ON THE ACCURACY OF HOTELLING-TYPE TENSOR DEFLATION: A RANDOM TENSOR ANALYSIS

Mohamed El Amine Seddik, Maxime Guillaud, Alexis Decurninge

Mathematical and Algorithmic Sciences Laboratory, Huawei Technologies France

ABSTRACT

Leveraging on recent advances in random tensor theory, we consider in this paper a rank-r asymmetric spiked tensor model of the form $\sum_{i=1}^{r} \beta_i \mathcal{A}_i + \mathcal{W}$ where $\beta_i \geq 0$ and the $\mathcal{A}_i$'s are rank-one tensors such that $\langle \mathcal{A}_i, \mathcal{A}_j \rangle \in [0, 1]$ for $i \neq j$, based on which we provide an asymptotic study of Hotelling-type tensor deflation in the large dimensional regime. Specifically, our analysis characterizes the singular values and alignments at each step of the deflation procedure, for asymptotically large tensor dimensions. This can be used to construct consistent estimators of different quantities involved in the underlying problem, such as the signal-to-noise ratios β_i or the alignments between the different signal components $\langle \mathcal{A}_i, \mathcal{A}_j \rangle$.

Index Terms— Random Tensor Theory, Hotelling Deflation, Low-rank Tensor Decomposition, Parameter Estimation.

1. INTRODUCTION

The analysis of random tensors has attracted significant attention in the last decade since the introduction of the concept of tensor PCA, which generalizes principal component analysis to high-order arrays. The first model introduced in [1] is the so-called spiked tensor model of the form $\beta \boldsymbol{x}^{\otimes d} + \mathcal{W}/\sqrt{n}$ where $\boldsymbol{x} \in \mathbb{R}^n$ is some high-dimensional unit vector referred to as a *spike*, $\mathcal{W}$ is a *symmetric* random tensor of order d having standard Gaussian entries and $\beta \geq 0$ is a parameter controlling the signal-to-noise ratio.

Follow-up results have improved the understanding of the behavior of the spiked model and allowed to identify theoretical and/or algorithmic guarantees in terms of efficient signal recovery. In particular, several works [2, 3, 4, 5, 6] have focused on the asymptotic (large dimensional) regime $n \to \infty$. We briefly summarize their main findings as follows: for $d \geq 3$, it has been shown that there exists a statistical threshold $\beta_{stat} = O(1)$ below which it is information-theoretically impossible to recover or even detect the spike, while above β_{stat} recovery is theoretically possible finding a critical point of the square loss. Moreover, the asymptotic alignment $\langle \boldsymbol{x}, \boldsymbol{u} \rangle$ between $\boldsymbol{x}$ and a critical point $\boldsymbol{u}$ is given in terms of β. Besides, since almost all tensor problems, e.g. finding the critical points, are NP-hard [7], many researchers were interested in exhibiting an algorithmic threshold for β above which re-

covery could be possible with a polynomial-time algorithm. The authors in [1] introduced a method for estimating $\boldsymbol{x}$ based on tensor unfolding and showed that spike recovery is possible above the algorithmic threshold $\beta_{algo} = O(n^{\frac{d-2}{4}})$.

These ideas were further generalized to the *asymmetric* spiked tensor model of the form $\beta \boldsymbol{x}_1 \otimes \cdots \otimes \boldsymbol{x}_d + \mathcal{W}/\sqrt{\sum_i n_i}$, where $\boldsymbol{x}_i \in \mathbb{R}^{n_i}$ are unit vectors and $\mathcal{W}$ is a random tensor with standard Gaussian i.i.d. entries. In particular, [8] provided an analysis of the unfolding method for asymmetric tensors and determined the algorithmic threshold to be $\beta_{algo} = O(n^{\frac{d-2}{4}})$ when $n_i = n$ for all i, while [9] showed the existence of a statistical threshold $\beta_s = O(1)$ above which a local solution $\boldsymbol{u}_i$ of the MLE aligns with the signal, and further quantified the asymptotic alignments $\langle \boldsymbol{x}_i, \boldsymbol{u}_i \rangle$.

Contribution: In this paper, we address the extension of these ideas to a more general setting, namely asymmetric *low-rank* spiked tensors of the form $\sum_{i=1}^{r} \beta_i \boldsymbol{x}_{i,1} \otimes \cdots \otimes \boldsymbol{x}_{i,d} + \mathcal{W}/\sqrt{\sum_i n_i}$ where $\boldsymbol{x}_{i,j} \in \mathbb{R}^{n_j}$ are unit vectors. Specifically, we consider the study of a simple deflation procedure, first introduced by Hotelling in the context of matrix principal component analysis [10] and still used in modern applications such as the recent AlphaTensor model [11], which consists in iterated rank-one approximations followed by subtraction of the estimated rank-one component.

We focus on the case where the spike components are not orthogonal to each other. While the orthogonal case (i.e. $\langle \boldsymbol{x}_{i,k}, \boldsymbol{x}_{j,k} \rangle = 0$ for all $i \neq j$) trivially boils down to the rank-one model studied in [5, 9], in the considered setting, i.e. $\langle \boldsymbol{x}_{i,k}, \boldsymbol{x}_{j,k} \rangle \in (0, 1)$ for $i \neq j$, the behavior of the deflation method is more complex, as depicted in Figure 1 in the simplified case $\boldsymbol{x}_{i,1} = \cdots = \boldsymbol{x}_{i,d} = \boldsymbol{x}_i$ for $i = 1, 2$.

In this work, we characterize this behavior by estimating the alignments $\langle \boldsymbol{x}_{i,k}, \boldsymbol{u}_{j,k} \rangle$ in the asymmetric case, in the high-dimensional regime when $n_i \to \infty$, using the random matrix approach developed in [9]. Furthermore, as a by-product of this analysis, we introduce a method to consistently estimate the underlying SNRs β_i and the alignments $\langle \boldsymbol{x}_{i,k}, \boldsymbol{x}_{j,k} \rangle$ and $\langle \boldsymbol{x}_{i,k}, \boldsymbol{u}_{j,k} \rangle$ from the quantities computed at each step of the deflation procedure.
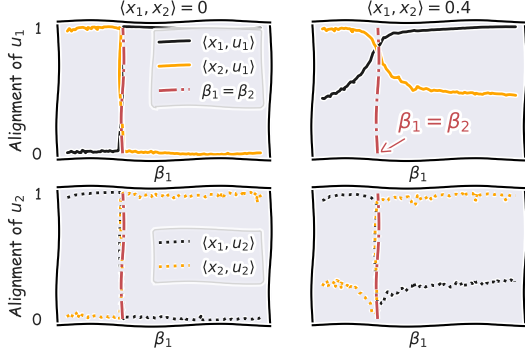
Fig. 1: Result of tensor deflation applied to a noisy rank-two spiked model $\beta_1 x_1^{\otimes 3} + \beta_2 x_2^{\otimes 3} + \mathcal{W}/\sqrt{n}$. The first row corresponds to the first deflation step (yielding u_1) while the second row corresponds to the second deflation step (yielding u_2). The plots depict the alignments $\langle x_i, u_j \rangle$ in terms of β_1 for a fixed β_2. In the orthogonal case $\langle x_1, x_2 \rangle = 0$ (left column), deflation is shown to successfully identify the strongest and second strongest components, i.e. u_1 aligns with x_1 and u_2 aligns with x_2 when $\beta_1 > \beta_2$, while u_1 aligns with x_2 and u_2 aligns with x_1 when $\beta_1 < \beta_2$. For the correlated setting, on the other hand (right column, $\langle x_1, x_2 \rangle = 0.4$), the alignment of (u_1, u_2) with (x_1, x_2) is imperfect in the region where $\beta_1 \approx \beta_2$.

Notations: Scalars are denoted by lowercase letters. Vectors are denote by bold lowercase letters. Tensors are denoted as $\mathcal{A}, \mathcal{B}, \mathcal{C}$. $T_{i_1 \dots i_d}$ denotes the entry $(i_1, \dots, i_d)$ of tensor $\mathcal{T}$. The inner product between two order-d tensors $\mathcal{A}$ and $\mathcal{B}$ is denoted $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1, \dots, i_d} A_{i_1 \dots i_d} B_{i_1 \dots i_d}$. The ℓ_2-norm of $\mathcal{A}$ is $\|\mathcal{A}\| = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$. For any vectors $u_1, \dots, u_d$, contractions of a tensor $\mathcal{A}$ are denoted by $\mathcal{A}(u_1, \dots, u_d) = \sum A_{i_1 \dots i_d} u_{1 i_1} \dots u_{d i_d}$. The notation $\|\mathcal{A}\|_{op} = \sup_{\|u_i\|=1} |\mathcal{A}(u_1, \dots, u_d)|$ stands for the spectral norm. $[n]$ denotes the set $\{1, \dots, n\}$.

2. MODEL AND MAIN RESULTS

We start by describing formally our considered model. Let $r \geq 1$ and $d \geq 3$, we consider the following rank-r order-d spiked tensor model

$$\mathcal{T}_1 = \sum_{i=1}^r \beta_i x_{i,1} \otimes \dots \otimes x_{i,d} + \frac{1}{\sqrt{n}} \mathcal{W} \qquad (1)$$

where $W_{i_1 \dots i_d} \sim \mathcal{N}(0,1)$ i.i.d., $x_{i,j} \in \mathbb{R}^{n_j}$ are unit vectors, $n = \sum_{i=1}^d n_i$ and $\beta_i \geq 0$.

Tensor deflation model: In order to recover the underlying signal components (i.e. the terms $\beta_i x_{i,1} \otimes \dots \otimes x_{i,d}$), Hotelling deflation consists in successive rank-one approximations followed by subtraction of the rank-1 estimate. This can be implemented by computing $\mathcal{T}_2, \mathcal{T}_3, \dots$ sequentially through

$$\mathcal{T}_{i+1} = \mathcal{T}_i - \hat{\lambda}_i \hat{u}_{i,1} \otimes \dots \otimes \hat{u}_{i,d} \quad \text{for} \quad i \in [r] \qquad (2)$$

where $\hat{\lambda}_i \hat{u}_{i,1} \otimes \dots \otimes \hat{u}_{i,d}$ is a critical point of the loss $\|\mathcal{T}_i - \lambda_i u_{i,1} \otimes \dots \otimes u_{i,d}\|_F^2$ which corresponds to the rank-one approximation of $\mathcal{T}_i$ [?]. The critical points satisfy the Karush-Kuhn-Tucker conditions derived from the Lagrangian of the latter objective, i.e.

$$\mathcal{T}_i(\hat{u}_{i,1}, \dots, \hat{u}_{i,j-1}, \cdot, \hat{u}_{i,j+1}, \dots, \hat{u}_{i,d}) = \hat{\lambda}_i \hat{u}_{i,j} \qquad (3)$$

with $\|u_{i,j}\| = 1$, for all $j \in [d]$. In the following, we aim at computing the limits of $\hat{\lambda}_i$ and $\langle x_{i,k}, \hat{u}_{j,k} \rangle$ for $i, j \in [r]$ and $k \in [d]$ when the dimensions n_i grow large. For notational convenience, in the sequel the notation $\lim Q$ stands for the limit of the quantity Q when $n_i \to \infty$.

Sketch of the analytical approach: We follow the approach developed in [9], whereby each tensor $\mathcal{T}_i$ is associated to a structured random matrix $\Phi_d(\mathcal{T}_i, \hat{u}_{i,1}, \dots, \hat{u}_{i,d}) \in \mathbb{R}^{n \times n}$ where the mapping Φ_d is defined in [9, Section 5]. Then, the characterization of the limits of $\hat{\lambda}_i$ and the alignments $\langle x_{i,k}, \hat{u}_{j,k} \rangle$ when $n_i \to \infty$ boils down to the computation of the Stieltjes transform of the limiting spectral measure of $\Phi_d(\mathcal{T}_i, \hat{u}_{i,1}, \dots, \hat{u}_{i,d})$. Hence, we need the following definition and technical assumptions.

Definition 2.1. Let μ be the probability measure with Stieltjes transform $g(z) = \sum_{i=1}^d g_i(z)$ verifying $\Im[g(z)] > 0$ for $\Im[z] > 0$, where $g_i(z)$ satisfies $g_i^2(z) - (g(z) + z) g_i(z) - c_i = 0$, for $z \notin \mathcal{S}(\mu)$ and $\mathcal{S}(\mu)$ stands for the support of μ.

Assumption 2.2. We assume that as $n_i \to \infty$, $r = O(1)$ and denote $c_i = \lim \frac{n_i}{\sum_{j=1}^d n_j}$. We further assume that there exists a sequence of critical points such that $\hat{\lambda}_i \xrightarrow{\text{a.s.}} \lambda_i$, $|\langle x_{i,k}, \hat{u}_{j,k} \rangle| \xrightarrow{\text{a.s.}} \rho_{ijk}$ and $|\langle \hat{u}_{i,k}, \hat{u}_{j,k} \rangle| \xrightarrow{\text{a.s.}} \eta_{ijk}$ such that $\lambda_i \notin \mathcal{S}(\mu)$ and $\rho_{ijk} > 0$.

We therefore have the following result[1] which characterizes the limiting spectral measure of $\Phi_d(\mathcal{T}_i, \hat{u}_{i,1}, \dots, \hat{u}_{i,d})$.

Theorem 2.3. *Under Assumption 2.2, the empirical spectral measure of $\Phi_d(\mathcal{T}_i, \hat{u}_{i,1}, \dots, \hat{u}_{i,d})$ converges to the deterministic measure μ defined in Definition 2.1.*

As shown in [9], in the case $c_i = \frac{1}{d}$ for all $i \in [d]$, the measure μ describes a semi-circle or Wigner-type law of compact support $\mathcal{S}(\mu) = [-2\sqrt{\frac{d-1}{d}}, 2\sqrt{\frac{d-1}{d}}]$, the Stieltjes transform of which writes explicitly as

$$g(z) = \frac{-zd + d\sqrt{z^2 - \frac{4(d-1)}{d}}}{2(d-1)}, \quad z \notin \mathcal{S}(\mu) \qquad (4)$$

---

[1] The proof of Theorem 2.3 follows similar arguments as in [9] and requires some additional arguments for controlling the statistical dependencies between the $\hat{u}_{i,j}$'s and the noise $\mathcal{W}$.
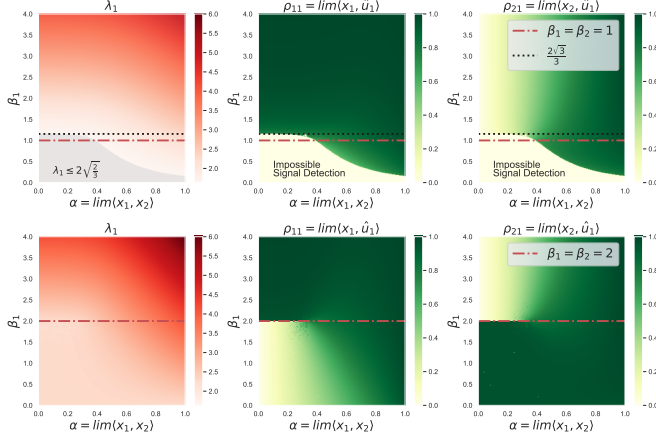
Fig. 2: Phase diagram of the two spikes model in (5). First row corresponds to $\beta_2 = 1$ and second row for $\beta_2 = 2$. The first column depicts λ_1 varying β_1 and α, while the remaining columns depict the asymptotic alignments between the $\boldsymbol{x}_i$'s and $\hat{\boldsymbol{u}}_1$. The figures were obtained by solving the first three equations in (6).

Limiting spectral norms and alignments: We introduce the quantities $\alpha_{ijk} = \lim |\langle \boldsymbol{x}_{i,k}, \boldsymbol{x}_{j,k}\rangle|$, $f(z) = z + g(z)$, $h_i(z) = -\frac{c_i}{g_i(z)}$ that shall be used subsequently. The main result brought by this paper describes the asymptotic singular values and alignments obtained after each tensor deflation step as stated by the following theorem.

Theorem 2.4. *Assume that Assumption 2.2 holds. Then, λ_i, ρ_{ijk} and η_{ijk} satisfy the following system of equations*

$$
\begin{cases}
f(\lambda_j) + \sum_{i=1}^{j-1} \lambda_i \prod_{k=1}^{d} \eta_{ijk} - \sum_{i=1}^{r} \beta_i \prod_{k=1}^{d} \rho_{ijk} = 0\, , \ 1 \leq j \leq r \\[2mm]
h_\ell(\lambda_j)\rho_{kj\ell} + \sum_{i=1}^{j-1} \lambda_i \rho_{ki\ell} \prod_{m\neq\ell}^{d} \eta_{ijm} - \sum_{i=1}^{r} \beta_i \alpha_{ik\ell} \prod_{m\neq\ell}^{d} \rho_{ijm} = 0 \\[1mm]
\qquad 1 \leq \ell \leq d, 1 \leq j, k \leq r \\[2mm]
h_\ell(\lambda_j)\eta_{kj\ell} + g_\ell(\lambda_k) \prod_{m\neq\ell}^{d} \eta_{kjm} + \sum_{i=1}^{j-1} \lambda_i \eta_{ik\ell} \prod_{m\neq\ell}^{d} \eta_{ijm} + \ldots \\[2mm]
\qquad - \sum_{i=1}^{r} \beta_i \rho_{ik\ell} \prod_{m\neq\ell}^{d} \rho_{ijm} = 0,\, 1 \leq \ell \leq d, 1 \leq j < k \leq r
\end{cases}
$$

Sketch of the proof. We use similar arguments as in [9]. We first show that $\mathrm{Var}[\hat{\lambda}_j] = O(n^{-1})$ and use a concentration argument to show that $\hat{\lambda}_j$ concentrates around its expectation. Simlarly, the same property holds for the alignments. Then, we evaluate the expectation of the scalar product between (3) and $\boldsymbol{x}_i$ or $\hat{\boldsymbol{u}}_i$ using Stein's Lemma[2]. $\qquad\square$

---

[2] $\mathbb{E}[Wf(W)] = \mathbb{E}[f'(W)]$ for $W \sim \mathcal{N}(0,1)$.

Particular case of a rank-2, order-3 tensor: For the sake of clarity, let us consider the example of a rank-2 order-3 spiked tensor with $n_1 = n_2 = n_3$, thus

$$
\mathcal{T}_1 = \sum_{i=1}^{2} \beta_i \boldsymbol{x}_{i,1} \otimes \boldsymbol{x}_{i,2} \otimes \boldsymbol{x}_{i,3} + \frac{1}{\sqrt{n}} \mathcal{W} \tag{5}
$$

Furthermore, we assume that for all $i \neq j$ and each $k \in [3]$, $\lim |\langle \boldsymbol{x}_{i,k}, \boldsymbol{x}_{j,k}\rangle| = \alpha \in [0,1]$. In this case, since all the dimensions n_i are equal, the limits of $|\langle \boldsymbol{x}_{i,k}, \hat{\boldsymbol{u}}_{j,k}\rangle|$ and of $|\langle \hat{\boldsymbol{u}}_{1,k}, \hat{\boldsymbol{u}}_{2,k}\rangle|$ are both independent from k by symmetry. Therefore, we drop their dependence on k in our notations. Hence, the system of equations in Theorem 2.4 reduces to seven equations, detailed in the following corollary.

Corollary 2.5. *Denote $\rho_{ij} = \lim |\langle \boldsymbol{x}_{i,k}, \hat{\boldsymbol{u}}_{j,k}\rangle|$ for $i, j \in [2]$ and $\eta = \lim |\langle \hat{\boldsymbol{u}}_{1,k}, \hat{\boldsymbol{u}}_{2,k}\rangle|$ and suppose that Assumption 2.2 holds, then λ_i, ρ_{ij} and η satisfy $\psi(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\rho}) = 0$ with $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \eta)$, $\boldsymbol{\rho} = (\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22})$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \alpha)$ and*

$$
\psi(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\rho}) =
\begin{pmatrix}
f(\lambda_1) - \beta_1 \rho_{11}^3 - \beta_2 \rho_{21}^3 \\
h(\lambda_1)\rho_{11} - \beta_1 \rho_{11}^2 - \beta_2 \alpha \rho_{21}^2 \\
h(\lambda_1)\rho_{21} - \beta_1 \alpha \rho_{11}^2 - \beta_2 \rho_{21}^2 \\
f(\lambda_2) + \lambda_1 \eta^3 - \beta_1 \rho_{12}^3 - \beta_2 \rho_{22}^3 \\
h(\lambda_2)\rho_{12} + \lambda_1 \rho_{11}\eta^2 - \beta_1 \rho_{12}^2 - \beta_2 \alpha \rho_{22}^2 \\
h(\lambda_2)\rho_{22} + \lambda_1 \rho_{21}\eta^2 - \beta_1 \alpha \rho_{12}^2 - \beta_2 \rho_{22}^2 \\
h(\lambda_2)\eta + q(\lambda_1)\eta^2 - \beta_1 \rho_{11}\rho_{12}^2 - \beta_2 \rho_{21}\rho_{22}^2
\end{pmatrix}
\tag{6}
$$

where $h(z) = \frac{-1}{g(z)}$ and $q(z) = z + \frac{g(z)}{3}$ with $g(z)$ given by (4) for $d = 3$ and we recall that $f(z) = z + g(z)$.

Fixing $\boldsymbol{\beta} = (\beta_1, \beta_2, \alpha)$, one can solve $\psi(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\rho}) = 0$ in $(\boldsymbol{\lambda}, \boldsymbol{\rho})$ while ensuring that $0 \leq \eta, \rho_{ij} \leq 1$ and $\lambda_1, \lambda_2 > 2\sqrt{\frac{2}{3}}$. This provides a fixed point equation satisfied by the asymptotic limits of the spectral norms $\hat{\lambda}_i$ and the alignments $\langle \boldsymbol{x}_{i,k}, \hat{\boldsymbol{u}}_{j,k}\rangle$ and $\langle \hat{\boldsymbol{u}}_{1,k}, \hat{\boldsymbol{u}}_{2,k}\rangle$.

Note that the first three equations in (6) only involve λ_1, ρ_{11} and ρ_{21} and are decoupled from the last four equations. Therefore, solving them allows to obtain the *phase diagram* related to the dominant singular mode $(\hat{\lambda}_1, \hat{\boldsymbol{u}}_1)$, depicted in Figure 2. It shows that when β_2 is not large enough (e.g. $\beta_2 = 1$, top row), there exists a region (varying β_1 and α) where it is information-theoretically impossible to detect a signal, while outside this region estimation becomes possible with the MLE (in that case, the estimated $\hat{\boldsymbol{u}}_1$ is shown to be correlated with both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$). For β_2 sufficiently large (e.g. $\beta_2 = 2$, bottom row of Figure 2), signal detection is always possible and the singular vector $\hat{\boldsymbol{u}}_1$ presents a higher alignment with the signal components having the highest SNR β_i (see the two columns on the right).

Moreover, we illustrate in Figure 3 the matching between $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\eta})$, $\hat{\boldsymbol{\rho}} = (\hat{\rho}_{11}, \hat{\rho}_{12}, \hat{\rho}_{21}, \hat{\rho}_{22})$ (where the rank-one approximations are performed using tensor power iteration initialized by tensor SVD [12]) and their asymptotic limits $\boldsymbol{\lambda}, \boldsymbol{\rho}$. Note that, for some β_1, the equation $\psi(\cdot, \boldsymbol{\beta}, \cdot) = 0$ has two distinct solutions. These two solutions correspond to two
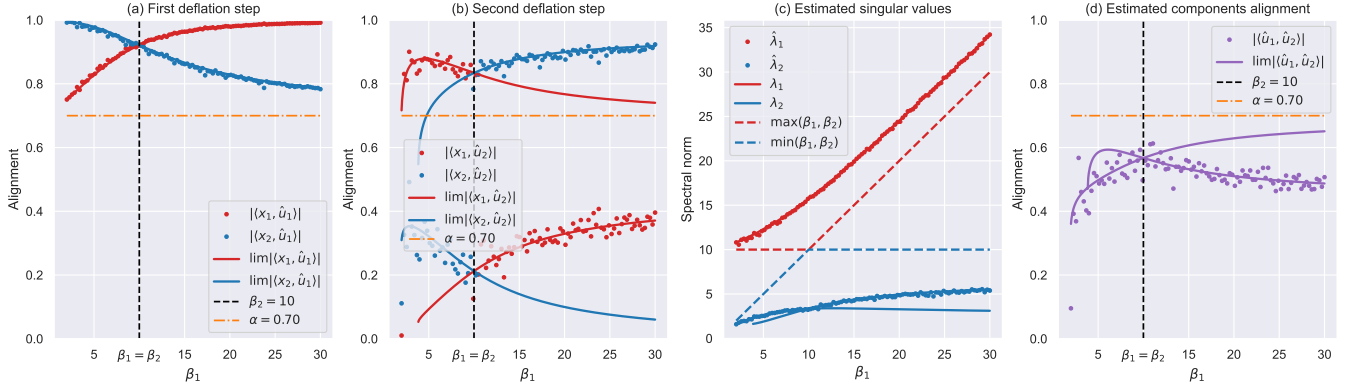
Fig. 3: Illustration of $\langle x_i, \hat{u}_j \rangle$ for $(i,j) \in \{1,2\}$, $\langle \hat{u}_1, \hat{u}_2 \rangle$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ vs. their limits for $\beta_2 = 10$ and $\alpha = \langle x_1, x_2 \rangle = 0.7$ of the two spikes tensor model in (5). (a) shows the alignments between the signal components and the first singular vectors corresponding to the best rank-one approximation of $\mathcal{T}_1$. (b) shows the alignments with the second singular vectors computed after deflation. (c) depicts the singular values. (d) shows the alignments between the singular vectors computed at each step of the deflation procedure. Simulations were performed on a tensor of dimensions $(50, 50, 50)$.

different sequences of critical points. However, in practice, the chosen initialization favor one sequence of critical points as can be seen in Figure 3.

Fig. 4: Estimation of the β_i's by solving the system $\psi(\hat{\boldsymbol{\lambda}}, \cdot, \cdot) = \mathbf{0}$ given $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\eta} = \langle \hat{u}_1, \hat{u}_2 \rangle$ estimated from a two steps deflation on a tensor distributed as in (5). First row corresponds to $n_i = 30$ and second row to $n_i = 100$.

3. CONSISTENT SNR ESTIMATION

Having set the relationship between the β_i's and the limits of the different spectral norms and alignments in our problem, one can exploit this mapping to design a consistent estimator of the underlying SNRs β_i's. Indeed, $\hat{\boldsymbol{\lambda}}$ can directly be estimated from $\hat{u}_i$ obtained with the deflation.

Then we denote $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\rho}}$ the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$ as the vectors satisfying $\psi(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\rho}}) = \mathbf{0}$ with ψ defined in (6). The additional condtions required for the existence and the uniqueness of such solution is not studied in the present paper and shall be considered in an extended version. In our simulations, we find one solution for $\psi(\hat{\boldsymbol{\lambda}}, \cdot, \cdot) = \mathbf{0}$ if $\hat{\lambda}_1, \hat{\lambda}_2 > 2\sqrt{2/3}$. We illustrate the result of such estimation in Figure 4 where we see that $(\hat{\beta}_1, \hat{\beta}_2)$ consistently estimate $(\max(\beta_1, \beta_2), \min(\beta_1, \beta_2))$, while the naive estimator $(\hat{\beta}_1, \hat{\beta}_2) = (\hat{\lambda}_1, \hat{\lambda}_2)$ exhibits a large error in the non-orthogonal case ($\alpha = 0.5$, second column). Moreover, as the dimensions of the tensor increase, Fig. 4 shows (comparing first and second row) that the estimation is consistent. This consistency can be related to a classical concentration phenomenon [13].

4. CONCLUSION

We have provided an analysis of a tensor deflation method in the high-dimensional regime and assuming a low-rank spiked tensor model with correlated signal components. Our analysis allows precise description of the asymptotic behavior of such models and provides consistent estimation of its parameters as shown in the last part of the paper. This paves a new way for analysis of more sophisticated tensor decomposition methods and the understanding of more general tensor models through random tensor theory.
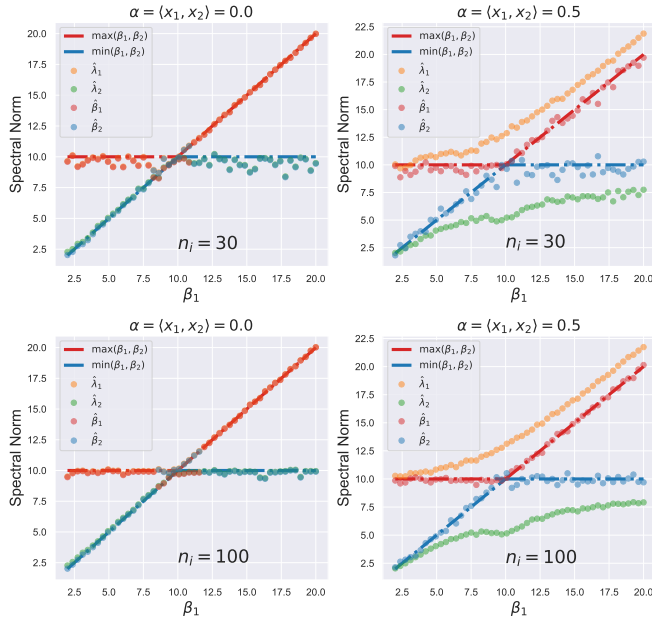
5. REFERENCES

[1] Emile Richard and Andrea Montanari, "A statistical model for tensor PCA," *Advances in neural information processing systems*, vol. 27, 2014.

[2] Amelia Perry, Alexander S Wein, and Afonso S Bandeira, "Statistical limits of spiked tensor models," in *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. Institut Henri Poincaré, 2020, vol. 56, pp. 230–264.

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

[4] Aukosh Jagannath, Patrick Lopatto, and Leo Miolane, "Statistical thresholds for tensor PCA," *The Annals of Applied Probability*, vol. 30, no. 4, pp. 1910–1933, 2020.

[5] Wei-Kuo Chen, Madeline Handschy, and Gilad Lerman, "Phase transition in random tensors with multiple independent spikes," *The Annals of Applied Probability*, vol. 31, no. 4, pp. 1868–1913, 2021.

[6] José Henrique de Morais Goulart, Romain Couillet, and Pierre Comon, "A random matrix perspective on random tensors," *arXiv preprint arXiv:2108.00774*, 2021.

[7] Christopher J Hillar and Lek-Heng Lim, "Most tensor problems are NP-hard," *Journal of the ACM (JACM)*, vol. 60, no. 6, pp. 1–39, 2013.

[8] Gérard Ben Arous, Daniel Zhengyu Huang, and Jiaoyang Huang, "Long random matrices and tensor unfolding," *arXiv preprint arXiv:2110.10210*, 2021.

[9] Mohamed El Amine Seddik, Maxime Guillaud, and Romain Couillet, "When random tensors meet random matrices," *arXiv preprint arXiv:2112.12348*, 2021.

[10] Harold Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, Sept. 1933.

[11] Alhussein Fawzi, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Francisco J R Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, et al., "Discovering faster matrix multiplication algorithms with reinforcement learning," *Nature*, vol. 610, no. 7930, pp. 47–53, 2022.

[12] Arnab Auddy and Ming Yuan, "On estimating rank-one spiked tensors in the presence of heavy tailed errors," *IEEE Transactions on Information Theory*, 2022.

[13] Florent Benaych-Georges, Alice Guionnet, and Mylène Maida, "Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices," *Electronic Journal of Probability*, vol. 16, pp. 1621–1662, 2011.