Statistical and computational thresholds for the planted k-densest sub-hypergraph problem

Luca Corinzia
Department of Computer Science
ETH Zürich, Switzerland
Email: luca.corinzia@inf.ethz.ch

Paolo Penna
Department of Computer Science
ETH Zürich, Switzerland
Email: paolo.penna@inf.ethz.ch

Wojciech Szpankowski
Department of Computer Science
Purdue University, USA
Email: szpan@purdue.edu

Joachim M. Buhmann
Department of Computer Science
ETH Zürich, Switzerland
Email: jbuhmann@inf.ethz.ch

Abstract

In this work, we consider the problem of recovery a planted k-densest sub-hypergraph on a d-uniform hypergraph. This fundamental problem appears in different contexts, e.g., community detection, average-case complexity, and neuroscience applications as a *structural* variant of *tensor* PCA problem. We provide tight statistical upper and lower bounds for the exact recovery threshold by the maximum-likelihood estimator, as well as algorithmic bounds based on approximate message passing algorithms. The problem exhibits a typical *statistical-computational gap* observed in analogous *sparse* settings that widens with increasing sparsity of the problem. The bounds show that the signal structure impacts the location of the statistical and computational phase transition not captured by the known existing bounds for the tensor PCA model. This effect is due to the generic planted signal prior that this latter model addresses.

1 INTRODUCTION

High dimensional inference problems play a key role in recent machine learning and data analysis applications. Typical scenarios exhibit problem dimensions that are comparable to the sample size, hence precluding effective estimation with no further structure imposed on the underlying signal, such as low rank or sparsity. Examples of such problems include sparse mean estimation, compressive sensing, sparse phase retrieval, low-rank matrix estimation, community detection, planted clique and densest subgraph recovery problems.

In this work we study the problem of recovery a *planted* k-densest sub-hypergraph on d-uniform hypergraphs over p nodes studied in (Corinzia et al., 2019). This is the *planted* version of the classical k-densest subgraph (Chlamtac et al., 2016), informally defined as follows. The *planted* solution consists of a set of k randomly chosen nodes among the total of p nodes. The weights of all $\binom{k}{d}$ hyperedges [1] *inside the planted solution* have some initial *bias*, while the others have zero bias. After some Gaussian *noise* is added to these initial weights, we would like to recover the planted solution given only these noisy observations.

> *Under what conditions is it* possible *to recover the planted solution?*
> *What about computationally* efficient *algorithms for this task?*

In the simpler *graph* setting the problem is closely related to detecting a core structure in community detection, it resembles the well-known planted clique problem, and other community detection models such as the stochastic block model (SBM). Recent work also suggest that the k-densest subgraph problem is related

Proceedings of the 25[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151.

[1] Recall that in an d-uniform hypergraph we have one hyperedge for each subset of d nodes, with $d = 2$ being the standard undirected graph.

to long term memory mechanism in the brain (Legenstein et al., 2018). The more general *hypergraph* version is closely related to the so-called *tensor* Principal Component Analysis (tensor PCA). Higher order interactions among nodes is also rather natural in several applications, including modeling brain regions (Gu et al., 2017; Wang et al., 2012; Zu et al., 2016) and in computer vision applications (Jolion and Kropatsch, 2012).

Our goal is to understand the *information-theoretic* and *computational* algorithmic limits for this class of recovery problems parameterized by d and k, where these parameters possibly depend on the number of nodes p. This will establish the regimes for which algorithms have the ability to recover a hidden structure or signal from noisy measurements and partial information. In particular, we are interested in the existence of the so called *statistical-computation gaps* observed in several problems (see Section 1.2 below). Note that the *maximum likelihood estimator (MLE)* is the optimal estimator for the 0-1 loss (Jagannath et al., 2020), and thus it characterizes the statistical limits of the problem (whether recovery is possible). For a start, consider the two extremes, $d = 1$ and $d = k$, where the MLE is *computationally efficient* though for *different* reasons:

- For $d = k$ the solutions[2] are statistically *independent* (like in the Random Energy Model by Derrida (1981)). The MLE can be implemented by searching exhaustively through all $\binom{p}{k} = \binom{p}{d}$ solutions, which is computationally efficient since $\binom{p}{d}$ is the input size.
- For $d = 1$ exhaustive search through all $\binom{p}{k}$ solutions is clearly inefficient, unless k is constant. Nonetheless simply selecting greedily the k nodes with highest weights implements efficiently the MLE. Unlike the previous case, the solutions' weights are statistically *dependent*.

Thus, in both these extreme cases there is *no statistical-computation gap*, while their information-theoretic nature is very different. The recoverability conditions (whether the MLE succeeds) for $d = 1$ boils down to analyzing $N = \binom{p}{d}$ *independent* Gaussian random variables. The presence of *dependencies* between solutions makes the analysis of the information-theoretic boundary conditions significantly more complex, even for $d = k$.

We are interested in the *non-extreme* cases (the spectrum $1 < d \leq k \leq p$), where most of the practical and theoretical applications arise. These are also the most challenging cases since they present two challenges: information-theoretical due to *dependencies* and algorithmic since the MLE is *not* computationally efficient in general. In particular, we shall focus on the *sparse* regime $k \ll p$ which, in several recovery problems, seems to be the major source of statistical-computation gaps. We mainly consider $k \approx p^{\alpha}$ with arbitrary $\alpha \in (0, 1)$ constant (though some of our results extend to the full spectrum).

[2]Each subset of k nodes is a possible solution, and two solutions are statistically independent if they do not share any random variable, i.e., any hyperedge.

1.1 Our contribution

Despite tensor PCA recovery bounds have been studied in depth, these results either do not yield tight information-theoretical bounds for our problem (because the tensor PCA assumes a more general prior), or they do not apply (because the prior is different from ours).[3] Moreover, the best known information-theoretic bounds for the problem we consider are far from tight (Corinzia et al., 2019).

Our first contribution settles this issue as we provide (essentially) tight information-theoretic recovery bounds. This allows for a direct comparison with existing bounds of more general or different recovery problems studied in the literature (thus highlighting similarities and differences). Intuitively speaking, our bounds depend on the so-called *signal-to-noise ratio (SNR)*, that is, the ratio between the initial *bias* β of the weights in the *planted* solution, and the "overall" noise. We express our results via the following scale-normalized SNR γ,

$$\gamma \equiv \beta \sqrt{\frac{\binom{k}{d}}{k} \cdot \frac{1}{2 \log p}} \ , \tag{1}$$

and prove the following:

Theorem (informal) *Consider the scale-normalized SNR γ and the following two constants,*

$$\begin{aligned} \gamma_{LB} &= \begin{cases} \sqrt{1-\alpha} & \text{for any } d = \omega(1) \\ \sqrt{1-\alpha}/2 & \text{in general} \end{cases} \\ \gamma_{UB} &= \sqrt{1+2\alpha} + \sqrt{\alpha} \ , \end{aligned} \tag{2}$$

where $\alpha \in (0, 1)$ is a constant satisfying $k \approx p^{\alpha}$. For $\gamma < \gamma_{LB}$, exact recovery is impossible *regardless of which estimator (algorithm) we use, while for $\gamma > \gamma_{UB}$ exact recovery is possible via the (computationally intractable) maximum likelihood estimator.*

[3]Intuitively, the planted signal may or may not obey a certain structure which corresponds to a specific prior distribution, e.g., the uniform distribution over all subsets of k nodes as in our problem (see Remark 1 below).

The scale-normalized SNR in (1) has a rather natural interpretation (see Remark 2 below). It is worth to observe that this normalization incorporates the problem parameters p, k, d and that the *information-theoretic recoverability* thresholds in (2) correspond to a *finite* γ. That is, to some initial bias

$$\beta = \beta(p) = \Theta\left(\sqrt{\frac{2k \log p}{\binom{k}{d}}}\right) .$$

Note that both k and d may depend on p, that is, we consider $k = k(p)$ and $h(p)$ for $p \to \infty$.

Our second contribution is a new *algorithmic* threshold based on Approximate Message Passing (AMP) algorithms. This class of algorithms represent the best known technique for several other high dimensional statistical estimation problems (see e.g. (Deshpande and Montanari, 2015; Richard and Montanari, 2014; Lesieur et al., 2017a,b; Wein et al., 2019) and references therein).

Specifically, we provide a heuristic derivation of AMP together with a state evolution analysis for our problem. Our analysis extends the AMP algorithm for tensor PCA with non-factorisable prior distribution, and results in the following non-trivial **computationally-efficient threshold**:

$$\gamma_{\text{AMP}} := \sqrt{\frac{1}{2e}\left(\frac{p}{k}\right)^{d-1}\frac{1}{d(d-1)\log p}} . \qquad (3)$$

Intuitively, for $k \approx p^{\alpha}$ and by ignoring low order terms, the above bound corresponds to

$$\gamma_c^{AMP} \approx \sqrt{p^{(1-\alpha)(d-1)}/(d(d-1)\log p)} ,$$

which obviously grows with p, hence a gap with the information-theoretic threshold γ_{LB} in (2). Our AMP algorithm has time complexity $\Theta(p^d)$, which is the cost of the application of the tensor to a vector, and is the scaling of the elements of the tensor.

The novel element in our AMP algorithm is the use of a *vectorial* threshold function (see Section 3). Intuitively, the threshold function is one of the key components of AMP algorithm, as AMP generalizes in a non-trivial way spectral algorithms which correspond to a particular threshold function. As we also show experimentally (see Section 4), this yields an improvement compared to the scalar threshold functions used in prior works, especially with informative initializations of the algorithm.

We stress that our analysis is heuristic in the sense that in the analysis it introduces certain approximations, which may not hold for all regimes. Indeed, rigorous analysis of the AMP is only available for a handful of problems like compressed sensing (Donoho et al., 2009), planted clique (Deshpande and Montanari, 2015), and matrix PCA in the dense (Deshpande and Montanari, 2014) and sparse (Barbier et al., 2020) regime. To the best of our knowledge, the only other non-rigorous analysis of AMP in tensor case is given in (Lesieur et al., 2017b) for dense tensor PCA with factorized prior. For these reasons, we consider the heuristic derivation of the AMP and its state evolution as one of the major contributions in this paper.

Our experiments show the that empirical behavior of the AMP algorithm matches the analytical threshold γ_{AMP} in (3). Interestingly, when initialized with the correct planted solution, the AMP algorithm (with our vectorial threshold function) approaches the information-theoretic threshold γ_{LB} in (2).

1.2 Related Work

The study of the recovery of a planted signal in a probabilistic generative model has received much attention recently, as it constitutes a fertile ground for the analysis of the *statistical-computational (SC) gaps*. Many variations of the planted problem have been addressed in the literature. The stochastic block model for community detection in graphs (Abbe et al., 2016; Mossel et al., 2015; Chen and Xu, 2014) or hypergraphs (Barak et al., 2016; Ghoshdastidar and Dukkipati, 2014)) has been one of the first model to be studied and does not present any SC gap (Abbe et al., 2016; Kim et al., 2018). Despite these models have been mainly used with discrete Bernoulli random variables, recent extension to *weighted* edges have been proposed (see (Aicher et al., 2014; Peixoto, 2018)). Analogously, the dense matrix-PCA problem (Barbier et al., 2016; Deshpande and Montanari, 2014) has been shown not to have any SC gap, as the proposed AMP algorithms match the statistical thresholds (Deshpande and Montanari, 2014).

SC gaps have been first observed in the context of dense tensor-PCA (Richard and Montanari, 2014; Hopkins et al., 2015; Montanari et al., 2016; Jagannath et al., 2020; Arous et al., 2020; Perry et al., 2020) and recently in the *sparse* matrix (Barbier et al., 2020) and tensor extension (Niles-Weed and Zadik, 2020; Corinzia et al., 2021). The last two works address only the statistical phase transition and are the closest to our work. In (Niles-Weed and Zadik, 2020), the statistical threshold for the MMSE estimator is fully characterized for the sparse tensor-PCA problem, and it is shown that the estimator undergoes an all-or-nothing transition. After proper rescaling, the SNR threshold given in (Niles-Weed and Zadik, 2020) for the MMSE estimator is located exactly at the lower bound threshold proved here in Theorem 2. The characterization is, however, performed for the tensorial (multi-dimensional) MMSE estimator. This estimator is allowed to return any tensor in the unit ball, with no guarantee whether this can be used to recover the *vectorial* planted signal. Moreover, the precise relation between the MMSE and the MLE addressed here is generally still unknown, as mentioned by Niles-Weed and Zadik (2020). Indeed, Corinzia et al. (2021) showed that a stronger condition on the MMSE behavior produces an equivalent transition on the tensorial MLE, which imposes limitations on the sparsity regime for which the results apply.

Other results on tensor-PCA bounds (Richard and Montanari, 2014; Hopkins et al., 2015; Montanari et al., 2016) with generic signal prior are not tight in most of the cases if applied to our problem in a black-box fashion (see, e.g. the comparison with prior upper and lower bounds given in Supplementary Material). This is due to the specific combinatorial structure of the planted vector (that is, in turn, a restriction on the signal prior) that is not specified for the generic spherical prior tensor-PCA. Further variants and extensions that fall into the class of tensor PCA have been considered in (Han et al., 2021; Luo and Zhang, 2021; Brennan and Bresler, 2020). We note that the tensor order d is not constant in our work, unlike in these works. This is crucial when considering the "missing diagonal" entries in our setting that are negligible for $d = O(1)$, but not in our setting (for $d = k$ our problem is equivalent to the independent Gaussian problem, unlike the tensor counterpart). The technique in (Brennan and Bresler, 2020) based on a "filling entries" argument is performed on (i) tensor PCA with a *specific* prior (Rademacher, that is, dense and factorisable), which seems to influence the statistical/computational thresholds, and (ii) negatively correlated sparse PCA in the regime $k = o(p^{1/6})$. The other two referenced papers study different priors, which makes these results/techniques not directly applicable. The algorithmic guarantees for a more general prior are suboptimal for our problem, and the recovery threshold is lower.

Regarding the computational thresholds, the algorithm here described is an extension of the tensor-PCA AMP algorithm to sparse hypergraph settings. These algorithms have been shown to be information-theoretic optimal in numerous high dimensional statistical estimation problems (dense matrix-PCA (Deshpande and Montanari, 2014), SBM (Abbe and Sandon, 2018) etc.) and to outperform other class of algorithms in the case of SC gaps (see e.g., planted clique problem (Deshpande and Montanari, 2015) and sparse matrix-PCA (Barbier et al., 2020)). Nonetheless, AMP algorithms have been shown to underperform in the tensor-PCA problem (Lesieur et al., 2017b) to the sum-of-squares class (Hopkins et al., 2015) and recently to averaged gradient descent (Biroli et al., 2020). A recent work suggests that a hierarchy of such AMP algorithms may actually match the performance of the best known efficient algorithms (Wein et al., 2019). Finally, Choo and d'Orsi (2021) provides a sum-of-squares type algorithm for the tensorial version of our problem when $k = O(\sqrt{p})$ and whose running time is polynomial for *constant* d provided the SNR γ is large enough (their algorithm applies to the full spectrum and its running time increases for smaller γ).

1.3 Model and formal definitions

We study the problem of recovery a *planted* sub-hypergraph on a d-uniform hypergraph over p nodes. Every subset of d nodes $\{i_1, i_2, \ldots, i_d\}$ is an *hyperedge* whose weight $Y_{i_1,i_2,\ldots,i_d}$ is a Gaussian random variable, defined according to the following process. We denote by $\mathbf{x} \in \mathcal{C}_{p,k} \subset \{0,1\}^p$ the vector of selected nodes, with

$$\mathcal{C}_{p,k} \coloneqq \{\mathbf{x} \in \{0,1\}^p \colon \sum_i x_i = k\}.$$

Furthermore, we assume that $\mathbf{x}$ is drawn uniformly at random in $\mathcal{C}_{p,k}$ with probability $\mathbb{P}_p$. The resulting weights are given by the d-order tensor $\mathbf{Y} \coloneqq \mathbf{Y}(\mathbf{x})$ in which all edges indicated by $\mathbf{x}$ have a *bias* $\beta \geq 0$, and the weights are perturbed by adding Gaussian *noise* across all hyperedges $(i_1 i_2 \cdots i_d)$ with $i_1 < i_2 < \cdots < i_d$.

For convenience let us now introduce the following tensor notation. The outer product of two tensors $\mathbf{U} \in \bigotimes^{d_1} \mathbb{R}^p$ and $\mathbf{V} \in \bigotimes^{d_2} \mathbb{R}^p$ is denoted by $\mathbf{U} \otimes \boldsymbol{V}$ with entries $(\mathbf{U} \otimes \boldsymbol{V})_{i_1 i_2 \ldots i_{d_1} j_1 j_2 \cdots j_{d_2}} = \mathbf{U}_{i_1 i_2 \cdots i_{d_1}} \mathbf{V}_{j_1 j_2 \ldots j_{d_2}}$. For $\boldsymbol{x} \in \mathbb{R}^p$, we define $\boldsymbol{x}^{\otimes d} = \boldsymbol{x} \otimes \cdots \otimes \boldsymbol{x} \in \bigotimes^d \mathbb{R}^p$ as the d-th outer power of $\boldsymbol{x}$. The inner product of the two tensors $\mathbf{U} \in \bigotimes^{d_1} \mathbb{R}^p$ and $\mathbf{V} \in \bigotimes^{d_2} \mathbb{R}^p$ with $d_2 \leq d_1$ is defined as $\langle \mathbf{U}, \mathbf{V} \rangle = \sum_{j_1,\ldots,j_{d_2}} \mathbf{U}_{i_1,\ldots,i_{d_2-d_1},j_1,\ldots,j_{d_2}} \mathbf{V}_{j_1,\ldots,j_{d_2}}$. Given a d-th order tensor $\mathbf{U} \in \bigotimes^d \mathbb{R}^p$, we define the map $\mathbf{U} : \mathbb{R}^p \to \mathbb{R}^p$ as

$$\boldsymbol{U}\{\boldsymbol{x}\}_i = \sum_{i_2,\ldots,i_d} U_{i,i_2,\ldots,i_d}\ x_{i_2} \cdot \ldots \cdot x_{i_d}. \tag{4}$$

Using tensor notation, the observation model reads:

$$\mathbf{Y} = (\beta \mathbf{x}^{\otimes d} + \mathbf{Z}) \mathbb{1}_{\{i_1 < \cdots < i_d\}} \tag{5}$$

where $\mathbb{1}_{\{P\}} = 1$ if P is true, and 0 otherwise, and $Z_{i_1 i_2 \cdots i_h} \sim \mathcal{N}(0,1)$.

Remark 1. *Note the main important difference with the tensor PCA formulation of the problem, where all the elements of the tensor are observed. Also note that* $\mathbf{x}$ *consists of randomly chosen vector with exactly* k

entries equals 1*, which gives a particular prior distribution on the tensor in* (5)*. Different priors have been considered like, e.g., the Rademacher prior where* $\mathbf{x}$ *is chosen uniformly in the set* $\{-1, +1\}^p$*. The more general (unrestricted) prior in tensor PCA corresponds to the tensor*

$$\mathbf{Y} = \beta\mathbf{X} + \mathbf{Z}$$

where $\mathbf{X} \in \mathbb{R}^N$ *is a tensor chosen uniformly among all tensors of unit length, and* $\mathbf{Z} \in \mathbb{R}^N$ *has i.i.d. standard Guaussian entries similarly to* (5)*.*

For any signal $\mathbf{x}$, we also consider the sum of all $\binom{k}{d}$ weights of the hyperedges with nodes in $\mathbf{x}$ as:

$$\begin{aligned} S(\mathbf{x}) &:= \sum_{i_1<i_2<\ldots<i_d} Y_{i_1,i_2\ldots,i_d} x_{i_1} \cdot x_{i_2} \cdot \cdots \cdot x_{i_d} \\ &= \sum_i \mathbf{Y}\{\mathbf{x}\}_i = \langle \mathbf{Y}, \mathbf{x}^{\otimes d} \rangle \end{aligned} \tag{6}$$

Definition 1 (Partial and exact receovery)**.** *A* k'*-partial recovery is achieved if there exists an estimator* $\hat{\mathbf{x}}$ *that, with input the weight tensor* $\mathbf{Y}(\mathbf{x})$ *given by* (5)*, returns* $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\mathbf{Y})$ *such that*

$$\mathbb{P}(\langle \hat{\mathbf{x}}, \mathbf{x} \rangle \geq k') = 1 - o(1).$$

Exact recovery is achieved if

$$\mathbb{P}(\langle \hat{\mathbf{x}}, \mathbf{x} \rangle = k) = \mathbb{P}(\hat{\mathbf{x}} = \mathbf{x}) = 1 - o(1).$$

Definition 2 (Maximum-likelihood estimator)**.** *The vectorial maximum-likelihood estimator is defined as*

$$\mathbf{x}_{\mathrm{MLE}}(\mathbf{Y}) = \operatorname*{argmax}_{\hat{\mathbf{x}}\colon \mathbf{Y}\to\hat{\mathbf{x}}(\mathbf{Y})\in\mathcal{C}_{p,k}} \mathbb{P}(\mathbf{Y}|\mathbf{x}(\hat{\mathbf{Y}}))$$

We define as $P_r^{(k')} = \mathbb{P}(\langle \mathbf{x}_{\mathrm{MLE}}, \mathbf{x} \rangle \geq k')$ *and* P_r *accordingly.*

It is easy to see that the vectorial MLE estimator for the problem in Equation (5) corresponds to the k-densest sub-hypergraph (from this the name of the problem, see Theorem 4 in (Corinzia et al., 2019) for a proof)

$$\begin{aligned} \mathbf{x}_{\mathrm{MLE}}(\mathbf{Y}) &= \operatorname*{argmax}_{\hat{\mathbf{x}}\colon \mathbf{Y}\to\hat{\mathbf{x}}(\mathbf{Y})\in\mathcal{C}_{p,k}} \sum_{i_1<\cdots<i_d} Y_{i_1,\ldots,i_d} \hat{x}_{i_1} \cdot \cdots \cdot \hat{x}_{i_d} \\ &= \operatorname*{argmax}_{\hat{\mathbf{x}}\colon \mathbf{Y}\to\hat{\mathbf{x}}(\mathbf{Y})\in\mathcal{C}_{p,k}} \langle \mathbf{Y}, \hat{\mathbf{x}}^{\otimes d} \rangle = \operatorname*{argmax}_{\hat{\mathbf{x}}\colon \mathbf{Y}\to\hat{\mathbf{x}}(\mathbf{Y})\in\mathcal{C}_{p,k}} S(\hat{\mathbf{x}}) \end{aligned} \tag{7}$$

Our bounds depend on the scale-normalized SNR:

$$\gamma := \beta\sqrt{\frac{\binom{k}{d}}{k} \cdot \frac{1}{2\log p}}. \tag{8}$$

This scaling incorporates the parameters k and d of the problem and it will result in information-theoretic thresholds located at finite values.

Remark 2. *The scale-normalized SNR* γ *in Equation* (8) *can be seen as the effective SNR of the problem, given by* total signal / total noise*. The total signal is* β*, times the number of planted edges, hence* $\beta\binom{k}{d}$*. The total noise is the standard deviation* $\sqrt{\binom{k}{d}}$ *times the scale of the number of solutions* $\sqrt{2\log\binom{p}{k}} \approx \sqrt{2k\log p}$*. The latter rescale has the following intuitive justification. If we assume that* $\binom{p-k}{k} \approx \binom{p}{k}$ unbiased *solutions are* independent*, then their* maximum *is located at* $\sqrt{\binom{k}{d}}\sqrt{2\log p}$*. The total signal has then to exceed this quantity, in order for the recovery to be possible. This argument ignores the dependencies between solutions, but it provides the right scaling of the SNR.*

Note that all these parameters may depend on p, that is, we consider $k = k_p$ and $d = d_p$. Throughout the paper, we hide the dependency on p for readability. In most of the analysis, the following rescaling of the involved quantities will appear naturally

$$\alpha_q := \lim_{p\to+\infty} \frac{\log q}{\log p} \tag{9}$$

which intuitively means that $q \approx p^{\alpha_q}$. We call α_q the *rate* of a generic $q = q_p$.

2 INFORMATION THEORETIC BOUNDS

By the MLE estimator's characterisation given in Equation (7), the recovery regime is regulated by the weight $S(\mathbf{x})$ of the planted solution and how it compares to the best among all other solutions' weights. For the analysis, it is useful to partition the latter according to their overlap with the planted solution.

Lemma 1. *For any* $m \in \{0, \ldots, k\}$*, let*

$$\mathcal{S}_m = \{\hat{\mathbf{x}} \in \mathcal{C}_{p,k} \colon \langle \mathbf{x}, \hat{\mathbf{x}} \rangle = m\}$$

denote the set of all solutions that share exactly m *nodes with the planted solution* $\mathbf{x}$*. Then the following bound holds for all* $k' \in \{0, \ldots, k\}$*:*

$$1 - P_r^{(k')} \leq \sum_{m=0}^{k'-1} \mathbb{P}\left(S(\mathbf{x}) < \max_{\hat{\mathbf{x}}\in\mathcal{S}_m} S(\hat{\mathbf{x}})\right). \tag{10}$$

For all $m < k$ *it further holds:*

$$1 - P_r \geq \mathbb{P}\left(S(\mathbf{x}) < \max_{\hat{\mathbf{x}}\in\mathcal{S}_m} S(\hat{\mathbf{x}})\right). \tag{11}$$

The proof of this lemma is given by a union bound and is given in the Supplementary Material. Given Definition 2 and the latter inequalities, we can reduce the analysis of the recovery regime to the determination of the values of γ for which $\mathbb{P}\left(S(\mathbf{x}) < \max_{\hat{\mathbf{x}} \in \mathcal{S}_m} S(\hat{\mathbf{x}})\right)$ vanishes or has limit 1, for $p \to +\infty$. In the first scenario (recovery regime), these probabilities need to vanish *sufficiently fast* to apply the union bound in Equation (10) over all different m.

2.1 Upper bound (partial or exact recovery)

We here provide upper bounds on the failure probability of partial and exact recover. The proof is given in appendix.

Theorem 1. *For any k and any $k' \in \{1, \ldots, k\}$, and for any $\gamma > \gamma_{\mathrm{UB}}^{(k')}$, the MLE estimator achieves k'-partial recovery according to Definition 1. The critical gamma is defined as*

$$\gamma_{\mathrm{UB}}^{(k')} \coloneqq \sqrt{1 + \alpha_k - 2\alpha_{k-k'} + \alpha_{k'}} + \sqrt{\alpha_k - \alpha_{k-k'} + \alpha_{k'}}$$

where α_k, $\alpha_{k'}$ and $\alpha_{k-k'}$ are defined according to Equation (9).

It follows easily from the latter theorem, the following on exact recovery.

Lemma 2. *Exact recovery is achieved for $\gamma > \gamma_{\mathrm{UB}}$, with*

$$\gamma_{\mathrm{UB}} = \sqrt{1 + 2\alpha_k} + \sqrt{2\alpha_k}.$$

Recovery of a constant fraction of k nodes is achieved with $k' = \lambda k$, with λ constant, for $\gamma > \gamma_{\mathrm{UB}}^{(\lambda k)}$ with

$$\gamma_{\mathrm{UB}}^{(\lambda k)} = 1 + \sqrt{\alpha_k}.$$

2.2 Lower bounds (impossibility of recovery)

We here provide a characterisation of the regime where recovery is impossible, given by the following two theorems, valid in different regimes.

Theorem 2. *For $\alpha_k \in (0, 1)$, and for $d \in \omega(1)$ the following holds:*

$$\lim_{p \to +\infty} P_r = 0$$

for any $\gamma < \gamma_{\mathrm{LB,G}}$, with $\gamma_{\mathrm{LB,G}}$ given by:

$$\gamma_{\mathrm{LB,G}} \coloneqq \begin{cases} \sqrt{1 - \alpha_k} & \text{for } d \in o(\sqrt{k}) \\ \sqrt{1 - \alpha_k}/\sqrt{e} & \text{otherwise} \end{cases} . \quad (12)$$

Theorem 3. *For $\alpha_k \in (0, 1)$ and any d, the following holds:*

$$\limsup_{p \to +\infty} P_r < 1$$

for any $\gamma < \gamma_{\mathrm{LB,F}}$, with $\gamma_{\mathrm{LB,F}}$ given by:

$$\gamma_{\mathrm{LB,F}} \coloneqq \sqrt{\frac{1 - \alpha_k}{2}} . \quad (13)$$

The first theorem gives tighter bounds, but its validity is confined to the case where the order of the hypergraph (or tensor) d grows to infinity with p. The second theorem is valid in any regime. However, it provides only an impossibility result as $\limsup P_r < 1$, and achieves a lower threshold. The proofs in the two regimes use two main arguments that are, respectively:

1. A recent tail bound on the maximum of *dependent* Gaussians with bounded correlation by Lopes (2018) (see Supplementary Material).
2. The generalised Fano's inequality (Han and Verdú, 1994).

Both proofs consider a coverage set of *weakly overlapping* solutions defined below. Intuitively, recovery in the original problem is at least as difficult as the recovery restricted to this set of weakly dependent solutions if the coverage is *sufficiently large*.

We now sketch the proof of the theorem starting with the following definition.

Definition 3. *For any $r \in \{0, \ldots, k\}$ define the coverage with overlap r a subset $\mathcal{C}(r) \subset \mathcal{C}_{p,k}$ of solutions satisfying the following conditions: (i) Any two solutions in $\mathcal{C}(r)$ share less than r nodes. (ii) For any solution $\mathbf{x}' \in \mathcal{C}_{p,k}, \mathbf{x}' \notin \mathcal{C}(r)$ there exists a solution $\mathbf{x}'' \in \mathcal{C}(r)$ such that $\mathbf{x}'$ and $\mathbf{x}''$ have at least r nodes in common. We denote by $C(r) \coloneqq |\mathcal{C}(r)|$ the cardinality of the coverage.*

In the proof of Theorem 2, we use the above mentioned result (Lopes, 2018) to show that the maximum among solutions in $\mathcal{C}(r)$ concentrates tightly around $\sigma_k\sqrt{\alpha_{C(r)} \cdot 2\log p}$ where $\sigma_k = \sqrt{\binom{k}{d}}$, while the planted solution concentrates around its expectation $\beta_k = \beta\binom{k}{d}$. Then, the condition for recovery translates into $\beta_k > \sigma_k\sqrt{\alpha_{C(r)} \cdot 2\log p}$ which corresponds to $\gamma < \sqrt{\frac{\alpha_{C(r)}}{k}}$. A crucial point is that, in order to prove the concentration of the maximum over $\mathcal{C}(r)$, the overlap r has to be small enough such that the maximum correlation between solutions weights $S(\hat{\mathbf{x}})$ vanishes, which gives the conditions on d in Theorem 2 and the corresponding bound on $\alpha_{C(r)}$.

The proof of Theorem 3 is based on Fano's inequality. By restricting to the solutions in $\mathcal{C}(r)$, we use the generalized Fano's inequality (Han and Verdú, 1994) to show that P_r satisfies the bound $P_r \lesssim 1 - \mathrm{I}(\mathbf{x}; \mathbf{Y}) \cdot \log^{-1} C(r)$ where $\mathrm{I}(\mathbf{x}; \mathbf{Y})$ is the mutual information between the planted solution $\mathbf{x}$ and the observations $\mathbf{Y}(\mathbf{x})$. By using (Han and Verdú, 1994) and the combinatorial structure of $\mathcal{C}_{p,k}$, we can upper bound the mutual information as $\binom{k}{d}\beta^2$. The rescaling of γ and a lower bound on the cardinality of the coverage set $C(r)$ gives the claim. The proof of both theorems is given in Supplementary Material.

3 COMPUTATIONAL THRESHOLDS VIA APPROXIMATE MESSAGE PASSING

Approximate message passing algorithms are a class of algorithms for high dimensional statistical estimation that approximate belief propagation in the large system limit (Donoho et al., 2009). Intuitively, it iteratively estimates the mean $\boldsymbol{x}^{(t)}$ and the variance $\boldsymbol{a}^{(t)}$ of the classic belief propagation messages in the factor graph of the estimation problem, discarding low order terms that depend on the target factor node of the messages. The algorithm results in a slight modification of a general spectral algorithm for tensor-PCA. In the following, we use the operator '$\circ$' to denote an operation that is performed elementwise on a vector or tensor. Let introduce the following probability distribution

$$g(\boldsymbol{y}|\boldsymbol{a},\boldsymbol{x}) = \frac{\mathbb{P}_p(\boldsymbol{y})\exp(\boldsymbol{x}\cdot\boldsymbol{y} - (\boldsymbol{y}\circ\boldsymbol{y})\cdot\boldsymbol{a})/2}{Z(\boldsymbol{a},\boldsymbol{x})}$$

where $Z(\boldsymbol{a},\boldsymbol{x})$ is the normalizing constant. Then, we define the following threshold function which will be used in the definition of AMP algorithm below:

$$f(\boldsymbol{a},\boldsymbol{x}) = \mathbb{E}_{g(\cdot|\boldsymbol{a},\boldsymbol{x})}[y]. \tag{14}$$

3.1 The AMP procedure

The iterative AMP procedure reads (derivation given in Supplementary Material):

$$\begin{aligned}
&\boldsymbol{b}^{(t)} = \beta^2(d-1)\langle \boldsymbol{Y}^{\circ 2}, \boldsymbol{\sigma}^{(t)} \otimes (\hat{\boldsymbol{x}}^{(t)} \circ \hat{\boldsymbol{x}}^{(t-1)})^{\otimes d-2}\rangle \\
&\boldsymbol{x}^{(t)} = \beta \mathbf{Y}\{\hat{\boldsymbol{x}}^{(t)}\} - \langle \boldsymbol{b}^{(t)}, \hat{\boldsymbol{x}}^{(t-1)}\rangle \\
&\boldsymbol{a}^{(t)} = \beta^2 \boldsymbol{Y}^{\circ 2}\{(\hat{\boldsymbol{x}}^{(t)})^{\circ 2}\} \\
&\boldsymbol{\sigma}^{(t+1)} = \boldsymbol{diag}(\boldsymbol{J}_x f(\boldsymbol{a}^{(t)}, \hat{\boldsymbol{x}}^{(t)})) \\
&\hat{\boldsymbol{x}}^{(t+1)} = f(\boldsymbol{a}^{(t)}, \boldsymbol{x}^{(t)})
\end{aligned} \tag{15}$$

where $\boldsymbol{J}_x f$ is the Jacobian w.r.t variable x of the function f and $\boldsymbol{diag}(\cdot)$ extracts its diagonal entries. The threshold function f becomes tractable when the prior distribution $\mathbb{P}_p$ *factorizes* as $\mathbb{P}_p(\boldsymbol{x}) = \prod_i p(x_i)$. In this case, the threshold function f factorizes as well into independent components, and the AMP equations proposed here are equivalent to those in Lesieur et al. (2017b) (details in Supplementary Material). In our problem, however, the prior distribution does *not* factorize since $\mathbb{P}_p$ is the uniform distribution over $\mathcal{C}_{p,k}$. A similar issue arises in the planted clique problem studied in (Deshpande and Montanari, 2015), where the authors propose to approximate the prior distribution with a factorized Bernoulli distribution with parameter $\delta = k/p$. With this approximation, the threshold function reads $f(\boldsymbol{a},\boldsymbol{x}) = (f_i(a_i,x_i))_{i=1}^p$, where

$$f_i(a_i,x_i) = \frac{1}{1+\exp(-x_i + a_i/2 + \log(1/\delta - 1))}\ .$$

While this Bernulli approximation is effective for large k – like the regime $k = \Theta(\sqrt{p})$ studied in (Deshpande and Montanari, 2015) – it may be inaccurate for small k. We indeed observe experimentally (see Section 4) that for small values of k, the above approximation is no longer effective. Hence, we propose a finer approximation of the threshold function that is still tractable. Our parametrization is inspired by the independent Bernoulli approximation:

$$f(\boldsymbol{a},\boldsymbol{x}) = \left(\frac{1}{1+\exp(-x_i + a_i/2 + \lambda(\boldsymbol{x}))}\right)_{i=1}^p \tag{16}$$

with $\lambda(\boldsymbol{x})$ being a scalar used to enforce f to select k components equals to 1, hence given by $\sum_i f_i(\boldsymbol{a},\boldsymbol{x}) = k$. The experiments described in Section 4 below show that our finer approximation outperforms the Bernulli i.i.d. approximation.

3.2 State Evolution

The evolution of the approximated message passing algorithm in the special case of Bayesian-optimal inference, can be tracked by a one dimensional iterative equation of the overlap order parameter. In our setting, with a generic (non-factorizable) prior distribution $\mathbb{P}_p$ and threshold function f we can define the multidimensional overlap order parameter as

$$\boldsymbol{m}^{(t)} = \frac{1}{\binom{p-1}{d-1}}\mathbb{E}_{\boldsymbol{x}}[\mathbf{1}\{\boldsymbol{x}\circ\boldsymbol{x}^{(t)}\}]$$

where $\mathbf{1}$ is the tensor with components $\mathbf{1}_{i,i_2,\ldots,i_d} = 1$ if $i_2 < i_3 < \cdots < i_d$ and 0 otherwise. The multidimensional state evolution (SE) (generalizing (Lesieur et al., 2017b)) reads then (see Supplementary Material for a heuristic derivation):

$$\boldsymbol{m}^{(t+1)} = \frac{\mathbb{E}_{\boldsymbol{x},\boldsymbol{z}}\left[\mathbf{1}\{\boldsymbol{x}\circ f\left(\hat{\boldsymbol{m}}^{(t)}, \hat{\boldsymbol{m}}^{(t)}\circ\boldsymbol{x} + (\hat{\boldsymbol{m}}^{(\boldsymbol{t})})^{\circ 1/2}\circ\boldsymbol{z}\right)\}\right]}{\binom{p-1}{d-1}}\ ,$$

where $\hat{\boldsymbol{m}}^{(t)} = \beta^2\binom{p-1}{d-1}\boldsymbol{m}^{(t)}$ and $\boldsymbol{z}$ is a p-dimensional vector with i.i.d. standard Gaussian entries. Note again that assuming a factorized i.i.d. prior distribution $\mathbb{P}_p(\boldsymbol{x}) = \prod_i p(x_i)$ the SE is equivalent to the single

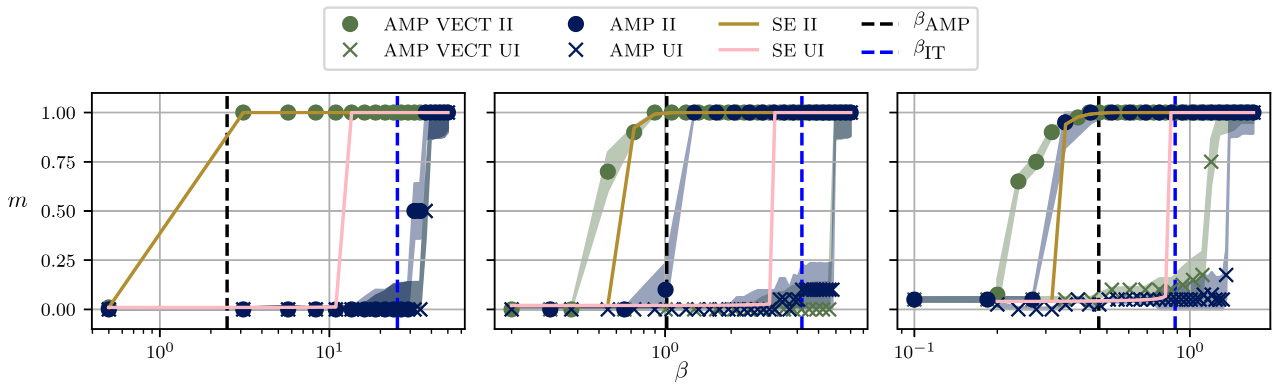

Figure 1: Empirical performance of the AMP algorithm for fixed $p = 500$, $d = 3$ and different values of k (respectively, from left to right, $k = 20, 10, 5$). The experiment is repeated 20 times, and we report the median overlap achieved m. The AMP and AMP VECT refer respectively to using a scalar and a vectorial thresholding function as in Equation (16).

letter evolution of the *scalar* overlap $m_t = \frac{1}{p}\langle \boldsymbol{x}, \hat{\boldsymbol{x}}^{(t)}\rangle$ described in (Lesieur et al., 2017b),

$$m_{t+1} = \mathbb{E}_{x,z}\left[x \cdot f(\hat{m}_t, \hat{m}_t x + \sqrt{\hat{m}_t} z)\right], \qquad (17)$$

where $x \sim p$, $z \sim \mathcal{N}(0,1)$, and $\hat{m}_t = \beta^2\binom{p-1}{d-1}m_t$. In the following we derive an analytical threshold for the AMP algorithm to succeed, approximating the true prior distribution with a factorized Bernoulli prior with parameter $\delta = \frac{k}{p}$ and using the simplified SE in Equation (17). A simple heuristic argument based on the study of the fixed points of the SE is given in the Supplementary Material.

Claim 1. *The recovery threshold for the AMP algorithm reads:*

$$\gamma_{\text{AMP}} \coloneqq \sqrt{\frac{1}{2e}\left(\frac{p}{k}\right)^{d-1}\frac{1}{d(d-1)\log p}}. \qquad (18)$$

4 EXPERIMENTS AND DISCUSSION

In Figure 1, we report both the empirical performance of the AMP algorithm and the factorized SE fixed point (according to Equation (17)) for uninformative initialization (with a random overlap with the planted solution) and informative initialization (respectively UI and II) for different values of k. For a high value of k, the Bernoulli i.i.d. approximation of the prior function and factorized thresholding function well matches the SE fixed point, with the UI empirical performance slightly worse than the SE prediction due to finite-size effects (as already observed for the planted clique problem in graphs (Deshpande and Montanari, 2015)). At smaller values of k, the statistical dependency between signal components increases and the AMP empirical performance badly mismatch the SE prediction. In this setting, the proposed multivariate threshold function significantly outperforms the naive factorized AMP, matching the SE in the II setting correctly. Interestingly, the dynamic phase transition of the AMP indicated by the critical signal at which the SE with II fails approaches the conjectured[4] information-theoretic threshold γ_{LB}. This finding suggests a possibility of analysis of the IT thresholds with statistical physics-inspired techniques (e.g., interpolation methods by Barbier et al. (2020)) that so far have never been applied rigorously to the case of structured priors in the sparse setting.

In Figure 2 we also report the factorized SE fixed point for $d = 3$ and $d = 4$ for both UI and II. We can observe the good agreement between the analytical computational threshold given by Claim 1 and the SE's empirical fixed point. The dynamical phase transition indicated by the m^* transition for II (bottom row) is close to the information-theoretic lower bound (as shown already for the tensor-PCA problem in Lesieur et al. (2017b)).

Acknowledgements

Szpankowski was in part supported by NSF Center on Science of Information Grants CCF-0939370 and NSF Grants CCF-1524312, CCF- 2006440, CCF-2007238, and Google Research Award.

[4]Clearly, γ_{LB} may no longer be a lower bound for the problem, since now we initialize the algorithm with the correct planted solution. Hence the term "conjectured".

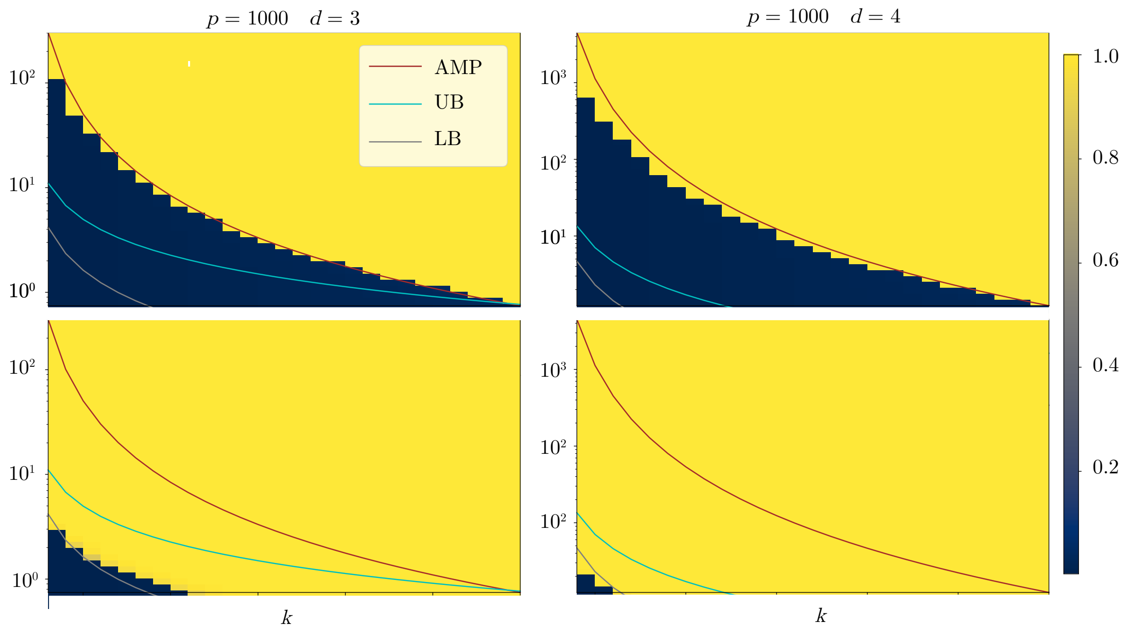

Figure 2: State evolution fixed point overlap m^* reached by the factorized equation in Equation (17). The first row reports the overlap obtain with uninformative initialization (UI, $m_0 \approx 0$), while the bottom row reports the overlap with informative initialization (II, $m_0 = 1$). The empirical fixed point with UI well matches the analytical AMP threshold given by Claim 1 (brown line, expressed in terms of the β parameter), while the II fixed point approaches the conjectured IT threshold located at the lower bound.

References

Emmanuel Abbe and Colin Sandon. Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics*, 71(7):1334–1406, 2018.

Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory (ISIT)*, 62(1):471–487, 2016.

Christopher Aicher, Abigail Z Jacobs, and Aaron Clauset. Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3(2):221–248, 2014.

Gerard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Algorithmic thresholds for tensor PCA. *The Annals of Probability*, 48(4):2052–2087, 2020.

Boaz Barak, Samuel B Hopkins, Jonathan Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *57th Symposium on Foundations of Computer Science (FOCS)*, pages 428–437. IEEE, 2016.

Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems (NIPS)*, pages 424–432, 2016.

Jean Barbier, Nicolas Macris, and Cynthia Rush. All-or-nothing statistical and computational phase transitions in sparse spiked matrix estimation. *Advances in Neural Information Processing Systems (NeurIPS)*, 33, 2020.

Giulio Biroli, Chiara Cammarota, and Federico Ricci-Tersenghi. How to iron out rough landscapes and get optimal performances: averaged gradient descent and its application to tensor PCA. *Journal of Physics A: Mathematical and Theoretical*, 53(17): 174003, 2020.

Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory (COLT)*, pages 648–847. PMLR, 2020.

Yudong Chen and Jiaming Xu. Statistical-computational phase transitions in planted models: The high-dimensional setting. In *31st International Conference on Machine Learning (ICML)*, PMLR, pages 244–252, 2014.

Eden Chlamtac, Michael Dinitz, Christian Konrad, Guy Kortsarz, and George Rabanca. The Densest k-Subhypergraph Problem. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, volume 60 of *LIPIcs*, pages 6:1–6:19, 2016.

Davin Choo and Tommaso d'Orsi. The Complexity of Sparse Tensor PCA. *Advances in Neural Information Processing Systems (NeurIPS)*, 34, 2021.

Luca Corinzia, Paolo Penna, Luca Mondada, and Joachim M. Buhmann. Exact recovery for a family of community-detection generative models. In *IEEE International Symposium on Information Theory (ISIT)*, pages 415–419. IEEE, 2019.

Luca Corinzia, Paolo Penna, Wojciech Szpankowski, and Joachim M Buhmann. On maximum-likelihood estimation in the all-or-nothing regime. In *IEEE International Symposium on Information Theory (ISIT)*, pages 1106–1111. IEEE, 2021.

Thomas M Cover. *Elements of information theory.* John Wiley & Sons, 1999.

Bernard Derrida. Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, 24(5):2613, 1981.

Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse pca. In *IEEE International Symposium on Information Theory (ISIT)*, pages 2197–2201. IEEE, 2014.

Yash Deshpande and Andrea Montanari. Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Foundations of Computational Mathematics*, 15(4):1069–1128, 2015.

David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

Willliam Feller. *An introduction to probability theory and its applications*, volume 2. John Wiley & Sons, 2008.

Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral partitioning of uniform hypergraphs under planted partition model. In *Advances in Neural Information Processing Systems (NIPS)*, pages 397–405, 2014.

Shi Gu, Muzhi Yang, John D Medaglia, Ruben C Gur, Raquel E Gur, Theodore D Satterthwaite, and Danielle S Bassett. Functional hypergraph uncovers novel covariant structures over neurodevelopment. *Human brain mapping*, 38(8):3823–3835, 2017.

Rungang Han, Rebecca Willett, and Anru R. Zhang. An optimal statistical and computational framework for generalized tensor estimation. *arXiv preprint arXiv: 2002.11255*, 2021.

Te Sun Han and Sergio Verdú. Generalizing the fano inequality. *IEEE Transactions on Information Theory (ISIT)*, 40(4):1247–1251, 1994.

Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory (COLT)*, pages 956–1006, 2015.

Yukito Iba. The Nishimori line and Bayesian statistics. *Journal of Physics A: Mathematical and General*, 32 (21):3875, 1999.

Aukosh Jagannath, Patrick Lopatto, Leo Miolane, et al. Statistical thresholds for tensor PCA. *Annals of Applied Probability*, 30(4):1910–1933, 2020.

Jean-Michel Jolion and Walter Kropatsch. *Graph based representations in pattern recognition*, volume 12. Springer Science & Business Media, 2012.

Chiheon Kim, Afonso S Bandeira, and Michel X Goemans. Stochastic block model for hypergraphs: Statistical limits and a semidefinite programming approach. *arXiv preprint arXiv:1807.02884*, 2018.

Robert Legenstein, Wolfgang Maass, Christos H. Papadimitriou, and Santosh S. Vempala. Long Term Memory and the Densest K-Subgraph Problem. In *9th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 94 of *LIPIcs*, pages 57:1–57:15, 2018. ISBN 978-3-95977-060-6.

Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017a.

Thibault Lesieur, Léo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *IEEE International Symposium on Information Theory (ISIT)*, pages 511–515. IEEE, 2017b.

Miles E Lopes. On the maximum of dependent gaussian random variables: A sharp bound for the lower tail. *arXiv preprint arXiv:1809.08539*, 2018.

Yuetian Luo and Anru R. Zhang. Tensor clustering with planted structures: Statistical optimality and computational limits. *arXiv preprint arXiv: 2005.10743*, 2021.

Andrea Montanari. Statistical estimation: from denoising to sparse regression and hidden cliques. *Statistical Physics, Optimization, Inference, and Message-Passing Algorithms: Lecture Notes of the Les Houches School of Physics: Special Issue*, 2013.

Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. On the limitation of spectral methods: From the gaussian hidden clique problem to rank one perturbations of gaussian tensors. *IEEE Transactions on Information Theory (ISIT)*, 63(3):1572–1579, 2016.

Elchanan Mossel, Joe Neeman, and Allan Sly. Consistency thresholds for the planted bisection model. In *47th ACM Symposium on Theory of Computing (STOC)*, pages 69–75. ACM, 2015.

Jonathan Niles-Weed and Ilias Zadik. The all-or-nothing phenomenon in sparse tensor PCA. *Advances in Neural Information Processing Systems (NeurIPS)*, 33, 2020.

Tiago P Peixoto. Nonparametric weighted stochastic block models. *Physical Review E*, 97(1):012306, 2018.

Amelia Perry, Alexander S. Wein, and Afonso S. Bandeira. Statistical limits of spiked tensor models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(1):230–264, 02 2020.

Emile Richard and Andrea Montanari. A statistical model for tensor pca. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2897–2905, 2014.

Zhijiang Wang, Jiming Liu, Ning Zhong, Yulin Qin, Haiyan Zhou, Jian Yang, and Kuncheng Li. A naive hypergraph model of brain networks. In *International Conference on Brain Informatics*, pages 119–129. Springer, 2012.

Alexander S. Wein, Ahmed El Alaoui, and Cristopher Moore. The Kikuchi Hierarchy and Tensor PCA. In *IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1446–1468, 2019.

Chen Zu, Yue Gao, Brent Munsell, Minjeong Kim, Ziwen Peng, Yingying Zhu, Wei Gao, Daoqiang Zhang, Dinggang Shen, and Guorong Wu. Identifying high order brain connectome biomarkers via learning on hypergraph. In *International Workshop on Machine Learning in Medical Imaging*, pages 1–9. Springer, 2016.

Supplementary Material:
Statistical and computational thresholds for the planted k-densest sub-hypergraph problem

A Postponed proofs

Proof of Lemma 1. By definition of $P_r^{(k')}$, $1 - P_r^{(k')} = \mathbb{P}\left(\langle \mathbf{x}_{\text{MLE}}, \mathbf{x}\rangle < k'\right)$. By the characterization of the MLE in Equation (7) and the definition of $\mathcal{S}_m$, the latter event is equal to

$$\{S(\mathbf{x}) > S(\hat{\mathbf{x}}),\ \forall\ \hat{\mathbf{x}} \in \mathcal{S}_0 \cup \cdots \cup \mathcal{S}_{k'-1}\}.$$

The claim in Equation (10) follows then from the union-bound on the probability

$$\mathbb{P}\left(\bigcup_{m=0}^{k'-1}\left\{S(\mathbf{x}) \le \max_{\hat{\mathbf{x}}\in\mathcal{S}_m}(S(\hat{\mathbf{x}}))\right\}\right).$$

To prove the lower bound in Equation (11), we can observe that if $S(\mathbf{x}) < \max_{\hat{\mathbf{x}}\in\mathcal{S}_m}(S(\hat{\mathbf{x}}))$ for some $m < k$, then the $\mathbf{x}_{\text{MLE}} \neq \mathbf{x}$, and thus it fails to exactly recover the planted solution. □

A.1 Proofs for the Lower Bound

We shall use the following basic fact. For any random variables A and B and for any $t \in \mathbb{R}$ the following inequality holds:

$$\mathbb{P}(A \ge B) \le \mathbb{P}(A > t) + \mathbb{P}(t \ge B) \tag{19}$$

Proof of Theorem 1. For the analysis, we define the following quantities depending on $k' \in \{0, \dots, k-1\}$ (we consider p, k, d to be the parameters of the problem, hence their dependency is not highlighted):

$$\begin{aligned} Q(k') &:= \binom{p-k}{k-k'} \\ M(k') &:= \binom{k}{k'}\binom{p-k}{k-k'} = \binom{k}{k-k'}\binom{p-k}{k-k'} \\ D(k') &:= \binom{k}{d} - \binom{k'}{d}. \end{aligned}$$

For each fixed subset of k' nodes of the planted solution, there are $Q(k')$ solutions that share exactly these k' nodes with the planted solution. Moreover, there are exactly $M(k')$ solutions that share any k' nodes with the planted solution. Each solution sharing k' nodes with the planted solution differs in $D(k')$ edges with the latter. We use the union bound given in Equation (10), and control the quantities

$$\mathbb{P}\left(\max_{\hat{\mathbf{x}}\in\mathcal{S}_m} S(\hat{\mathbf{x}}) > S(\mathbf{x})\right)$$

using the tail bounds of the Gaussian from Lemma 9 and spitting the inequality into the sum of two independent terms as in Equation (19). We hence get the following lemma (full proof given below in this section):

Lemma 3. *For every $\epsilon > 0$ and for every $k' < k$ and $k > 1$, let $\gamma_{UB_\epsilon}^{(k')} := \sqrt{\frac{\binom{k}{d}}{kD(k')}} \cdot UB_\epsilon(k')$, where*

$$UB_\epsilon(k') := \sqrt{\frac{\log M(k')}{\log p} + \epsilon} + \sqrt{\frac{\log \binom{k}{k'}}{\log p} + \epsilon}\ . \tag{20}$$

Then, for any $\gamma > \gamma_{UB_\epsilon}^{(k')}$ it holds that

$$\mathbb{P}\left(\max_{\hat{\mathbf{x}}\in\mathcal{S}_{k'}} S(\hat{\mathbf{x}}) > S(\mathbf{x})\right) \le \frac{1}{\sqrt{\pi}p^\epsilon}.$$

Plugging the result of the latter Lemma into Equation (10) we get:

$$1 - P_r^{(k')} \in \mathcal{O}\left(\frac{k'}{p^{\epsilon}}\right) .$$

Using now Lemma 15 to characterize further the bound $\gamma_{UB_\epsilon}^{(k')}$, and reparametrazing $\epsilon = \alpha_{k'} + \tilde{\epsilon}$, with $\tilde{\epsilon} > 0$ an arbitrarily constant, we have

$$\frac{k'}{p^{\epsilon}} = \frac{k'}{p^{\alpha_{k'}}} \cdot \frac{1}{p^{\tilde{\epsilon}}} = \exp\left(\log p \left[-\tilde{\epsilon} + \frac{\log k'}{\log p} - \alpha_{k'}\right]\right) \in o(1).$$

The asymptotics is due to the fact that $\frac{\log k'}{\log p} - \alpha_{k'} \to 0$, by definition of $\alpha_{k'}$ in Equation (9), and hence the expression $-\tilde{\epsilon} + \frac{\log k'}{\log p} - \alpha_{k'}$ is negative for sufficiently large p. Hence, the probability $1 - P_r^{(k')}$ tends to 0. □

Proof of Lemma 3. For an arbitrary subset F of k' nodes of the planted solution $\mathbf{x}$, denoted by $\mathbf{x}_F$, with $k' \in \{0, \ldots, k-1\}$, let $\mathcal{S}_{k'}^F$ be the set of all solutions that share exactly the set F with the planted solution. Let $S^{-F}(\hat{\mathbf{x}})$ denote the sum of the weights in $\hat{\mathbf{x}}$ but not in $\mathbf{x}_F$, as

$$S^{-F}(\hat{\mathbf{x}}) = S(\hat{\mathbf{x}}) - S(\hat{\mathbf{x}} \circ \mathbf{x}_F).$$

By the union bound over the $\binom{k}{k'}$ possible fixed subsets F of k' nodes, we get

$$\begin{aligned}
&\mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{S}_{k'}} S(\hat{\mathbf{x}}) > S(\mathbf{x})\right) \leq \\
&\qquad \leq \binom{k}{k'} \mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{S}_{k'}^F} S(\hat{\mathbf{x}}) > S(\mathbf{x})\right) \\
&\qquad = \binom{k}{k'} \mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{S}_{k'}^F} S^{-F}(\hat{\mathbf{x}}) > S^{-F}(\mathbf{x})\right)
\end{aligned} \tag{21}$$

where the equality follows from subtracting from both sides the quantity $S(\mathbf{x}_F)$. Using Equation (19), we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{S}_{k'}^F} S^{-F}(\hat{\mathbf{x}}) > S^{-F}(\mathbf{x})\right) \leq \\
&\qquad \leq \mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{S}_{k'}^F} S^{-F}(\hat{\mathbf{x}}) > t\right) + \mathbb{P}\left(t \geq S^{-F}(\mathbf{x})\right) \\
&\qquad \leq \mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{S}_{k'}^F} S^{-F}(\hat{\mathbf{x}}) > t_{\Delta'}\right) + \\
&\qquad\qquad + \mathbb{P}\left(D(k')\beta - t_{\Delta''} > S^{-F}(\mathbf{x})\right) \\
&\qquad \leq Q(k') p(k', \Delta') + p(k', \Delta'')
\end{aligned} \tag{22}$$

where the second inequality follows from Lemma 14 (together with the definition of $t_{\Delta'}$ and $t_{\Delta''}$) and the latter inequality follows from Equation (34) and Equation (33) in Lemma 13. Combining Equation (21) and Equation (22), we get

$$\begin{aligned}
&\mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{S}_{k'}} S(\hat{\mathbf{x}}) > S(\mathbf{x})\right) \leq \binom{k}{k'} \left(Q(k') p(k', \Delta') + p(k', \Delta'')\right) \\
&= \binom{k}{k'} \frac{1}{\sqrt{4\pi \log p}} \left(\left(\frac{1}{p}\right)^{\frac{\Delta'}{D(k')}} \frac{Q(k')}{\sqrt{\Delta'}} + \left(\frac{1}{p}\right)^{\frac{\Delta''}{D(k')}} \frac{1}{\sqrt{\Delta''}}\right) \\
&= \frac{1}{\sqrt{4\pi \log p}} \Bigg(\left(\frac{1}{p}\right)^{\frac{\Delta'}{D(k')} - \frac{\log M(k')}{\log p}} \frac{1}{\sqrt{\Delta'}} + \\
&\qquad + \left(\frac{1}{p}\right)^{\frac{\Delta''}{D(k')} - \frac{\log \binom{k}{k'}}{\log p}} \frac{1}{\sqrt{\Delta''}} \Bigg)
\end{aligned}$$

where in the first equality we used Equation (32) and in the last equality the identity $z = p^{\frac{\log z}{\log p}}$ and the fact that

$$\binom{k}{k'} Q(k') = \binom{k}{k'}\binom{p-k}{k-k'} = M(k').$$

Then, by Lemma 14, we get

$$\begin{aligned}\mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{S}_{k'}} S(\hat{\mathbf{x}}) > S(\mathbf{x})\right) &\leq \frac{1}{p^\epsilon}\frac{1}{\sqrt{4\pi \log p}}\left(\frac{1}{\sqrt{\Delta'}} + \frac{1}{\sqrt{\Delta''}}\right) \\ &\leq \frac{1}{p^\epsilon}\frac{1}{\sqrt{4\pi\log p}}\frac{1}{D(k')}\left(\sqrt{\frac{\log p}{\log M(k')}} + \right. \\ &\qquad \left. + \sqrt{\frac{\log p}{\log\binom{k}{k'}}}\right) \\ &\leq \frac{1}{\sqrt{\pi} p^\epsilon}\end{aligned}$$

where in the last inequality we used $D(k') \geq 1$, $M(k') \geq \binom{k}{k'} \geq k$, and $\log k > 1$, that valid for $k > k'$ and $k > 1$. □

A.2 Proofs of the Lower Bound

Proof of Theorem 2. The proof of this lower bound is based on the result on tail bounds on the *maximum* of Gaussians with *bounded* correlation (Lopes, 2018) given for convenience in Lemma 10. We here outline a road map of the proof: (i) Given an overlap $r \leq k$ between two solutions $\mathbf{x}', \mathbf{x}''$, we define the correlation $\rho(r)$ as the correlation between the random variables $S(\mathbf{x}')$ and $S(\mathbf{x}'')$. We first analyze the conditions for vanishing correlation $\rho(r) \to 0$. (ii) Using the tail bound on correlated Gaussians in Lemma 10, we show that vanishing correlation implies the concentration of the maximum of a given coverage of solutions $\mathcal{C}(r)$. (iii) Using a lower bound on the rate of the cardinality of the coverage, we provide a respective lower bound of the recovery threshold. We hence first give a Lemma that characterizes the condition to have vanishing correlations.

Lemma 4. *For any two solutions $\mathbf{x}', \mathbf{x}'' \in \mathcal{C}_{p,k}$, that share r nodes as $\langle \mathbf{x}', \mathbf{x}'' \rangle = r$, the correlation of the weights $\rho \coloneqq \rho_{S(\mathbf{x}'), S(\mathbf{x}'')}$ reads*

$$\rho = \frac{\binom{r}{d}}{\binom{k}{d}}$$

and it vanishes in each of the following two regimes:

1. *For $d \in \omega(1)$ and $r \leq \lambda k$ for any constant λ satisfying*

$$\begin{cases} \lambda < 1 & \text{for } d \in o(\sqrt{k}) \\ \lambda < 1/e & \text{otherwise} \end{cases} \tag{23}$$

2. *For $d \in \mathcal{O}(1)$ and for any $r \in o(k)$.*

We can hence show in the following that, given a coverage with vanishing correlation $\rho(r)$, the maximum of such coverage in bounded from below in high probability.

Lemma 5. *Given a number of shared nodes r and a coverage $\mathcal{C}(r)$ that satisfies Definition 3, if the correlation between the weights of two solutions $\mathbf{x}', \mathbf{x}'' \in \mathcal{C}(r)$, $\rho(r) = \rho_{S(\mathbf{x}'), S(\mathbf{x}'')} \to 0$ vanishes, then for any constant $\epsilon \in (0, 1)$ the following upper bound on the maximum weight in the coverage holds:*

$$\lim_{p \to +\infty} \mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{C}(r)} S(\hat{\mathbf{x}}) \leq (1-\epsilon)\cdot \sigma_k \sqrt{\alpha_{C(r)} \cdot 2\log p}\right) = 0 \tag{24}$$

where $\sigma_k = \sqrt{\binom{k}{d}}$.

We further provide a bound on the opposite direction for the weight of the planted solution, such that the two quantities can be well separated in high probability.

Lemma 6. *For any sequence* $\Delta = \Omega\left(\frac{\binom{k}{d}}{\log p}\right)$*, it holds that*

$$\lim_{p\to+\infty} \mathbb{P}_z\left(S(\mathbf{x}) > \beta_k + \sqrt{\Delta\cdot 2\log p}\right) = 0$$

where $\beta_k = \binom{k}{d}\beta$ *and* $\mathbf{x}$ *is the planted solution.*

We can prove the main Lemma that connects the recovery threshold to the rate of any coverage $\alpha_{C(r)}$.

Lemma 7. *Given the assumptions of Lemma 5 and given that* $\alpha_{C(r)} \in \Omega(1)$*, for any* $\epsilon > 0$ *constant, if* $\gamma \leq (1-\epsilon)\sqrt{\frac{\alpha_{C(r)}}{k}}$*, then the probability of recovery vanishes:*

$$\lim_{p\to+\infty} P_r = 0\ .$$

Proof. From the hypothesis that $\alpha_{C(r)} \in \Omega(1)$, $\log p \to +\infty$, and given any constant $\delta_0 > 0$ the condition on γ above implies that for p large enough:

$$\gamma \leq (1-\epsilon)\sqrt{\frac{\alpha_{C(r)}}{k}} - \sqrt{\frac{\delta_0}{k\log p}}\ .$$

Using the definition of γ and defining $\Delta = \delta_0 \cdot \frac{\binom{k}{d}}{\log p}$ we get with simple manipulation:

$$(1-\epsilon)\sqrt{\binom{k}{d}}\sqrt{\alpha_{C(r)}\cdot 2\log p} > \binom{k}{d}\beta + \sqrt{\Delta\cdot 2\log p} \tag{25}$$

Using Equation (19) and Equation (11), we can write an upper-bound on the recovery probability as:

$$\begin{aligned} P_r &\leq \mathbb{P}\left(\max_{\hat{\mathbf{x}}\in\mathcal{C}(r)} S(\hat{\mathbf{x}}) \leq S(\mathbf{x})\right) \\ &\leq \mathbb{P}\left(\max_{\hat{\mathbf{x}}\in\mathcal{C}(r)} S(\hat{\mathbf{x}}) \leq t\right) + \mathbb{P}\left(t < S(\mathbf{x})\right). \end{aligned}$$

We can now use the condition in Equation (25), to get a t such that the conditions for both Lemma 5 and Lemma 6 are satisfied. We hence obtain

$$\mathbb{P}\left(\max_{\hat{\mathbf{x}}\in\mathcal{C}(r)} S(\hat{\mathbf{x}}) > t\right) \to 1 \qquad \text{and} \qquad \mathbb{P}\left(S(\mathbf{x}) \leq t\right) \to 1 \tag{26}$$

and so the claim follows. □

Consider any $r = \lambda k$ with λ being any constant satisfying Equation (23), according to the regime of d and k. By Lemma 4 and the assumption of the theorem $d \in \omega(1)$, the correlation $\rho(r)$ vanishes. We can hence bound the recovery threshold with a respective bound on the rate $\alpha_{C(r)}$, that is given in the following Lemma.

Lemma 8. *There exist a coverage* $\mathcal{C}(r)$ *according to Definition 3 with cardinality* $\mathcal{C}(r)$*,* $C(r) := |\mathcal{C}(r)|$ *at least*

$$C(r) \gtrsim \frac{\binom{p}{k}}{B(r)} \qquad \textit{for} \qquad B(r) = \sum_{l=r}^{k}\binom{k}{l}\binom{p-k}{k-l}. \tag{27}$$

For any $\alpha_k \in (0,1)$ *and* $r \leq \lambda k$*, with constant* $\lambda \in (0,1)$*, and for* $r \in \omega(\frac{k}{\log n})$ *it holds that* $\alpha_{C(r)} \gtrsim r(1-\alpha_k)$*.*

We can thus apply Lemma 7 and Lemma 8 with $r = \lambda k$ and obtain the desired result as follows:

$$(1-\epsilon)\sqrt{\frac{\alpha_{C(r)}}{k}} \geq (1-\epsilon)\sqrt{\lambda(1-\alpha_k)} > \gamma$$

where the last inequality follows from the condition on γ in Equation (12) and the hypothesis that λ satisfies Equation (23). The above inequality and Lemma 7 implies the claim. $\square$

Proof of Lemma 4. Both $S(\mathbf{x}')$ and $S(\mathbf{x}'')$ are Gaussian random variables $\mathcal{N}(\mu, \beta^2\binom{k}{d})$, where μ is the mean, and depends on the amount of nodes that the solution shares with the planted one $\mathbf{x}$. The correlation is hence

$$\rho = \frac{\mathbb{E}[(S(\mathbf{x}') - \mathbb{E}S(\mathbf{x}'))(S(\mathbf{x}'') - \mathbb{E}S(\mathbf{x}''))]}{\beta^2\binom{k}{d}}.$$

Observe that $S(\mathbf{x}') - \mathbb{E}S(\mathbf{x}')$ and $S(\mathbf{x}'') - \mathbb{E}S(\mathbf{x}'')$ are the sum of $\binom{k}{d} - \binom{r}{d}$ independent terms, and $\binom{r}{d}$ identical terms that are Gaussian distributed $\mathcal{N}(0, \beta^2)$. Hence, we get $\rho = \binom{r}{d}/\binom{k}{d}$. We can now upper bound the correlation as

$$\begin{aligned} k(k-1)\cdots(k-d+1) \geq (k-d)^d &= \left(1-\frac{d}{k}\right)^d k^d \\ &\geq \left(1-\frac{d^2}{k}\right) k^d \end{aligned}$$

and therefore

$$\begin{aligned} \frac{\binom{r}{d}}{\binom{k}{d}} = \frac{r!}{d!(r-d)!}\frac{d!(k-d)!}{r!} &= \frac{r(r-1)\cdots(r-d+1)}{k(k-1)\cdots(k-d+1)} \\ &\leq \left(\frac{r}{k}\right)^d \frac{1}{(1-\frac{d^2}{k})} \\ &\leq \frac{\lambda^d}{(1-\frac{d^2}{k})} . \end{aligned}$$

For $d \in o(\sqrt{k})$ and $d \in \omega(1)$ the latter quantity converges to 0 for any constant $\lambda < 1$. As for the other case in Equation (23), we have

$$\frac{\binom{r}{d}}{\binom{k}{d}} \leq \frac{e^d r^d}{d^d} \cdot \frac{d^d}{k^d} = \left(\frac{e \cdot r}{k}\right)^d \leq (e\lambda)^d \to 0$$

where the asymptotics follows from $e < c_0$ and $d \to \infty$. As for the second case, since $d \in \Theta(1)$ we have $\binom{r}{d} \in \Theta(r^d)$ and $\binom{k}{d} \in \Theta(k^d)$, and thus $r \in o(k)$ implies $\binom{r}{d} \in o(\binom{k}{d})$. $\square$

Proof of Lemma 5. We apply the tail bound to the maximum of correlated Gaussians given in (Lopes, 2018) (given also in Lemma 10 for convenience) with $N = C(r)$. In particular, since by assumption the correlation vanishes, we can choose *any* $\rho_0 < 1$ and, for sufficiently large p, satisfy $\delta_0\sqrt{1-\rho_0} \geq 1-\epsilon$ and $\rho(r) \leq \rho_0$. We can hence obtain:

$$\lim_{p\to+\infty} \mathbb{P}\left(\max_{\hat{\mathbf{x}}\in\mathcal{C}(r)} S(\hat{\mathbf{x}}) \leq (1-\epsilon)\cdot\sigma_k\sqrt{2\log C(r)}\right) = 0 . \tag{28}$$

Then, Equation (28) follows from Equation (30) as both η and ξ in Equation (31) are constants and $C(r) = |\mathcal{C}(r)| \to \infty$. Finally, recall that $\alpha_{C(r)} = \lim_p \log C(r)/\log p$, and hence, for p large enough, $\log C(r) \leq \alpha_{C(r)} \log p(1+\epsilon)$. The claim follows from the arbitrariness of ϵ. $\square$

Proof of Lemma 6. Simply observe that, fixing $\mathbf{x}$ as the planted solution, $S(\mathbf{x}) \sim \mathcal{N}(\binom{k}{d}\beta, \binom{k}{d})$. Hence, by applying Lemma 9 with $\mu = \binom{k}{d}\beta = \beta_k$, $\sigma^2 = \binom{k}{d}$, and $c = t_\Delta = \sqrt{\Delta 2 \log n}$ we have

$$\begin{aligned}\mathbb{P}_z\left(S(\mathbf{x}) > \beta_k + t_\Delta\right) &\leq \frac{1}{t_\Delta} \cdot \frac{e^{-t_\Delta^2/2\binom{k}{d}}}{\sqrt{2\pi}} \\ &= \frac{e^{-(\Delta \log p)/\binom{k}{d}}}{t_\Delta\sqrt{2\pi}} \\ &= \frac{p^{-\Delta/\binom{k}{d}}}{\sqrt{2\pi\Delta}\cdot\sqrt{2\log p}} ,\end{aligned}$$

and the latter quantity goes to 0 for $\Delta = \Omega\left(\frac{\binom{k}{d}}{\log p}\right)$. □

Proof of Lemma 8. We construct $\mathcal{C}(r)$ by a iterative greedy procedure. Starting from an arbitrary $\hat{\mathbf{x}} \in \mathcal{C}_{p,k}$, include $\hat{\mathbf{x}}$ into $\mathcal{C}(r)$ and remove all solutions in $\mathcal{C}_{p,k}$ with overlap with the planted solution $\langle \mathbf{x}, \hat{\mathbf{x}} \rangle \geq r$. Iterate this step with the remaining solutions in $\mathcal{C}_{p,k}$ not considered before, until there are none with this property. At every step we include one new solution in $\mathcal{C}(r)$ we remove at most $B(r)$ solutions from $\mathcal{C}_{p,k}$. Hence, in total we can collect in $\mathcal{C}(r)$ at least $\lceil |\mathcal{C}_{p,k}|/B(r)\rceil = \lceil \binom{p}{k}/B(r)\rceil$ many solutions. Note that this construction satisfies Definition 3 as (i) holds by construction, and (ii) follows from the fact that if $\hat{\mathbf{x}} \notin \mathcal{C}(r)$, then we can perform another greedy step of the procedure and included it. Hence this proves Equation (27). To prove the second part of the lemma, note that $\binom{k}{d} \leq 2^k$ and $\binom{p-k}{k-l} \leq \binom{p}{k-r}$, hence we can get the following upper bound on $B(r)$.

$$B(r) \leq 2^k \binom{p}{k-r}(k-r) .$$

Plugging the latter in Equation (27) we get:

$$\begin{aligned}C(r) \geq \log \frac{\binom{p}{k}}{B(r)} \geq \log \binom{p}{k} + \\ - \left[k + \log \binom{p}{k-r} + \log(k-r)\right] .\end{aligned}$$

Using the standard concentration inequalities on the binomial coefficients in Lemma 11, we get

$$\log \binom{p}{k} \geq k(\log p - \log k) = k(1-\alpha_k)\log p$$

and

$$\begin{aligned}\log \binom{p}{k-r} &\leq (k-r)(1 + \log p - \log(k-r)) \\ &\lesssim (k-r)(1-\alpha_k)\log p ,\end{aligned}$$

where the last inequality follows from $r \leq \lambda k$ and, in particular, $\log(k-r) \geq \log(k-\lambda k) = \log k + \log(1-\lambda) \approx \log k \approx \alpha_k \log p$. Hence we finally get the bound for the rate of the cardinality $C(r)$:

$$\begin{aligned}\alpha_{C(r)} &\coloneqq \frac{\log C(r)}{\log p} \\ &\gtrsim k(1-r) - (k-r)(1-\alpha_k) - \frac{k + \log(k-r)}{\log p} \\ &= r(1-\alpha_k) - o(r) \approx r(1-\alpha_k) ,\end{aligned}$$

where the last equality is due to the fact that $r \in \omega(\frac{k}{\log p})$. □

Proof of Theorem 3. Let $\mathbf{Y}^{(u)}$ denote the $\binom{p}{d}$-dimensional vector obtained by the unfolding of the d-tensor $\mathbf{Y}$ into a vector containing its non-zero components (all $\mathbf{Y}_{i_1 i_2 \cdots i_d}$ for distinct d-tuples $i_1 < i_2 < \cdots < i_d$). Since each component of vector $\mathbf{Y}^{(u)}$ is a Gaussian r.v. according to Equation (5), the vector $\mathbf{Y}^{(u)}$ is also distributed as a Gaussian,

$$\mathbf{Y}^{(u)}(\hat{\mathbf{x}}) \sim \mathcal{N}\left(\beta(\hat{\mathbf{x}}^{\otimes d})^{(u)}, \mathrm{I}_{\binom{p}{d}}\right) := P_{\hat{\mathbf{x}}} \tag{29}$$

where $\mathrm{I}_{\binom{p}{d}}$ is the identity matrix in $\mathbb{R}^{\binom{p}{d}}$. For any $l \in \{0, \ldots, k\}$ and for $r = k - l$, let $\mathcal{C}(r)$ be a maximum-cardinality coverage defined according to Lemma 8. For any $\mathbf{x}' \in \mathcal{C}_{p,k}$ (possibly $\mathbf{x}' \notin \mathcal{C}(r)$) we consider its closest solution in $\mathcal{C}(r)$ according to the scalar product:

$$\mathcal{P}_{\mathcal{C}(r)}(\mathbf{x}') := \underset{\mathbf{x}'' \in \mathcal{C}(r)}{\operatorname{argmax}} \langle \mathbf{x}', \mathbf{x}'' \rangle \ .$$

Let $\tilde{\mathbf{x}} := \mathcal{P}_{\mathcal{C}(r)}(\mathbf{x}_{\mathrm{MLE}}(\mathbf{Y}))$, where $\mathbf{x}_{\mathrm{MLE}}(\mathbf{Y})$ is the MLE characterized in Equation (7). By Fano's inequality (Cover, 1999) we have that, for $\mathbf{x}'$ be chosen uniformly at random in $\mathcal{C}(r)$,

$$\begin{aligned} 1 - P_r &\geq \mathbb{P}\left(\tilde{\mathbf{x}} \neq \mathbf{x}'\right) \\ &\geq 1 - \frac{\mathrm{I}(\mathbf{x}; \mathbf{Y}) + \log 2}{\log C(r)} \end{aligned}$$

where $\mathrm{I}(\cdot;\cdot)$ denotes the mutual information. Using the generalized Fano's inequality (Han and Verdú, 1994) we further have:

$$\mathrm{I}(\mathbf{x}; \mathbf{Y}) \leq \frac{1}{C(r)^2} \sum_{\mathbf{x}'' \neq \mathbf{x}' \in \mathcal{C}(r)} D(P_{\mathbf{x}'} \| P_{\mathbf{x}''}) \leq \binom{k}{d} \beta^2 \ ,$$

where $D(\cdot||\cdot)$ denotes the Kullback–Leibler divergence and the second inequality follows from Lemma 16. Hence, from the definition of γ, we get

$$\begin{aligned} 1 - P_r \geq 1 - \frac{\binom{k}{d}\beta^2 + \log 2}{\log C(r)} &\approx 1 - \frac{2\gamma^2 k \cdot \log p}{\log C(r)} \\ &= 1 - \frac{2\gamma^2 k}{\alpha_{C(r)}} \end{aligned}$$

where in the last equality we used the definition of $\alpha_{C(r)}$ in Equation (9) and the previous approximation comes from the fact that $C(r) \to \infty$. The bound $\alpha_{C(r)} \gtrsim r(1 - \alpha_k)$ in Lemma 8, with $r = \lambda k$ and $\lambda \in (0, 1)$ constant, yields

$$P_r \lesssim 2\gamma^2 \frac{1}{\lambda(1 - \alpha_k)}$$

from which the claim follows easily from the arbitrariness of λ. □

B Useful lemmas

Lemma 9. *For any $X \sim \mathcal{N}(\mu, \sigma^2)$ and any $c > 0$, the following concentration inequalities hold:*

$$\left(\frac{1}{c} - \frac{1}{c^2}\right) \cdot \frac{e^{-c^2/2\sigma^2}}{\sqrt{2\pi}} \leq \mathbb{P}(X \geq \mu + c) \leq \frac{1}{c} \cdot \frac{e^{-c^2/2\sigma^2}}{\sqrt{2\pi}}.$$

Proof. See (Feller, 2008, Section 7.1). □

Lemma 10 (Theorem 2.2 in (Lopes, 2018))**.** *For any constant $\delta_0 \in (0, 1)$, the maximum of N possibly dependent Gaussian random variables $X_1, \ldots, X_N \sim \mathcal{N}(0, \sigma_X^2)$ satisfies*

$$\mathbb{P}\left(\max(X_1, \ldots, X_N) \leq \sigma_X \cdot \delta_0 \sqrt{2(1 - \rho_0) \log N}\right) \leq C \cdot \frac{\log^\eta N}{N^\xi} \tag{30}$$

where $C = C(\delta_0, \rho_0)$ is a constant depending only on δ_0 and ρ_0 and

$$\eta = \frac{1-\rho_0}{\rho_0}(1-\delta_0) \ , \qquad \text{and} \qquad \xi = \frac{1-\rho_0}{\rho_0}(1-\delta_0)^2 \ . \tag{31}$$

Lemma 11. *For any integers a and b the following concentration inequalities hold for the log binomial:*

$$a(\log b - \log a) \leq \log \binom{b}{a} \leq a(1 + \log b - \log a).$$

Lemma 12. *For any integers k, d and k' with $k' \leq k$ the following holds:*

$$\frac{k-k'}{\binom{k}{d} - \binom{k'}{d}} \leq \frac{k}{\binom{k}{d}}.$$

Proof. With simple manipulation, observe that this inequality is equivalent to

$$\frac{\binom{k'}{d}}{\binom{k}{d}} \leq \frac{k'}{k} \ .$$

For $k' < d$ this inequality is trivially satisfied since $\binom{k'}{d} = 0$. Otherwise we can write the previous inequality as

$$\begin{aligned}\frac{\binom{k'}{d}}{\binom{k}{d}} = \frac{k'!}{d!(k'-d)!}\frac{d!(k-d)!}{k!} &= \frac{k'(k'-1)\cdots(k'-d+1)}{k(k-1)\cdots(k-d+1)} \\ &\leq \frac{k'}{k}\end{aligned}$$

which is satisfied for any $k' \leq k$ since all these terms satisfy $\frac{k'-i}{k-i} \leq 1$, for $1 \leq i \leq d-1$. □

Lemma 13. *For any $\Delta > 0$, define $t_\Delta := \sqrt{2\Delta \log p}$. Then the following bound holds for any $\hat{\mathbf{x}} \in \mathcal{S}^F_{k'}$:*

$$\mathbb{P}\left(S^{-F}(\hat{\mathbf{x}}) \geq t_\Delta\right) \leq p(k', \Delta) := \left(\frac{1}{p}\right)^{\frac{\Delta}{D(k')}} \frac{1}{\sqrt{4\pi\Delta \log p}}. \tag{32}$$

Moreover, the following inequalities hold:

$$\mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{S}^{-F}_{k'}} S^{-F}(\hat{\mathbf{x}}) \geq t_\Delta\right) \leq Q(k') \cdot p(k', \Delta) \tag{33}$$

$$\mathbb{P}\left(S^{-F}(\mathbf{x}) < D(k')\beta - t_\Delta\right) \leq p(k', \Delta). \tag{34}$$

Proof. Observe that each $\hat{\mathbf{x}} \in \mathcal{S}^{(-F)}$, $S^{-F}(\hat{\mathbf{x}})$ consists of $D(k')$ non-biased edges, and therefore $S^{-F}(\hat{\mathbf{x}}) \sim \mathcal{N}(0, D(k'))$. Hence, by using the tail bound in Lemma 9 with $t = t_\Delta$ we have:

$$\begin{aligned}\mathbb{P}\left(S^{-F}(\hat{\mathbf{x}}) \geq t_\Delta\right) \leq \frac{1}{t_\Delta} \cdot \frac{e^{-t_\Delta^2/2D(k')}}{\sqrt{2\pi}} &= \frac{e^{-\Delta \log p/D(k')}}{t_\Delta\sqrt{2\pi}} \\ &= \frac{p^{-\Delta/D(k')}}{\sqrt{2\pi\Delta} \cdot \sqrt{2\log p}}\end{aligned}$$

which proves Equation (32). By the union bound and Equation (32) we obtain:

$$\begin{aligned}\mathbb{P}\left(\max_{\hat{\mathbf{x}} \in \mathcal{S}^{-F}_{k'}} S^{-F}(\hat{\mathbf{x}}) \geq t_\Delta\right) =& \mathbb{P}\left(\bigcup_{\hat{\mathbf{x}} \in \mathcal{S}^{-F}_{k'}} S^{-F}(\hat{\mathbf{x}}) \geq t_\Delta\right) \\ &\leq Q(k')p(k', \Delta)\end{aligned}$$

from which Equation (33) follows noting that $|\mathcal{S}_{k'}^{-F}| = Q(k')$. Finally, since all $D(k')$ edges of $S^{-F}(\mathbf{x})$ are biased, we have $S^{-F}(\mathbf{x}) \sim \mathcal{N}(D(k')\beta, D(k'))$. Therefore, by using Lemma 9 with $t = D(k')\beta + t_\Delta$, we get

$$\begin{aligned}\mathbb{P}\left(S^{-F}(\mathbf{x}) \le D(k')\beta - t_\Delta\right) &= \mathbb{P}\left(S^{-F}(\mathbf{x}) \ge D(k')\beta + t_\Delta\right)\\ &\le \frac{1}{t_\Delta} \cdot \frac{e^{-t_\Delta^2/2D(k')}}{\sqrt{2\pi}}.\end{aligned}$$

The remainder of the proof is as above. □

Lemma 14. *For $\gamma > \gamma_{UB_\epsilon}^{(k')}$, there exists t such that*

$$t_{\Delta'} < t < D(k')\beta - t_{\Delta''}$$

with

$$\frac{\Delta'}{D(k')} = \frac{\log M(k')}{\log p} + \epsilon \quad \text{and} \quad \frac{\Delta''}{D(k')} = \frac{\log \binom{k}{k'}}{\log p} + \epsilon\ . \tag{35}$$

Proof. We can rewrite the inequality $t_{\Delta'} < D(k')\beta - t_{\Delta''}$ as follows:

$$D(k')\beta > t_{\Delta'} + t_{\Delta''} = \sqrt{2\log p}(\sqrt{\Delta'} + \sqrt{\Delta''})$$

hence, by the rescaling in Equation (1), we get the condition for the SNR:

$$\begin{aligned}\frac{\beta}{\sqrt{2\log p}} &> \frac{1}{D(k')}(\sqrt{\Delta'} + \sqrt{\Delta''})\\ &= \sqrt{\frac{1}{D(k')}}\left(\sqrt{\frac{\log M(k')}{\log p} + \epsilon} + \sqrt{\frac{\log\binom{k}{k'}}{\log p} + \epsilon}\right)\\ &= \sqrt{\frac{1}{D(k')}} \cdot UB_\epsilon(k')\end{aligned}$$

from which the claim follows. □

Lemma 15. *For every k and $k' \in \{0, \dots, k-1\}$, it holds that*

$$\begin{aligned}\gamma_{UB_\epsilon}^{(k')} &\le \left(\sqrt{1 + \alpha_k - 2\alpha_k' + \epsilon} + \sqrt{\alpha_k - \alpha_k' + \epsilon}\right) \times\\ &\qquad \times \sqrt{\frac{(k-k')\binom{k}{d}}{(\binom{k}{d} - \binom{k'}{d})k}} \qquad (36)\\ &\le \left(\sqrt{1 + \alpha_k - 2\alpha_k' + \epsilon} + \sqrt{\alpha_k - \alpha_k' + \epsilon}\right)\ . \qquad (37)\end{aligned}$$

Proof. Using the standard inequalities on binomial coefficients in Lemma 11 and the definition of rate in Equation (9) we have

$$\begin{aligned}\frac{\log\binom{k}{k'}}{\log p} &\le (k-k')\frac{1 + \log k - \log(k-k')}{\log p}\\ &\approx (k-k')(\alpha_k - \alpha_{k-k'}),\end{aligned}$$

and analogously

$$\begin{aligned}\frac{\log\binom{p-k}{k-k'}}{\log p} \le& (k-k')\frac{1 + \log(p-k) - \log(k-k')}{\log p}\\ &\approx (k-k')(1 - \alpha_{k-k'}),\end{aligned}$$

thus implying

$$\frac{\log M(k')}{\log n} = \frac{\log \left(\binom{k}{k'} \binom{p-k}{k-k'} \right)}{\log p} \lesssim (k-k')(1+\alpha_k - 2\alpha_{k-k'}) \ .$$

By plugging this into the definition of $UB_\epsilon(k')$ in Equation (20), we get

$$\begin{aligned} UB_\epsilon(k') &= \sqrt{\frac{\log M(k')}{\log p} + \epsilon} + \sqrt{\frac{\log \binom{k}{k'}}{\log p} + \epsilon} \\ &\lesssim \sqrt{1+\alpha_k - 2\alpha_{k-k'} + \epsilon} + \sqrt{\alpha_k - \alpha_{k-k'} + \epsilon} \ . \end{aligned}$$

Hence, using the definition of $\gamma_{UB_\epsilon}^{(k')}$ given in Lemma 3, we obtain Equation (36). To conclude the proof and have Equation (37) we finally use the simple manipulations of the Lemma 12. □

Lemma 16. *For any two vectors* $\mathbf{x}', \mathbf{x}'' \in \mathcal{C}_{p,k}$ *we have*

$$D(P_{\mathbf{x}'} \| P_{\mathbf{x}''}) \leq \binom{k}{d} \beta^2 \ ,$$

where $D(\cdot\|\cdot)$ *denotes the Kullback-Leiber divergence.*

Proof. Since $P_{\mathbf{x}'}$ and $P_{\mathbf{x}''}$ are a Gaussian probability distributions (29), we have

$$\begin{aligned} D(P_{\mathbf{x}'} \| P_{\mathbf{x}''}) &= \frac{1}{2}\beta^2 \|(\mathbf{x}'^{\otimes d})^{(u)} - (\mathbf{x}''^{\otimes d})^{(u)}\|^2 \\ &= \beta^2 (\|(\mathbf{x}'^{\otimes d})^{(u)}\|^2 - \langle (\mathbf{x}'^{\otimes d})^{(u)}, (\mathbf{x}''^{\otimes d})^{(u)} \rangle) \\ &\leq \beta^2 \|(\mathbf{x}'^{\otimes d})^{(u)}\|^2 = \beta^2 \binom{k}{d} \ . \end{aligned}$$

□

C Derivations for Section 3 (AMP algorithm)

C.1 Derivation of the AMP iterative equations

In the following we adopt conveniently the notation used in (Lesieur et al., 2017b), denoting as

$$\boldsymbol{S} = \beta \sqrt{\binom{p-1}{d-1}} \boldsymbol{Y}.$$

Throughout this section, we shall make the following assumption on the above rescaled tensor, i.e., on how β scales with respect to the entries of $\boldsymbol{Y}$.

Assumption 1. *For every d-tuple* $i_1 i_2 \cdots i_d$ *of distinct indices, it holds that* $S_{i_1 i_2 \cdots i_d} \leq 1$, *that is,* $\beta \leq 1/\sqrt{\binom{p-1}{d-1}} Y_{i_1 i_2 \cdots i_d}$.

A stronger assumption is the following one:

Assumption 2. *For every d-tuple* $i_1 i_2 \cdots i_d$ *of distinct indices, it holds that* $S_{i_1 i_2 \cdots i_d} \in o(1)$.

We next present a message passing algorithm which is a simple generalization of the algorithm for matrix (Lesieur et al., 2017a) and tensor-PCA (Lesieur et al., 2017b). The algorithm is described by the following iterative equations:

$$
\begin{aligned}
x^{(t)}_{i\to ii_2\dots i_d} &= \frac{1}{\sqrt{\binom{p-1}{d-1}}} \sum_{\substack{k_2<\dots<k_d \\ (k_2,\dots,k_d)\neq(i_2,\dots,i_d)}} \left(S_{ik_2\dots k_d}\cdot \hat{x}^{(t)}_{k_2\to ik_2\dots k_d}\cdot\dots\cdot \hat{x}^{(t)}_{k_d\to ik_2\dots k_d}\right) \\
A^{(t)}_{i\to ii_2\dots i_d} &= \frac{1}{\binom{p-1}{d-1}} \sum_{\substack{k_2<\dots<k_d \\ (k_2,\dots,k_d)\neq(i_2,\dots,i_d)}} \left(S^2_{ik_2\dots k_d}\cdot (\hat{x}^{(t)}_{k_2\to ik_2\dots k_d})^2\cdot\dots\cdot (\hat{x}^{(t)}_{k_d\to ik_2\dots k_d})^2\right) \\
\left(\hat{x}^{(t+1)}_{i\to ii_2\dots i_d}\right)_i &= f\left((A^{(t)}_{i\to ii_2\dots i_d})_i, (x^{(t)}_{i\to ii_2\dots i_d})_i\right)
\end{aligned}
$$

where $f(\cdot,\cdot)$ is the multidimensional threshold function defined in Equation (14), and $\left(\hat{x}^{(t+1)}_{i\to ii_2\dots i_d}\right)_i$ denotes the p-dimensional vector indexed by i. Note that the function f is applied element-wise on all indices other than i, that is, on $i_2,\dots,i_d$. Since the messages depend weakly on the target factor, we can compute the messages including the factor $(i_2,\dots,i_d)$ in the sums above, and dropping the factor node dependency as:

$$
\begin{aligned}
x^{(t)}_i &= \frac{1}{\sqrt{\binom{p-1}{d-1}}} \sum_{i_2<\dots<i_d} \left(S_{ii_2\dots i_d}\cdot \hat{x}^{(t)}_{i_2\to ii_2\dots i_d}\cdot\dots\cdot \hat{x}^{(t)}_{i_h\to ii_2\dots i_d}\right) && (38) \\
A^{(t)}_i &= \frac{1}{\binom{p-1}{d-1}} \sum_{i_2<\dots<i_d} \left(S^2_{ii_2\dots i_d}\cdot (\hat{x}^{(t)}_{i_2\to ii_2\dots i_d})^2\cdot\dots\cdot (\hat{x}^{(t)}_{i_d\to ii_2\dots i_d})^2\right) && (39) \\
\hat{\boldsymbol{x}}^{(t+1)} &= f\left(\boldsymbol{A}^{(t)}, \boldsymbol{x}^{(t)}\right).
\end{aligned}
$$

We can now analyze the error obtained by this simplification as:

$$
\begin{aligned}
x^{(t)}_i - x^{(t)}_{i\to ii_2\dots i_h} &= \frac{1}{\sqrt{\binom{p-1}{d-1}}} \left(S_{ii_2\dots i_d}\cdot \hat{x}^{(t)}_{i_2\to ii_2\dots i_d}\cdot\dots\cdot \hat{x}^{(t)}_{i_d\to ii_2\dots i_d}\right) && (40) \\
A^{(t)}_i - A^{(t)}_{i\to ii_2\dots i_d} &= \frac{1}{\binom{p-1}{d-1}} \left(S^2_{ii_2\dots i_d}\cdot (\hat{x}^{(t)}_{i_2\to ii_2\dots i_d})^2\cdot\dots\cdot (\hat{x}^{t}_{i_d\to ii_2\dots i_d})^2\right). && (41)
\end{aligned}
$$

Observe that by Assumption 1, the error in A is of lower order then the one in x. Hence, we can estimate the error in $\hat{\boldsymbol{x}}^{(t+1)}$ by focusing on the quantity in Equation (40):

$$
\begin{aligned}
\hat{x}^{(t)}_{i\to ii_2\dots i_d} - \hat{x}^{(t)}_i &= f\left((A^{(t-1)}_{i\to ii_2\dots i_d})_i, (x^{(t-1)}_{i\to ii_2\dots i_d})_i\right)_i - f(\boldsymbol{A}^{(t-1)}, \boldsymbol{x}^{(t-1)})_i \\
&= -\partial_{x_i} f(\boldsymbol{A}^{(t-1)}, \boldsymbol{x}^{(t-1)})_i \frac{1}{\sqrt{\binom{p-1}{d-1}}} \left(S_{ii_2\dots i_d}\cdot x^{(t-1)}_{i_2\to ii_2\dots i_d}\cdot\dots\cdot x^{(t-1)}_{i_d\to ii_2\dots i_d}\right) + o\left(\frac{S_{ii_2\dots i_d}}{\sqrt{\binom{p-1}{d-1}}}\right) \\
&= -\partial_{x_i} f(\boldsymbol{A}^{(t-1)}, \boldsymbol{x}^{(t-1)})_i \frac{1}{\sqrt{\binom{p-1}{d-1}}} \left(S_{ii_2\dots i_d}\cdot x^{(t-1)}_{i_2}\cdot\dots\cdot x^{(t-1)}_{i_d}\right) + o\left(\frac{S_{ii_2\dots i_d}}{\sqrt{\binom{p-1}{d-1}}}\right). && (42)
\end{aligned}
$$

Plugging this last expansion into Equation (38), and by recalling the definition of $\boldsymbol{\sigma}^{(t)}$ in Equation (15), we get to leading order:

$$
x^{(t)}_i = \frac{1}{\sqrt{\binom{p-1}{d-1}}} \sum_{i_2<\dots<i_d} S_{ii_2\dots i_d}\hat{x}^{(t)}_{i_2}\cdot\ldots\cdot\hat{x}^{(t)}_{i_d} \frac{d-1}{\binom{p-1}{d-1}} \sum_{i_2} \sigma^{(t)}_{i_2} \sum_{i_3<\dots<i_d} S^2_{ii_2\dots i_d}\hat{x}^{(t)}_{i_3}\hat{x}^{(t-1)}_{i_3}\cdot\ldots\cdot\hat{x}^{(t)}_{i_d}\hat{x}^{(t-1)}_{i_d} + o\left(\frac{S^2_{ii_2\dots i_d}}{\binom{p-1}{d-1}}\right)
$$

where the factor $d-1$ comes from the symmetry of choosing the index (i_2 in our case) for the lower order term in the product. Note that all the other terms are of lower order and are hence discarded. Plugging Equation (42) into Equation (39) we obtain:

$$
A^{(t)}_i = \frac{1}{\binom{p-1}{d-1}} \sum_{i_2<\dots<i_d} S^2_{ii_2\dots i_d}\cdot(\hat{x}^{(t)}_{i_2})^2\cdot\dots\cdot(\hat{x}^{(t)}_{i_d})^2 + o\left(\frac{S^2_{ii_2\dots i_d}}{\binom{p-1}{d-1}}\right).
$$

C.2 Derivation of the state evolution

We assume here a typical condition for belief propagation algorithm, that is the statistical independence between the rescaled observation $\boldsymbol{S}$ and the estimates $\hat{\mathbf{x}}$ (see for reference Montanari (2013); Lesieur et al. (2017a)). Given the independence, the central limit theorem applies for the right hand side of both Equation (38) and Equation (39), so that the messages behave in the large p limit as Gaussian random variables. We can thus describe the evolution of the AMP algorithm by tracking only the average and variance of such messages (recall that $\mathbf{x}$ is the planted solution):

$$
\begin{aligned}
\mathbb{E}_{\mathbf{Z}}[x_i^{(t)}] &= \frac{1}{\sqrt{\binom{p-1}{d-1}}} \sum_{i_2<\cdots<i_d} \left(\mathbb{E}_{\mathbf{Z}}[S_{ii_2\dots i_d}] \cdot \hat{x}^{(t)}_{i_2\to ii_2\dots i_d} \cdot \dots \cdot \hat{x}^{(t)}_{i_d\to ii_2\dots i_d}\right) \\
&= \beta^2 \sum_{i_2<\cdots<i_d} \left(x_i \cdot x_{i_2} \cdot \dots \cdot x_{i_d} \cdot \hat{x}^{(t)}_{i_2\to ii_2\dots i_d} \cdot \dots \cdot \hat{x}^{(t)}_{i_d\to ii_2\dots i_d}\right) \\
&= \beta^2 \sum_{i_2<\cdots<i_d} x_i \cdot x_{i_2} \cdot \dots \cdot x_{i_d} \cdot \hat{x}^{(t)}_{i_2} \cdot \dots \cdot \hat{x}^{(t)}_{i_d} + o(\beta^2) \\
&= \beta^2 x_i \cdot \mathbf{1}\{\boldsymbol{x} \circ \hat{\boldsymbol{x}}^{(t)}\}_i + o(\beta^2) \ .
\end{aligned} \tag{43}
$$

Analogously

$$
\begin{aligned}
\mathbb{V}_{\boldsymbol{Z}}[x_i^{(t)}] &= \frac{1}{\binom{p-1}{d-1}} \sum_{i_2<\cdots<i_d} \left(\mathbb{V}_{\boldsymbol{Z}}[S_{ii_2\dots i_d}] \cdot (\hat{x}^{(t)}_{i_2\to ii_2\dots i_d})^2 \cdot \dots \cdot (\hat{x}^{(t)}_{i_d\to ii_2\dots i_d})^2\right) \\
&= \beta^2 \sum_{i_2<\cdots<i_d} (\hat{x}^{(t)}_{i_2\to ii_2\dots i_d})^2 \cdot \dots \cdot (\hat{x}^{(t)}_{i_d\to ii_2\dots i_d})^2 \\
&= \beta^2 \sum_{i_2<\cdots<i_d} (x^{(t)}_{i_2})^2 \cdot \dots \cdot (x^{(t)}_{i_d})^2 + o(\beta^2) \\
&= \beta^2 \mathbf{1}\{\hat{\boldsymbol{x}}^t \circ \hat{\boldsymbol{x}}^t\}_i + o(\beta^2)
\end{aligned} \tag{44}
$$

and

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{Z}}[A_i^{(t)}] &= \frac{1}{\binom{n-1}{d-1}} \sum_{i_2<\cdots<i_d} \left(\mathbb{E}_{\boldsymbol{Z}}[S^2_{ii_2\dots i_d}] \cdot (\hat{x}^{(t)}_{i_2\to ii_2\dots i_d})^2 \cdot \dots \cdot (\hat{x}^{(t)}_{i_d\to ii_2\dots i_d})^2\right) \\
&= \beta^2 \sum_{i_2<\cdots<i_d} (\beta^2 (x_i)^2 \cdot (x_{i_2})^2 \cdot \dots \cdot (x_{i_d})^2 + 1) \cdot (\hat{x}^{(t)}_{i_2})^2 \cdot \dots \cdot (\hat{x}^{(t)}_{i_d})^2 + o(\beta^2) \\
&= \beta^4 x_i \cdot \mathbf{1}\{\boldsymbol{x} \circ \boldsymbol{x} \circ \hat{\boldsymbol{x}}^{(t)} \circ \hat{\boldsymbol{x}}^t\}_i + \beta^2 \mathbf{1}\{\hat{\boldsymbol{x}}^{(t)} \circ \hat{\boldsymbol{x}}^{(t)}\}_i + o(\beta^2) \\
&= \beta^2 \mathbf{1}\{\hat{\boldsymbol{x}}^{(t)} \circ \hat{\boldsymbol{x}}^{(t)}\}_i + o(\beta^2)
\end{aligned} \tag{45}
$$

where in the second equality we used the fact that

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{Z}}[S^2_{ii_2\dots i_h}] =& \beta^2 (x_i)^2 \cdot (x_{i_2})^2 \cdot \dots \cdot (x_{i_h})^2 + \mathbb{E}_{\boldsymbol{Z}}[Z^2_{ii_2\dots i_h}] + \beta x_i \cdot x_{i_2} \cdot \dots \cdot x_{i_h} \mathbb{E}_{\boldsymbol{Z}}[Z_{ii_2\dots i_h}] \\
=& \beta^2 (x_i)^2 \cdot (x_{i_2})^2 \cdot \dots \cdot (x_{i_h})^2 + 1.
\end{aligned}
$$

We now assume that the signal estimates $\hat{\boldsymbol{x}}^t$ are drawn from the true intractable posterior distribution $\mathbb{P}(\cdot|\boldsymbol{x}, \boldsymbol{Z})$. Given this assumption and using the Nishimori condition (see Iba (1999); Lesieur et al. (2017b)) we obtain easily from Equation (44) and Equation (45):

$$
\begin{aligned}
\mathbb{V}_{\boldsymbol{x},\boldsymbol{Z}}[\hat{\boldsymbol{x}}^{(t)}] = \mathbb{E}_{\boldsymbol{x},\boldsymbol{Z}}[\boldsymbol{A}^{(t)}] &= \beta^2 \mathbb{E}_{\boldsymbol{x}} \mathbf{1}\{\hat{\boldsymbol{x}}^{(t)} \circ \hat{\boldsymbol{x}}^{(t)}\} + o(\beta^2) \\
&= \beta^2 \mathbb{E}_{\boldsymbol{x}} \mathbf{1}\{\boldsymbol{x} \circ \hat{\boldsymbol{x}}^{(t)}\} + o(\beta^2) \\
&= \hat{\boldsymbol{m}}^{(t)} + o(\beta^2).
\end{aligned}
$$

It can be easily seen that the *variance* of the messages A is of lower order, hence $\boldsymbol{A}^{(t)}$ can be approximated by only its mean. With the assumed gaussianity of messages, we can hence write using Equation (43) and

Equation (44) $\boldsymbol{x}^{(t)} = \hat{\boldsymbol{m}}^{(t)} \circ \boldsymbol{x} + \sqrt{\hat{\boldsymbol{m}}^{(t)}} \circ \boldsymbol{z}$, with $\boldsymbol{z}$ being a p-dimensional standard Gaussian vector. The state evolution finally reads:

$$\begin{aligned}\boldsymbol{m}^{(t+1)} &= \frac{1}{\binom{p-1}{d-1}} \mathbb{E}_{\boldsymbol{x},\boldsymbol{z}}[\mathbf{1}\{\boldsymbol{x} \circ \hat{\boldsymbol{x}}^{(t+1)}\}] \\ &= \frac{1}{\binom{p-1}{d-1}} \mathbb{E}_{\boldsymbol{x},\boldsymbol{z}}[\mathbf{1}\{\boldsymbol{x} \circ f(\hat{\boldsymbol{m}}^{(t)}, \hat{\boldsymbol{m}}^{t} \circ \boldsymbol{x} + \sqrt{\hat{\boldsymbol{m}}^{(t)}} \circ \boldsymbol{z})\}] \,.\end{aligned}$$

C.3 Analytical threshold for AMP recovery

To get an analytical threshold, we start from the SE for factorizable prior as in Equation (17) and we study the fixed point of the recursive equation for the parameter of the Bernulli distribution $\delta = k/p \to 0$ in the large system limit. In particular the threshold function reads in the limit $f(a,x) = \delta e^{x-a/2} + O(\delta^2)$, hence the factorized SE becomes $m_{t+1} = \delta^2 \mathbb{E}_z[e^{\hat{m}_t/2+\sqrt{\hat{m}_t}z}] = \delta^2 e^{\hat{m}_t}$ with $\hat{m}_t = \beta^2 \binom{p-1}{d-1} m_t$. It is easy to see that the critical bias β to have a perfect overlap with the planted signal is

$$\beta_{\mathrm{AMP}} \coloneqq \sqrt{\frac{1}{e(d-1)} \frac{p^{2(h-1)}}{\binom{p-1}{d-1}} \frac{1}{k^{2(d-1)}}}.$$

Using the definition of normalized SNR γ in Equation (1), we obtain the threshold γ_{AMP} in Claim 1.

Remark 3 (Validity of the AMP approximations)**.** *We can observe from the derivations above that the AMP equations and state evolution are carried out assuming the quantity*

$$\frac{S_{i_1,\dots i_h}}{\sqrt{\binom{p-1}{d-1}}} = \beta \sqrt{\binom{p-1}{d-1}} Y_{i_1,\dots i_d}$$

being small (which corresponds to Assumption 2 above). This is the case for the classic tensor-PCA for both dense and sparse signal with linear sparsity as in Lesieur et al. (2017b). However in the scenario here considered the effective sparsity of the problem defined by the parameter $\delta = k/p$ can be sub-linear, and hence non-trivial estimation requires a β (hence a SNR) such that the quantity above is not in $o(1)$. For this reason the derivations have to be considered non-rigorous in the regime used in this analysis. The rigorous presentation of AMP-like algorithms in effectively sub-linear sparse estimation problem, also in line of the approach proposed recently in Barbier et al. (2020), is left for future developments.

D Comparison with Bounds in the literature

D.1 Tensor-PCA formulation

We follow the terminology of (Hopkins et al., 2015) to explain the similarities/differences between tensor-PCA and our problem.

Definition 4 (Symmetric Tensor-PCA)**.** *Given an input tensor $\mathbf{Y} = \tau \cdot \mathbf{v}^{\otimes d} + \mathbf{Z}$, where $\mathbf{v} \in \mathbb{R}^p$ is an arbitrary* unit *vector, $\tau \geq 0$ is the signal-to-noise ratio, and $\mathbf{Z}$ is a random noise tensor with iid standard Gaussian entries, recover the signal v approximately. Moreover, the noise tensor is symmetric and thus so is the input tensor as well, that is, $\mathbf{Z}_{\pi(i_1)\pi(i_2)\cdots\pi(i_h)} = \mathbf{Z}_{i_1 i_2 \cdots i_d}$ and $\mathbf{Y}_{\pi(i_1)\pi(i_2)\cdots\pi(i_h)} = \mathbf{Y}_{i_1 i_2 \cdots i_d}$ for any permutation π.*

Definition 5 (Planted k-Densest Sub-Hypergraph)**.** *This problem is a variant of symmetric tensor-PCA in which we impose the following additional structure:*

1. *We consider only the $\binom{p}{d}$ entries with* distinct *indices, that is, $\mathbf{Y}_{i_1 i_2 \cdots i_d} = 0$ whenever $i_a = i_b$ for some a and b.*

2. *The vector $\mathbf{v}$ encodes a planted k-Subgraph and thus has* exactly *k entries equal to $1/\sqrt{k}$, and all other entries are equal to 0.*

Remark 4 (Impact of diagonal entries – Item 1)**.** *Dropping Condition 1 leads to the variant in which we make the substitution $\binom{p}{d} \mapsto p^d$ in Item 1 above.*

Remark 5 (Rescalings). *Different papers consider different rescaling of the signal-to-noise ratio, that are here reported for convenience in Table 1. In general, consider a tensor*

$$\mathbf{Y} = \mu \cdot \mathbf{v}^{\otimes d} + \mathbf{Z} \qquad\qquad \mathbf{Z} \sim \mathcal{N}(0, \sigma^2) \tag{46}$$

where $\mathbf{v}$ *is a vector of unit length, and* μ *and* σ^2 *(signal and noise, respectively) determine the snr. By simply rescaling so that we have normally distributed Gaussian noise, this is the same as*

$$\mathbf{Y} = \mu/\sigma \cdot \mathbf{v}^{\otimes d} + \mathbf{Z} \qquad\qquad \mathbf{Z} \sim \mathcal{N}(0, 1)$$

and since in our formulation (snr in Equation (1)*) we consider the planted solution as a 0-1 vector* $\mathbf{x}$ *consisting of* k *ones and* $p-k$ *zeros, we are effectively considering a planted signal* $\beta \cdot \mathbf{x}^{\otimes d} = \beta\sqrt{k^d} \cdot \mathbf{v}^{\otimes d}$ *where* $\mathbf{v} = \mathbf{x}/\sqrt{k}$ *is a unitary vector. Therefore, the tensor-PCA formulation in Equation* (46) *corresponds to*

$$\beta\sqrt{k^d} = (\mu/\sigma) \qquad \Leftrightarrow \qquad \gamma = (\mu/\sigma)\sqrt{\frac{\binom{k}{d}}{k^{d+1}} \cdot \frac{1}{2\log p}}$$

We use this relation to convert the existing bounds for tensor-PCA in the literature to our snr γ *as shown in Table 1. In [*$\star$*] we ignore the diagonal entries (see Remark 4). As we are implicitly considering the easier problem with the additional entries, the bounds that one obtains are in a sense "optimistic" for our original problem.*

Table 1: Signal to noise ratio scaling in the literature

Tensor	**Noise**	**SNR ($\star$=ours)**
$\mathbf{Y} = \beta \cdot \mathbf{v}^{\otimes d} + \mathbf{Z}$	$Z_{i_1 i_2 \cdots i_d} \sim \mathcal{N}(0, 1/(p(d-1)!))$	β (Richard and Montanari, 2014)
$\mathbf{Y} = \beta' \cdot \mathbf{v}^{\otimes d} + \mathbf{Z}$	$Z_{i_1 i_2 \cdots i_d} \sim \mathcal{N}(0, 2/(p \cdot d!))$	β' (Montanari et al., 2016)
$\mathbf{Y} = \beta'' \cdot \mathbf{v}^{\otimes d} + \mathbf{Z}$	$Z_{i_1 i_2 \cdots i_d} \sim \mathcal{N}(0, 2/(p \cdot d!))$	β'' (Perry et al., 2020)
$\mathbf{Y} = \tau \cdot \mathbf{v}^{\otimes d} + \mathbf{Z}$	$Z_{i_1 i_2 \cdots i_d} \sim \mathcal{N}(0, 1)$	τ (Hopkins et al., 2015)
$\mathbf{Y} = \lambda\sqrt{p} \cdot \mathbf{v}^{\otimes d} + \mathbf{Z}$	$Z_{i_1 i_2 \cdots i_d} \sim \mathcal{N}(0, 1)$	λ (Jagannath et al., 2020; Arous et al., 2020)
$\mathbf{Y} = \sqrt{\lambda_p} \cdot \mathbf{v}^{\otimes d} + \mathbf{Z}$	$Z_{i_1 i_2 \cdots i_d} \sim \mathcal{N}(0, 1)$	λ_p (Niles-Weed and Zadik, 2020)
$\mathbf{Y} = \beta\sqrt{k^d} \cdot \mathbf{v}^{\otimes d} + \mathbf{Z}$	$Z_{i_1 i_2 \cdots i_d} \sim \mathcal{N}(0, 1)$	γ eq. (1) $\star$

The information-theoretic bounds translated in our scale read as follows:

$$\text{lower bound (Richard and Montanari, 2014):} \quad \gamma \leq \sqrt{\frac{p \cdot d!\binom{k}{d}}{k^{d+1}} \frac{1}{20\log p}}$$

$$\text{upper bound (Richard and Montanari, 2014):} \quad \gamma \geq \sqrt{\frac{p \cdot d!\binom{k}{d}}{k^{d+1}} \frac{\log d}{2\log p}}$$

Sharper bounds have been obtained for detection and (weak) recoverability:

$$\text{generic spherical prior (Perry et al., 2020):} \quad \gamma = \sqrt{\frac{p \cdot d!\binom{k}{d}}{k^{d+1}} \frac{\log d}{2\log p}}$$

$$\text{Radamacher prior (Perry et al., 2020):} \quad \gamma = \sqrt{\frac{p \cdot d!\binom{k}{d}}{k^{d+1}} \frac{\log d}{4\log p}}$$

where the Radamacher prior bounds apply to one of the following restrictions: (i) the dense regimes with any sparsity constant $\rho \in (0, 1]$ and $d \to \infty$ or (2) the vanishing sparsity regime $\rho \to 0$ and constant d. Note that in both cases (Richard and Montanari, 2014) and (Perry et al., 2020), the bound are located at a scale $\sqrt{\frac{p}{k\log p}}$ that diverges for any rate $\alpha_k < 1$ considered in this paper. From this result, we can observe that the recovery above the γ_{UB} given in this paper is possible only thanks to the exploitation of the prior constraint, and it is not possible in general.

Sharp bounds on the MMSE estimator have been also obtained (Niles-Weed and Zadik, 2020) with the same Bernoulli prior considered here, and translate into the threshold

$$\gamma_{\mathrm{MMSE}} = \sqrt{1-\alpha_k}$$

which can be obtained by a proper rescale of the planted vector $\mathbf{v}$ so that the resulting tensor (without diagonal entries) has unit length. Note that these bounds on the MMSE regard the problem of finding a vector with a positive non-vanishing correlation with the planted vector (weak recovery), and correspond to the lower bound γ_{LB} for the MLE provided here in the case of $d \to \infty, d \in o(k)$. Whether the MLE undergoes the same transition as the MMSE is in general open (Niles-Weed and Zadik, 2020) and, for our problem, it has been shown a relation only for the "ultra sparse" regime $k = o(\log^{\frac{1}{4d-1}} p)$ (Corinzia et al., 2021).

Algorithmic upper bounds provided by sum-of-squares (SOS) algorithms (Hopkins et al., 2015) are:

$$\gamma_{\mathrm{SOS}} := \sqrt{\frac{p^{d/2}\binom{k}{d}}{k^{d-1}} \frac{1}{2\log^{1/2} p}} \tag{47}$$

for any $d \geq 3$. Note that also for the computational threshold, this bound is higher than the AMP threshold $\gamma_{\mathrm{AMP}} \approx \sqrt{p^{(1-\alpha_k)(d-1)}}$ for

$$\alpha_k > 1/2 - 1/d.$$

A recent work by Choo and d'Orsi (2021) derives further SOS algorithmic bounds for the tensorial version of our problem when $\alpha_k \leq 1/2$. Their algorithm runs in time $\tilde{O}(p^{d+t})$ for any parameter $1 \leq t \leq k$ such that *their* SNR $\lambda \geq \tilde{O}(\sqrt{t} \cdot (k/t)^{d/2})$ where the $\tilde{O}(\cdot)$ notation hides multiplicative factors logarithmic in p. It is open whether their analysis extends to our problem (without diagonal entries available) as the authors discuss in (Choo and d'Orsi, 2021, Appendix A).

Further papers provide general bounds whose thresholds do *not* have a closed form and apply to the easier problem of detection or hypothesis testing:

$$\text{(Montanari et al., 2016)} \quad \beta_d^2 := \inf_{q \in (0,1)} \sqrt{-\frac{1}{q^d}\log(1-q^2)}$$

$$\text{(Jagannath et al., 2020)} \quad \lambda_c := \sup_{\lambda \geq 0} \left\{ \sup_{t \in [0,1)} f_\lambda(t) \leq 0 \right\}$$

with $f_\lambda(t) = \lambda^2 t^d + \log(1-t) + t$.

D.2 Prior bounds for the k densest subhypergraph

We report the prior upper and lower bounds on the very same problem in (Corinzia et al., 2019, Theorem 5). For the sake of comparison, we rewrite the upper and lower bound there according to our scaled-normalized snr γ, and denote these bounds as γ_{lb} and γ_{ub}, respectively. As we can see below, these bounds are very loose in most of the cases, $\gamma_{lb} \ll \gamma_{LB} \leq \gamma_{UB} \ll \gamma_{ub}$:

$$\gamma_{lb} = \sqrt{\frac{1}{d}}$$

and

$$\gamma_{ub} = \begin{cases} \sqrt{2} & \frac{\binom{k}{d}}{k}\frac{1}{\log p} \to 0 \\ 2\sqrt{1+c(1+\log 2)} & \frac{\binom{k}{d}}{k}\frac{1}{\log p} \to c \in (0,+\infty) \\ 2\sqrt{\frac{\binom{k}{d}}{k\log p} \cdot \frac{1+\log 2}{1-\alpha_k}} & \alpha_k \in (0,1) \end{cases}$$

For instance, when $1 \ll d \ll k$ we have $\gamma_{lb} \to 0$ and $\gamma_{ub} \to +\infty$.