistical and computational limits of tensor clustering with planted structures, including the constant high-order structure (CHC) and rank-one high-order structure (ROHC). We derive tight statistical lower bounds and tight computational lower bounds under the HPC/HPDS conjectures for both high-order cluster detection and recovery problems. For each problem, we also provide unconstrained-time algorithms and polynomial-time algorithms that respectively achieve these statistical and computational limits. The main results of this paper are summarized in the phase transition diagrams in Figure 1 and Table 1.

There are a few directions worth exploring in the future. First, this paper mainly focuses on the full high-order clustering in the sense that the signal tensor is sparse along all modes. In practice, the partial cluster also commonly appears (e.g., tensor biclustering (Feizi, Javadi and Tse (2017))), where the signal is sparse only in part of the modes. It is interesting to investigate the statistical and computational limits for high-order partial clustering. Second, in addition to the exact recovery discussed in this paper, we think our results can be extended to other variants of recovery, such as partial recovery and weak recovery. Third, in the ROHC model, the nonzero components of the signal are required to have the similar magnitudes as this assumption is essential for support recovery. Another interesting problem is to *estimate* $(\mathbf{v}_1,\dots,\mathbf{v}_d)$ *without the constraint on the component magnitudes of the signal*, which can be seen a rank-one case of the sparse tensor SVD/PCA problem (Niles-Weed and Zadik (2020), Sun et al. (2017), Zhang and Han (2019)). For this problem, the signal-to-noise ratio lower bounds we established in Theorems 10 and 17 still hold by virtue of the estimation-to-detection reduction. However, the ROHC_R Search and Power-iteration algorithms studied in this paper may no longer be suitable for estimating $(\mathbf{v}_1,\dots,\mathbf{v}_d)$. A natural unconstrained-time estimator is the maximum likelihood estimator, while to our best knowledge its guarantee is unexplored. Zhang and Han (2019) developed efficient algorithms which can achieve the minimax optimal error rate in sparse tensor estimation. However, it is unclear if the required signal-to-noise in Zhang and Han (2019) is tight. It is interesting to develop algorithms with optimal guarantees for sparse tensor SVD/PCA under the tight signal-to-noise ratio requirement. Finally, since our computational lower bounds of CHC and ROHC are based on HPC conjecture (Conjecture 1) and HPDS conjecture (Conjecture 2), it is interesting to provide more evidence for these conjectures.

Acknowledgment. We would like to thank Guy Bresler for the helpful discussions. We also thank the Editor, Associate Editor, and two anonymous referees for their helpful suggestions, which helped improve the presentation and quality of this paper.

Funding. This work was supported in part by NSF Grant CAREER-1944904, NSF Grants DMS-1811868 and DMS-2023239, NIH Grant R01 GM131399, and Wisconsin Alumni Research Foundation (WARF).

SUPPLEMENTARY MATERIAL

Supplement to "Tensor clustering with planted structures: Statistical optimality and computational limits" (DOI: 10.1214/21-AOS2123SUPP; .pdf). The supplementary materials contain all technical proofs of this paper.

REFERENCES

ALON, N., KRIVELEVICH, M. and SUDAKOV, B. (1998). Finding a large hidden clique in a random graph. *Random Structures Algorithms* **13** 457–466.

ALON, N., ANDONI, A., KAUFMAN, T., MATULEF, K., RUBINFELD, R. and XIE, N. (2007). Testing k-wise and almost k-wise independence. In *STOC'*07*—Proceedings of the* 39*th Annual ACM Symposium on Theory of Computing* 496–505. ACM, New York. MR2402475 https://doi.org/10.1145/1250790.1250863

AMAR, D., YEKUTIELI, D., MARON-KATZ, A., HENDLER, T. and SHAMIR, R. (2015). A hierarchical Bayesian model for flexible module discovery in three-way time-series data. *Bioinformatics* **31** i17–i26.

AMES, B. P. W. and VAVASIS, S. A. (2011). Nuclear norm minimization for the planted clique and biclique problems. *Math*. *Program*. **129** 69–89. MR2831403 https://doi.org/10.1007/s10107-011-0459-x

ANANDKUMAR, A., GE, R. and JANZAMIN, M. (2014). Guaranteed non-orthogonal tensor decomposition via alternating rank-1 updates. ArXiv preprint. Available at arXiv:1402.5180.

ARIAS-CASTRO, E. and VERZELEN, N. (2014). Community detection in dense random networks. *Ann*. *Statist*. **42** 940–969. MR3210992 https://doi.org/10.1214/14-AOS1208

ARORA, S., BARAK, B., BRUNNERMEIER, M. and GE, R. (2011). Computational complexity and information asymmetry in financial products. *Commun*. *ACM* **54** 101–107.

AWASTHI, P., CHARIKAR, M., LAI, K. A. and RISTESKI, A. (2015). Label optimal regret bounds for online local learning. In *Conference on Learning Theory* 150–166.

BALAKRISHNAN, S., KOLAR, M., RINALDO, A., SINGH, A. and WASSERMAN, L. (2011). Statistical and computational tradeoffs in biclustering. In *NIPS* 2011 *Workshop on Computational Trade-Offs in Statistical Learning* **4**.

BANDEIRA, A. S., KUNISKY, D. and WEIN, A. S. (2020). Computational hardness of certifying bounds on constrained PCA problems. *Innov*. *Theor*. *Comp*. *Sci*.

BARAK, B. and MOITRA, A. (2016). Noisy tensor completion via the sum-of-squares hierarchy. In *Conference on Learning Theory* 417–445.

BARAK, B., HOPKINS, S., KELNER, J., KOTHARI, P. K., MOITRA, A. and POTECHIN, A. (2019). A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM J*. *Comput*. **48** 687–735. MR3945259 https://doi.org/10.1137/17M1138236

BERTHET, Q. and BALDIN, N. (2020). Statistical and computational rates in graph logistic regression. In *International Conference on Artificial Intelligence and Statistics* 2719–2730.

BERTHET, Q. and RIGOLLET, P. (2013a). Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory* 1046–1066.

BERTHET, Q. and RIGOLLET, P. (2013b). Optimal detection of sparse principal components in high dimension. *Ann*. *Statist*. **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127

BOLLOBÁS, B. and ERDŐS, P. (1976). Cliques in random graphs. *Math*. *Proc*. *Cambridge Philos*. *Soc*. **80** 419–427. MR0498256 https://doi.org/10.1017/S0305004100053056

BRENNAN, M. and BRESLER, G. (2019a). Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness. In *Conference on Learning Theory* 469–470. PMLR.

BRENNAN, M. and BRESLER, G. (2019b). Average-case lower bounds for learning sparse mixtures, robust estimation and semirandom adversaries. ArXiv preprint. Available at arXiv:1908.06130.

BRENNAN, M. and BRESLER, G. (2020). Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory* 648–847. PMLR.

BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2018). Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference on Learning Theory* 48–166. PMLR.

BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2019). Universality of computational lower bounds for submatrix detection. In *Conference on Learning Theory* 417–468. PMLR.

BUSYGIN, S., PROKOPYEV, O. and PARDALOS, P. M. (2008). Biclustering in data mining. *Comput*. *Oper*. *Res*. **35** 2964–2987. MR2586410 https://doi.org/10.1016/j.cor.2007.01.005

BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. MR3160567 https://doi.org/10.3150/12-BEJ470

BUTUCEA, C., INGSTER, Y. I. and SUSLINA, I. A. (2015). Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM Probab*. *Stat*. **19** 115–134. MR3374872 https://doi.org/10.1051/ps/2014017

CAI, T. T., LIANG, T. and RAKHLIN, A. (2017). Computational and statistical boundaries for submatrix localization in a large noisy matrix. *Ann*. *Statist*. **45** 1403–1430. MR3670183 https://doi.org/10.1214/16-AOS1488

CAI, T. T. and WU, Y. (2020). Statistical and computational limits for sparse matrix detection. *Ann*. *Statist*. **48** 1593–1614. MR4124336 https://doi.org/10.1214/19-AOS1860

CANDOGAN, U. O. and CHANDRASEKARAN, V. (2018). Finding planted subgraphs with few eigenvalues using the Schur–Horn relaxation. *SIAM J*. *Optim*. **28** 735–759. MR3775143 https://doi.org/10.1137/16M1075144

CHANDRASEKARAN, V. and JORDAN, M. I. (2013). Computational and statistical tradeoffs via convex relaxation. *Proc. Natl. Acad. Sci. USA* **110** E1181–E1190. MR3047651 https://doi.org/10.1073/pnas.1302293110

CHARIKAR, M., NAAMAD, Y. and WU, J. (2018). On finding dense common subgraphs. ArXiv preprint. Available at arXiv:1802.06361.

CHEN, Y. (2015). Incoherence-optimal matrix completion. *IEEE Trans. Inf. Theory* **61** 2909–2923. MR3342311 https://doi.org/10.1109/TIT.2015.2415195

CHEN, Y. and XU, J. (2016). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *J. Mach. Learn. Res.* **17** 882–938.

CHI, E. C., ALLEN, G. I. and BARANIUK, R. G. (2017). Convex biclustering. *Biometrics* **73** 10–19. MR3632347 https://doi.org/10.1111/biom.12540

CHLAMTÁČ, E., DINITZ, M. and KRAUTHGAMER, R. (2012). Everywhere-sparse spanners via dense subgraphs. In 2012 *IEEE* 53*rd Annual Symposium on Foundations of Computer Science—FOCS* 2012 758–767. IEEE Computer Soc., Los Alamitos, CA. MR3186664

CHLAMTÁČ, E., DINITZ, M. and MAKARYCHEV, Y. (2017). Minimizing the union: Tight approximations for small set bipartite vertex expansion. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* 881–899. SIAM, Philadelphia, PA. MR3627785 https://doi.org/10.1137/1.9781611974782.56

CHLAMTÁC, E. and MANURANGSI, P. (2018). Sherali-Adams integrality gaps matching the log-density threshold. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* (*APPROX/RANDOM* 2018) Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

CICHOCKI, A., MANDIC, D., DE LATHAUWER, L., ZHOU, G., ZHAO, Q., CAIAFA, C. and PHAN, H. A. (2015). Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Process. Mag.* **32** 145–163.

DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000a). A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21** 1253–1278. MR1780272 https://doi.org/10.1137/S0895479896305696

DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000b). On the best rank-1 and rank-$(R_1, R_2, \dots, R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* **21** 1324–1342. MR1780276 https://doi.org/10.1137/S0895479898346995

DEKEL, Y., GUREL-GUREVICH, O. and PERES, Y. (2014). Finding hidden cliques in linear time with high probability. *Combin. Probab. Comput.* **23** 29–49. MR3197965 https://doi.org/10.1017/S096354831300045X

DESHPANDE, Y. and MONTANARI, A. (2015a). Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Found. Comput. Math.* **15** 1069–1128. MR3371378 https://doi.org/10.1007/s10208-014-9215-y

DESHPANDE, Y. and MONTANARI, A. (2015b). Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Conference on Learning Theory* 523–562.

DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2017). Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures (extended abstract). In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 73–84. IEEE Computer Soc., Los Alamitos, CA. MR3734219 https://doi.org/10.1109/FOCS.2017.16

DIAKONIKOLAS, I., KONG, W. and STEWART, A. (2019). Efficient algorithms and lower bounds for robust linear regression. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* 2745–2754. SIAM, Philadelphia, PA. MR3909639 https://doi.org/10.1137/1.9781611975482.170

DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2019). Subexponential-time algorithms for sparse PCA. ArXiv preprint. Available at arXiv:1907.11635.

DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2020). The average-case time complexity of certifying the restricted isometry property. ArXiv preprint. Available at arXiv:2005.11270.

DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.

DUDEJA, R. and HSU, D. (2021). Statistical query lower bounds for tensor PCA. *J. Mach. Learn. Res.* **22** 83. MR4253776

FAN, J., LIU, H., WANG, Z. and YANG, Z. (2018). Curse of heterogeneity: Computational barriers in sparse mixture models and phase retrieval. ArXiv preprint. Available at arXiv:1808.06996.

FAUST, K., SATHIRAPONGSASUTI, J. F., IZARD, J., SEGATA, N., GEVERS, D., RAES, J. and HUTTENHOWER, C. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* **8** e1002606.

FEIGE, U. and KRAUTHGAMER, R. (2000). Finding and certifying a large hidden clique in a semirandom graph. *Random Structures Algorithms* **16** 195–208. MR1742351 https://doi.org/10.1002/(SICI)1098-2418(200003)16:2<195::AID-RSA5>3.3.CO;2-1

FEIZI, S., JAVADI, H. and TSE, D. (2017). Tensor biclustering. In *Advances in Neural Information Processing Systems* 1311–1320.

FELDMAN, V., PERKINS, W. and VEMPALA, S. (2018). On the complexity of random satisfiability problems with planted solutions. *SIAM J. Comput.* **47** 1294–1338. MR3827195 https://doi.org/10.1137/16M1078471

FELDMAN, V., GRIGORESCU, E., REYZIN, L., VEMPALA, S. S. and XIAO, Y. (2017). Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM* **64** 8. MR3664576 https://doi.org/10.1145/3046674

FLORES, G. E., CAPORASO, J. G., HENLEY, J. B., RIDEOUT, J. R., DOMOGALA, D., CHASE, J., LEFF, J. W., VÁZQUEZ-BAEZA, Y., GONZALEZ, A. et al. (2014). Temporal variability is a personalized feature of the human microbiome. *Genome Biol*. **15** 531. https://doi.org/10.1186/s13059-014-0531-y

GAMARNIK, D. and SUDAN, M. (2014). Limits of local algorithms over sparse random graphs [extended abstract]. In *ITCS'*14*—Proceedings of the* 2014 *Conference on Innovations in Theoretical Computer Science* 369–375. ACM, New York. MR3359490

GAMARNIK, D. and ZADIK, I. (2019). The landscape of the planted clique problem: Dense subgraphs and the overlap gap property. ArXiv preprint. Available at arXiv:1904.07174.

GAO, C., MA, Z. and ZHOU, H. H. (2017). Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist*. **45** 2074–2101. MR3718162 https://doi.org/10.1214/16-AOS1519

GHOSHDASTIDAR, D. and DUKKIPATI, A. (2017). Consistency of spectral hypergraph partitioning under planted partition model. *Ann. Statist*. **45** 289–315. MR3611493 https://doi.org/10.1214/16-AOS1453

HAJEK, B., WU, Y. and XU, J. (2015). Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory* 899–928.

HAJEK, B., WU, Y. and XU, J. (2016). Information limits for recovering a hidden community. In 2016 *IEEE International Symposium on Information Theory* (*ISIT*) 1894–1898. IEEE, Los Alamitos.

HAN, R., WILLETT, R. and ZHANG, A. (2020). An optimal statistical and computational framework for generalized tensor estimation. ArXiv preprint. Available at arXiv:2002.11255.

HENRIQUES, R. and MADEIRA, S. C. (2019). Triclustering algorithms for three-dimensional data analysis: A comprehensive survey. *ACM Comput. Surv.* **51** 95.

HILLAR, C. J. and LIM, L.-H. (2013). Most tensor problems are NP-hard. *J. ACM* **60** 45. MR3144915 https://doi.org/10.1145/2512329

HOPKINS, S. B. K. (2018). Statistical inference and the sum of squares method.

HOPKINS, S. B., SHI, J. and STEURER, D. (2015). Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory* 956–1006.

HOPKINS, S. B. and STEURER, D. (2017). Efficient Bayesian estimation from few samples: Community detection and related problems. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 379–390. IEEE Computer Soc., Los Alamitos, CA. MR3734245 https://doi.org/10.1109/FOCS.2017.42

HOPKINS, S. B., KOTHARI, P. K., POTECHIN, A., RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2017). The power of sum-of-squares for detecting hidden structures. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 720–731. IEEE Computer Soc., Los Alamitos, CA. MR3734275 https://doi.org/10.1109/FOCS.2017.72

JERRUM, M. (1992). Large cliques elude the Metropolis process. *Random Structures Algorithms* **3** 347–359. MR1179827 https://doi.org/10.1002/rsa.3240030402

JI, L., TAN, K.-L. and TUNG, A. K. (2006). Mining frequent closed cubes in 3D datasets. In *Proceedings of the* 32*nd International Conference on Very Large Data Bases* 811–822. VLDB Endowment.

JIANG, D., PEI, J., RAMANATHAN, M., TANG, C. and ZHANG, A. (2004). Mining coherent gene clusters from gene-sample-time microarray data. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 430–439. ACM, New York.

KANNAN, R. and VEMPALA, S. (2017). The hidden hubs problem. In *Conference on Learning Theory* 1190–1213.

KIM, C., BANDEIRA, A. S. and GOEMANS, M. X. (2017). Community detection in hypergraphs, spiked tensor models, and sum-of-squares. In 2017 *International Conference on Sampling Theory and Applications* (*SampTA*) 124–128. IEEE, Los Alamitos.

KOIRAN, P. and ZOUZIAS, A. (2014). Hidden cliques and the certification of the restricted isometry property. *IEEE Trans. Inf. Theory* **60** 4999–5006. MR3245368 https://doi.org/10.1109/TIT.2014.2331341

KOLAR, M., BALAKRISHNAN, S., RINALDO, A. and SINGH, A. (2011). Minimax localization of structural information in large noisy matrices. In *Advances in Neural Information Processing Systems* 909–917.

KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056 https://doi.org/10.1137/07070111X

KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2019). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. ArXiv preprint. Available at arXiv:1907.11636.

LESIEUR, T., MIOLANE, L., LELARGE, M., KRZAKALA, F. and ZDEBOROVÁ, L. (2017). Statistical and computational phase transitions in spiked tensor estimation. In 2017 *IEEE International Symposium on Information Theory* (*ISIT*) 511–515. IEEE, Los Alamitos.

LI, A. and TUCK, D. (2009). An effective tri-clustering algorithm combining expression data with gene regulation information. *Gene Regul. Syst. Biol.* **3** 49–64. https://doi.org/10.4137/grsb.s1150

LÖFFLER, M., WEIN, A. S. and BANDEIRA, A. S. (2020). Computationally efficient sparse clustering. ArXiv preprint. Available at arXiv:2005.10817.

LUO, Y. and ZHANG, A. R. (2020a). Open problem: Average-case hardness of hypergraphic planted clique detection. *Conf. Learn. Theory* **125** 3852–3856.

LUO, Y. and ZHANG, A. R. (2022b). Supplement to "Tensor clustering with planted structures: Statistical optimality and computational limits." https://doi.org/10.1214/21-AOS2123SUPP

MA, T. and WIGDERSON, A. (2015). Sum-of-squares lower bounds for sparse PCA. In *Advances in Neural Information Processing Systems* 1612–1620.

MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. MR3346698 https://doi.org/10.1214/14-AOS1300

MADEIRA, S. C. and OLIVEIRA, A. L. (2004). Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **1** 24–45.

MANKAD, S. and MICHAILIDIS, G. (2014). Biclustering three-dimensional data arrays with plaid models. *J. Comput. Graph. Statist.* **23** 943–965. MR3270705 https://doi.org/10.1080/10618600.2013.851608

MCSHERRY, F. (2001). Spectral partitioning of random graphs. In 42*nd IEEE Symposium on Foundations of Computer Science* (*Las Vegas*, *NV*, 2001) 529–537. IEEE Computer Soc., Los Alamitos, CA. MR1948742

MEKA, R., POTECHIN, A. and WIGDERSON, A. (2015). Sum-of-squares lower bounds for planted clique [extended abstract]. In *STOC'*15*—Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 87–96. ACM, New York. MR3388186

MONTANARI, A. (2015). Finding one community in a sparse graph. *J. Stat. Phys.* **161** 273–299. MR3401018 https://doi.org/10.1007/s10955-015-1338-2

NILES-WEED, J. and ZADIK, I. (2020). The all-or-nothing phenomenon in sparse tensor PCA. *Adv. Neural Inf. Process. Syst.* **33**.

PERRY, A., WEIN, A. S. and BANDEIRA, A. S. (2020). Statistical limits of spiked tensor models. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 230–264. MR4058987 https://doi.org/10.1214/19-AIHP960

RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2018). High-dimensional estimation via sum-of-squares proofs. ArXiv preprint. Available at arXiv:1807.11419.

RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems* 2897–2905.

RON, D. and FEIGE, U. (2010). Finding hidden cliques in linear time. *Discrete Math. Theor. Comput. Sci.*.

ROSSMAN, B. (2008). On the constant-depth complexity of k-clique. In *STOC'*08 721–730. ACM, New York. MR2582693 https://doi.org/10.1145/1374376.1374480

ROSSMAN, B. (2014). The monotone complexity of k-clique on random graphs. *SIAM J. Comput.* **43** 256–279. MR3166976 https://doi.org/10.1137/110839059

SCHRAMM, T. and WEIN, A. S. (2020). Computational barriers to estimation from low-degree polynomials. ArXiv preprint. Available at arXiv:2008.02269.

SIM, K., AUNG, Z. and GOPALKRISHNAN, V. (2010). Discovering correlated subspace clusters in 3D continuous-valued data. In 2010 *IEEE International Conference on Data Mining* 471–480. IEEE, Los Alamitos.

SUN, X. and NOBEL, A. B. (2013). On the maximal size of large-average and ANOVA-fit submatrices in a Gaussian random matrix. *Bernoulli* **19** 275–294. MR3019495 https://doi.org/10.3150/11-BEJ394

SUN, W. W., LU, J., LIU, H. and CHENG, G. (2017). Provable sparse tensor decomposition. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 899–916. MR3641413 https://doi.org/10.1111/rssb.12190

TANAY, A., SHARAN, R. and SHAMIR, R. (2002). Discovering statistically significant biclusters in gene expression data. *Bioinformatics* **18** S136–S144.

VERZELEN, N. and ARIAS-CASTRO, E. (2015). Community detection in sparse random networks. *Ann. Appl. Probab.* **25** 3465–3510. MR3404642 https://doi.org/10.1214/14-AAP1080

WANG, T., BERTHET, Q. and PLAN, Y. (2016). Average-case hardness of RIP certification. In *Advances in Neural Information Processing Systems* 3819–3827.

WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. MR3546438 https://doi.org/10.1214/15-AOS1369

WANG, M., FISCHER, J. and SONG, Y. S. (2019). Three-way clustering of multi-tissue multi-individual gene expression data using semi-nonnegative tensor decomposition. *Ann. Appl. Stat.* **13** 1103–1127. MR3963564 https://doi.org/10.1214/18-AOAS1228

WANG, Z., GU, Q. and LIU, H. (2015). Sharp computational-statistical phase transitions via oracle computational model. ArXiv preprint. Available at arXiv:1512.08861.

WEIN, A. S., EL ALAOUI, A. and MOORE, C. (2019). The Kikuchi hierarchy and tensor PCA. In 2019 *IEEE* 60*th Annual Symposium on Foundations of Computer Science* 1446–1468. IEEE Comput. Soc. Press, Los Alamitos, CA. MR4228236

WU, Y. and XU, J. (2021). Statistical problems with planted structures: Information-theoretical and computational limits. *Inf. Theor. Methods Data Sci.* 383.

XIA, D. and ZHOU, F. (2019). The sup-norm perturbation of HOSVD and low rank tensor denoising. *J. Mach. Learn. Res.* **20** 61. MR3960915

ZHANG, T. and GOLUB, G. H. (2001). Rank-one approximation to high order tensors. *SIAM J. Matrix Anal. Appl.* **23** 534–550. MR1871328 https://doi.org/10.1137/S0895479899352045

ZHANG, A. and HAN, R. (2019). Optimal sparse singular value decomposition for high-dimensional high-order data. *J. Amer. Statist. Assoc.* **114** 1708–1725. MR4047294 https://doi.org/10.1080/01621459.2018.1527227

ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377