Selective Multiple Power Iteration for Tensor PCA

Mohamed Ouerfelli

CEA List
IJCLab, Paris-Saclay University

Under the supervision of
Vincent Rivasseau & Mohamed Tamaazousti

Tensor Club, October 13, 2021

Overview

Tensor PCA problem

- Tensor PCA is a statistical model introduced by [Richard and Montanari, 2014], it consists in inferring an unknown unit vector v_0 from a tensor $\mathbf{T}$ given by:

$$\mathbf{T} = \sqrt{n}\beta(\mathbf{v}_0)^{\otimes k} + \mathbf{Z}$$

with Z Gaussian noise tensor such that $Z_{i_1 \ldots i_k} \sim \mathcal{N}(0, 1)$ and β the signal-to-noise ratio.

- $\mathcal{H}(\mathbf{v}) = \langle \mathbf{T}, \mathbf{v}^{\otimes k} \rangle$ will be referred to as the landscape.

Tensor PCA motivations

Tensor PCA model has been extensively studied in the last years due to three main important motivations:

1. Algorithms for Tensor PCA may be adaptable for Tensor decomposition which has multiple important applications.

2. It is a simple model that allows the study of high dimensional non convex landscapes that arise in multiple fields as well as the gradient descent dynamics in such landscapes.

3. Tensor PCA may exhibit a statistical-algorithmic gap that are common in multiple other statistical inference models.

Tensor decomposition

- Tensor decomposition (generalization of matrix SVD) is motivated by the increasing number of problems in which it is crucial to exploit the tensorial structure ([Sidiropoulos et al., 2017]).
- Two main tensor decomposition models are commonly used: Tucker decomposition and CP decomposition.
- CP decomposition consists in approximating an initial tensor **T** by a sum of rank-one tensors:

$$\hat{\mathbf{T}} = \sum_i \mathbf{v}_i \otimes \mathbf{w}_i \otimes \mathbf{z}_i$$

- Example of papers that proposed algorithms that could be used as frames or basic bricks for more general algorithms:
Homotopy [Anandkumar et al., 2017], Averaged gradient descent [Biroli et al., 2020], Iterated gradient etc.

Applications in machine learning

Recently it was successfully used to address important problems in machine learning:

- Unsupervised learning (learning latent variable models, in particular latent Dirichlet allocation [Anandkumar et al., 2014], [Anandkumar et al., 2015])

- Supervised learning (training of two-layer neural networks [Janzamin et al., 2015])

- Reinforcement learning ([Azizzadenesheli et al., 2016])

Other applications
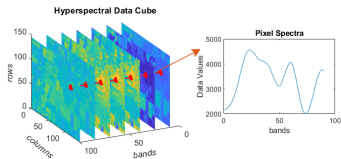
(a) Hyper-spectral images [Lin and Bourennane, 2013]
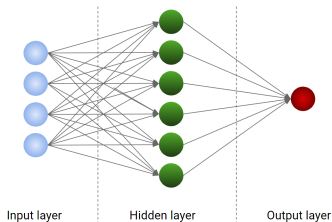
(b) Neural network compression [Astrid and Lee, 2017]

(c) Telecommunication [Decurninge et al., 2020]

(d) Medical diagnosis [Bharath et al., 2015]

High dimensional non convex landscapes

- It has been showed that the landscape of Tensor PCA has interesting properties [Arous et al., 2019, Ros et al., 2019, Auffinger et al., 2021]

- It is a typical case of a high dimensional landscape highly non-convex and that has an exponentially growing number of critical points and local minima.

- These landscapes occur in diverse areas: disordered systems, biology, neural networks, inference, string theory and cosmology.

- A line of research uses Tensor PCA to understand these landscapes [Arous et al., 2019, Mannelli et al., 2019a, Mannelli et al., 2019b, Mannelli et al., 2020, Ros et al., 2019, Ros, 2020]

Statistical threshold: MLE

- Maximum Likehood Estimator (MLE):

$$\hat{\boldsymbol{v}} = \arg\max_{\|\boldsymbol{v}\|=1} \langle \boldsymbol{\mathsf{T}}, \boldsymbol{v} \otimes \boldsymbol{v} \otimes \boldsymbol{v} \rangle$$

- The maximum likelihood estimator achieves the maximal correlation with the planted vector among measurable estimators above the estimation threshold. [Jagannath et al., 2020]
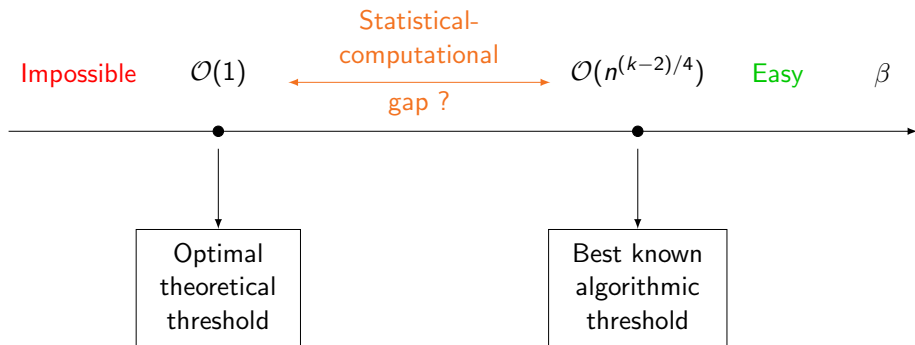
Impossible $\mathcal{O}(1)$ Theoretically possible β

- In [Evnin, 2020], trace invariants were recently used to study the highest eigenvalue of a real symmetric Gaussian tensor.

- Subsequently, [Gurau, 2020] provided a theoretical study on a function based on an infinite sum of these invariants. Their results suggest a transition phase for the highest eigenvalue of a tensor for β around the theoretical threshold $\mathcal{O}(1)$.

- However this method requires computing an integral over a n-dimensional space, which may not be possible in a polynomial time.

Existent algorithms and computational threshold

- Sum-of-squares: Semi-definite programming method, the first with a $\mathcal{O}(n^{(k-2)/4})$ theoretical guarantees.

- Tensor unfolding consists in reorganizing the tensor elements on a matrix and then using matrix PCA tools. It is equivalent to SoS.

- Homotopy-based method: consists in applying tensor power iteration $\boldsymbol{v} \leftarrow \mathbf{T}.\boldsymbol{v}^{\otimes k-1}$ with an initialization $\boldsymbol{v} = T_{ijj}$.

- Other studied methods have been inspired by different perspectives like statistical physics [Arous et al., 2020, Wein et al., 2019, Biroli et al., 2020], quantum computing [Hastings, 2020] as well as statistical query [Dudeja and Hsu, 2021].

Statistical-algorithmic gap

Impossible $\mathcal{O}(1)$ Statistical-computational gap ? $\mathcal{O}(n^{(k-2)/4})$ Easy β

Optimal theoretical threshold

Best known algorithmic threshold

Statistical-computational gap investigations

- There often exist conjectured intrinsic statistical-computational gaps in many problems, as observed in tensor completion (Barak and Moitra, 2016), high-order clustering (Luo and Zhang, 2020), but also planted clique, sparse PCA, community detection, etc.

- The analysis of statistical-computational gaps has attracted a lot of attention because of its crucial role in the understanding of the computational feasibility of a wide range of inference and tensor problems.

- Two main approaches:
  - Average case reduction [Luo and Zhang, 2020]: Evidence for the computational hardness developed by establishing the equivalence of the computational hardness commonly raised conjectures.
  - Analysis of restricted algorithmic models

Algorithmic limits

- The analysis of Sum-of-squares framework. This led to a new conjecture on low degree polynomial methods [Hopkins, 2018, Kunisky et al., 2019].

- Belief propagation, approximate message passing [Lesieur et al., 2017]

- Analysis of optimization landscape

Gradient-based algorithms

In particular for local algorithms like gradient descent (on which SMPI is based). Two main explanations are given for the failure of gradient-based methods in low SNR:

- The number of minima is exponentially large, thus the algorithm will get stuck in one of them. [Arous et al., 2019]

- Regardless of if it will get stuck or not, the signal is too weak anyway in the equator [Arous et al., 2020].

Simple Gradient descent

- In the power iteration, let's denote the part associated to the noise $\boldsymbol{g}_N$ and the one associated to the signal $\boldsymbol{g}_S$.

$$\mathbf{T}\boldsymbol{vv} = \mathbf{Z}\boldsymbol{vv} + \beta\langle\boldsymbol{v}, \boldsymbol{v}_0\rangle^2\boldsymbol{v}_0$$
$$\equiv \boldsymbol{g}_N \quad +\boldsymbol{g}_S$$

- A naive approach is to consider that $\boldsymbol{g}_N$ is at each step a random vector. And see in which case we can increase the correlation with $\boldsymbol{v}_0$ at each step.
- If $\beta \gg n^{(k-2)/2}$, gradient descent will always be successful.
- If $\beta \ll n^{(k-2)/2}$, gradient descent with a random initialization will fail with an exponential probability (which is the probability to have a non trivial initial correlation with $\boldsymbol{v}_0$).
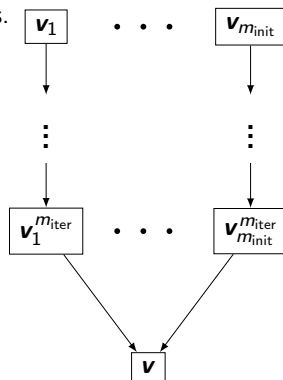
$$\text{Tensor } \mathbf{T} = \mathbf{Z} + \beta \mathbf{v}_0^{\otimes k}$$

First step: Generate m_{init} random vectors.

$$\boxed{\mathbf{v}_1} \quad \bullet \ \bullet \ \bullet \quad \boxed{\mathbf{v}_{m_{\text{init}}}}$$

Second step: Iterate m_{iter} times :

$$\mathbf{v}_i^{j+1} = \frac{\mathbf{T}\mathbf{v}_i^j \mathbf{v}_i^j}{\left\| \mathbf{T}\mathbf{v}_i^j \mathbf{v}_i^j \right\|}$$

$$\vdots \qquad\qquad \vdots$$

$$\boxed{\mathbf{v}_1^{m_{\text{iter}}}} \quad \bullet \ \bullet \ \bullet \quad \boxed{\mathbf{v}_{m_{\text{init}}}^{m_{\text{iter}}}}$$

Third step: Choose the vector $\mathbf{v} = \arg\max_{1 \leq i \leq m_{\text{init}}} \left(\mathbf{T}.\mathbf{v}_i^{m_{\text{iter}}}.\mathbf{v}_i^{m_{\text{iter}}}.\mathbf{v}_i^{m_{\text{iter}}} \right)$:

$$\boxed{\mathbf{v}}$$

The main features of SMPI

Simple algorithm that began by some surprising observations, followed by refinements that led to these four main features: Just take them as they are for now.

1. Large step size

2. Larger stopping condition

3. Polynomial number of iterations

4. Polynomial number of initialisations.

Power iteration and gradient descent

- The gradient on a sphere is given by
  $\nabla f(v) - (\nabla f(v).v)v = Tvv - (Tvvv)v$ (more mathematical details in [Ros et al., 2019])

- It will be in our case equal to $g = Tvv - (Tvvv)v$.

- The power iteration could be considered as a gradient descent with a step size equal to $1/(Tvvv)$ indeed
  $Tvv = Tvv - (Tvvv)v + (Tvvv)v = g + (Tvvv)v = (Tvvv)(v + \frac{g}{Tvvv})$.
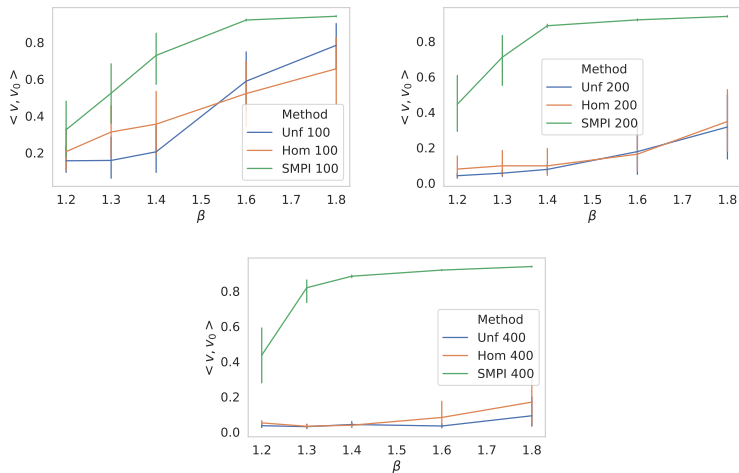
Figure: Comparison of the results of SMPI with the unfolding method (Unf) and Homotopy-based method (Hom)

- Why does it give different results that one should expect from the naive analysis we did in the beginning?

- Why do we require these four conditions?

- What about the previous explanations on the failure of local methods?

- Why does it give different results that one should expect from the naive analysis we did in the beginning?

- Why do we require these four conditions?

- What about the previous explanations on the failure of local methods?

$$\mathbf{T}\boldsymbol{v}\boldsymbol{v} = \mathbf{Z}\boldsymbol{v}\boldsymbol{v} \ + \ \beta\langle\boldsymbol{v}, \boldsymbol{v}_0\rangle^2 \boldsymbol{v}_0$$

$$\equiv \boldsymbol{g}_N \quad + \boldsymbol{g}_S$$

Figure: In blue, the total correlation of $\boldsymbol{v}_i$ with $\boldsymbol{v}_0$ at each iteration i. In orange in the left, the correlation between the normalized $\boldsymbol{g}_N$ and $\boldsymbol{v}_0$. In orange in the right, the ratio between contribution of the noise gradient to $\boldsymbol{v}_0$ and the signal gradient.

The plateau

$$\langle \frac{\mathbf{Z}\boldsymbol{vv}}{\|\mathbf{Z}\boldsymbol{vv}\|}, v_0 \rangle = \frac{\langle v, v_0 \rangle (\|\mathbf{T}\boldsymbol{vv}\| - \beta\langle \boldsymbol{v}, \boldsymbol{v}_0 \rangle)}{\sqrt{\|\mathbf{T}\boldsymbol{vv}\|^2 + \beta^2\langle \boldsymbol{v}, \boldsymbol{v}_0 \rangle^4 - 2\beta\langle \boldsymbol{v}, \boldsymbol{v}_0 \rangle^3\|\mathbf{T}\boldsymbol{vv}\|}}$$

For $\beta = 1.44\sqrt{n}$, the theoretical value is $\langle \frac{Z\boldsymbol{vv}}{\|\mathbf{Z}\boldsymbol{vv}\|}, v_0 \rangle_{\mathrm{th}} = 0.496$. The table 1 give the average and the standard deviation obtained experimentally for different n

Table: Experimental plateau

n	50	100	150	200	400
$\langle \frac{Z\boldsymbol{vv}}{\|\mathbf{Z}\boldsymbol{vv}\|}, v_0 \rangle$	0.469 $\pm$0.148	0.518 $\pm$0.074	0.507 $\pm$0.056	0.487 $\pm$0.038	0.511 $\pm$0.025

Noise gradient increases its correlation with v_0

Proposition

Let $\mathbf{Z}$ be a random tensor whose components follow $\mathcal{N}(0,1)$. Let v_0 a fixed vector. Let ρ a random unitary vector and $\varepsilon > 0$.

Then if we denote $w_0 = \dfrac{\rho + \varepsilon v_0}{\sqrt{1 + \varepsilon^2}}$ and $w_1 = \dfrac{\mathbf{Z} w_0 w_0}{\|\mathbf{Z} w_0 w_0\|}$, we have:

$$E(\langle \frac{\mathbf{Z} w_1 w_1}{\|\mathbf{Z} w_1 w_1\|}, v_0 \rangle^2) > E(\langle \frac{\mathbf{Z} w_0 w_0}{\|\mathbf{Z} w_0 w_0\|}, v_0 \rangle^2) \tag{1}$$

Simple illustration

$$\boldsymbol{v} \leftarrow \mathbf{T}\boldsymbol{v}\boldsymbol{v} \quad = \quad (\mathbf{Z}\boldsymbol{v}_0\boldsymbol{v}\boldsymbol{v})\boldsymbol{v}_0 \quad + \quad \sum_{i=1}^{n-1}(\mathbf{Z}\boldsymbol{e}_i\boldsymbol{v}\boldsymbol{v})\boldsymbol{e}_i \quad + \quad \beta\boldsymbol{v}_0(\boldsymbol{v}.\boldsymbol{v}_0)^2$$

$$\mathbf{T}\boldsymbol{v}\boldsymbol{v} \quad = \quad \underbrace{(\mathbf{Z}\boldsymbol{v}_0\boldsymbol{v}\boldsymbol{v})\boldsymbol{v}_0}_{\substack{\text{Noise contribution}\\\text{to the convergence}\\\text{towards } \boldsymbol{v}_0}} \quad + \quad \sum_{i=1}^{n-1}(\mathbf{Z}\boldsymbol{e}_i\boldsymbol{v}\boldsymbol{v})\boldsymbol{e}_i \quad + \quad \underbrace{\beta\boldsymbol{v}_0(\boldsymbol{v}.\boldsymbol{v}_0)^2}_{\substack{\text{Signal contribution}\\\text{to the convergence}\\\text{towards } \boldsymbol{v}_0}}$$

Questions on the algorithm

- Why does it give different results that one should expect from the naive analysis we did in the beginning? ⇒ The mechanism of convergence takes advantage of the noise

- Why do we require these four conditions?

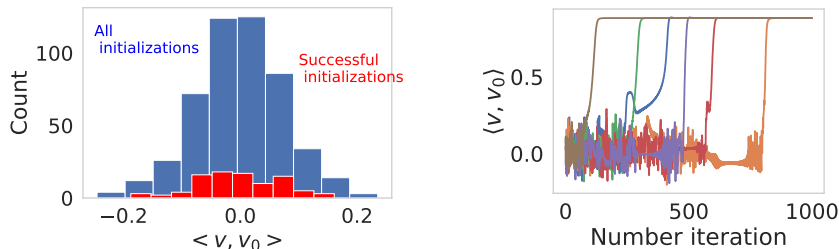- What about the previous explanations on the failure of local methods?

Figure: (a) : in blue the distribution of the correlation between the signal and all initial vectors, and in red the correlations of the initializations that succeeded. $n = 200$, $m_{iter} = 1000$ and $\beta = 1.2$ (b) each curve represents the trajectory of an initalization

Diverging out of a minimum

- Let's consider a minimum of the landscape that we denote $\mathbf{v}_0$. Let $\mathbf{v}_1$ be one of the eigenvector of the Hessian matrix at $\mathbf{v}_0$ and λ_1 its associated eigenvalue.

-

$$
\begin{aligned}
\mathbf{T}.(\mathbf{v}_0 + \varepsilon \mathbf{v}_1)^{\otimes 2} &= \mathbf{T}.(\mathbf{v}_0)^{\otimes 2} + 2\varepsilon \mathbf{T}.(\mathbf{v}_0 \otimes \mathbf{v}_1) + O(\epsilon^2) \\
&= f(\mathbf{v}_0)\mathbf{v}_0 + \varepsilon\, 2\lambda_1 \mathbf{v}_1 + O(\epsilon^2)
\end{aligned}
\tag{2}
$$

- Hence, if $2|\lambda_1| > f(\mathbf{v}_0)$, $\langle \mathbf{y}_1, \mathbf{v}_0 \rangle < \langle \mathbf{y}_0, \mathbf{v}_0 \rangle$ and $\langle \mathbf{y}_1, \mathbf{v}_1 \rangle > \langle \mathbf{y}_0, \mathbf{v}_0 \rangle$. Which means that if any of the eigenvalues is smaller than $-2f(\mathbf{v}_0)$, then the minimum is unstable under power iteration.
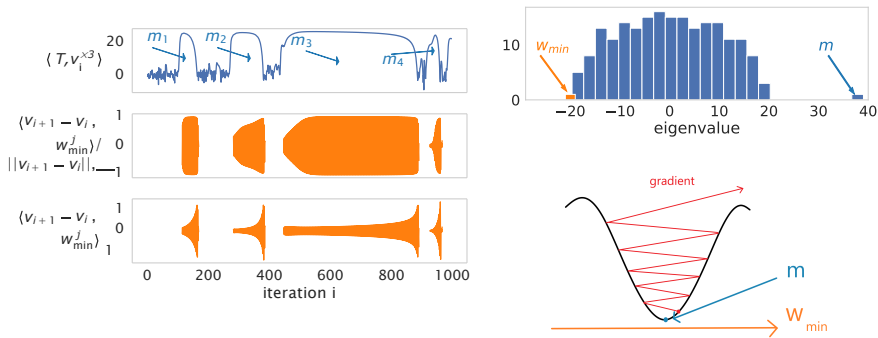
Figure: Escaping the local minima m_j by oscillating around it along the axis parallel to the vector w_{min}^j
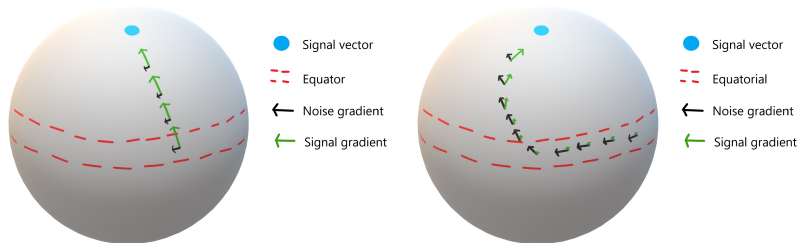
Schema

Figure: Figure illustrating the mechanism of convergence of power iteration in low (left) and large (right) Signal-to-Noise Ratio

It is thus very interesting to take advantage of the numerical analysis we performed on SMPI to understand how it would be able to bypass this these possible explanations to the failure of local iterative algorithms.

- The algorithm runs through many spurious minima but is still able to escape them thanks to its large step size.
- It is the gradient associated to the noise Zvv that not only triggers the convergence but also carries it. It is consistent with the fact that the signal is indeed too small for its associated gradient to converge towards the signal by itself.

Practical applications: Tensor decomposition

- Hyperspectral images consist in hundreds of narrow contiguous bands over a wide range of the electromagnetic spectrum.
- Hyperspectral sensors have applications in astronomy, agriculture, molecular biology, biomedical imaging, geosciences, physics, and surveillance.
- Denoising is an important preprocessing step to analyze a hyperspectral image (HSI).

Figure: Caption

Comparison of SMPI with ALS

- In [Liu et al., 2012], the authors showed that the CP decomposition model using the ALS algorithm performs better than other considered methods as a denoising procedure.
- In order to judge the performance of our algorithm, we perform the same experiment and compare it with the ALS algorithm using the Python TensorLy package [Kossaifi et al., 2016]. The hyperspectral image we use is the open source data given in [Miller et al., 2018] that we normalize.

Table: Comparison of methods

n	20	50	100	200
ALS (TensorLy)	43.58	54.24	60.58	66.53
SMPI	**44.24**	**54.53**	**60.93**	**66.91**

Non convex landscapes

Many observations that may bring new insights:

- The importance of the step size.

- There may be more than just counting the minima and the critical points.

- In inference problems, one has to study more carefully the behavior of the gradient.
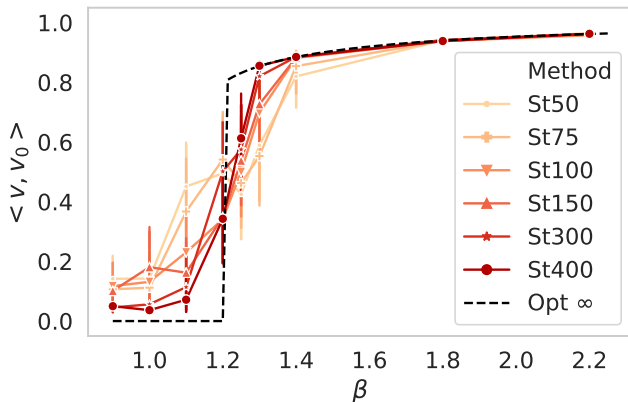
Statistical algorithmic gap

Figure: Asymptotic behavior of SMPI method illustrated by different results on various values of n, ranging from 50 to 400. The dashed line (Opt ∞) corresponds to the predicted optimal theoretical result assuming $n = \infty$ ([Jagannath et al., 2020]).

Scaling

We will make the assumption that the threshold can approximately be written as $\beta = cn^\alpha$ and our aim is to recover α empirically. Given n_1 and n_2, two values of n:

$$\alpha_{\text{emp}} = \frac{\log\left(\beta_{n_2}^{\text{emp}}\right) - \log\left(\beta_{n_1}^{\text{emp}}\right)}{\log(n_2) - \log(n_1)} \tag{3}$$

Table: Experimental scaling for each method

n_2	150	200	400	800
Unfolding	0.23	0.248	0.26	0.2516
SMPI	-0.063	-0.052	-0.053	-0.036

The two methods have an approximately constant α, which confirms that the threshold behaves like $\beta = cn^\alpha$ in this range of n.

Finite size effects?

- It is important to first consider the possibility that the experimental results that we obtained could be due to finite size effects.

- The main aim and motivation of this work is not closing the gap but rather to provide novel theoretical and experimental insights that will help us understand better this conjectured gap either to prove it rigorously or to rule it out.

Comparison with a matrix case

A recent paper [Bandeira et al., 2020] proved that an observation made by [Montanari and Richard, 2015] that an SDP can tightly optimize a spiked matrix Hamiltonian was actually due to finite size effect and the SDP is not tight. This emerges at around dimension $n \sim 10^4$

- Constant factors are not relevant and the difference is fundamentally different and differ by a factor of $n^{1/2}$
- We observe that as n grows, the average correlation obtained by SMPI actually improves
- We can't compare directly the dimension n of a tensor **T** with the one of a matrix **M**.

Conclusion

- The results obtained and the new insights opens the way to explore further questions.

- Possible approaches to imporve our understanding:
  - Statistical physics (Spin glass)
  - Renormalization
  - Advanced Probability tools
  - Trace invariants

Anandkumar, A., Deng, Y., Ge, R., and Mobahi, H. (2017).
Homotopy analysis for tensor pca.
In *Conference on Learning Theory*, pages 79–104. PMLR.

Anandkumar, A., Foster, D. P., Hsu, D., Kakade, S. M., and Liu, Y.-K. (2015).
A spectral algorithm for latent dirichlet allocation.
Algorithmica, 72(1):193–214.

Anandkumar, A., Ge, R., Hsu, D., Kakade, S. M., and Telgarsky, M. (2014).
Tensor decompositions for learning latent variable models.
Journal of Machine Learning Research, 15:2773–2832.

Arous, G. B., Gheissari, R., Jagannath, A., et al. (2020).
Algorithmic thresholds for tensor pca.
Annals of Probability, 48(4):2052–2087.

Arous, G. B., Mei, S., Montanari, A., and Nica, M. (2019).
The landscape of the spiked tensor model.

Communications on Pure and Applied Mathematics,
72(11):2282–2330.

📄 Astrid, M. and Lee, S.-I. (2017).
Cp-decomposition with tensor power method for convolutional neural
networks compression.
In *2017 IEEE International Conference on Big Data and Smart
Computing (BigComp)*, pages 115–118. IEEE.

📄 Auffinger, A., Arous, G. B., and Li, Z. (2021).
Sharp complexity asymptotics and topological trivialization for the (p,
k) spiked tensor model.
arXiv preprint arXiv:2106.09768.

📄 Azizzadenesheli, K., Lazaric, A., and Anandkumar, A. (2016).
Reinforcement learning of pomdps using spectral methods.
In *29th Annual Conference on Learning Theory*, volume 49 of
Proceedings of Machine Learning Research, publisher = PMLR, pages
193–256.

📄 Bandeira, A. S., Kunisky, D., and Wein, A. S. (2020).

Average-case integrality gap for non-negative principal component analysis.
arXiv preprint arXiv:2012.02243.

📄 Bharath, H. N., Sima, D. M., Sauwen, N., Himmelreich, U., De Lathauwer, L., and Van Huffel, S. (2015).
Tensor based tumor tissue type differentiation using magnetic resonance spectroscopic imaging.
In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7003–7006. IEEE.

📄 Biroli, G., Cammarota, C., and Ricci-Tersenghi, F. (2020).
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.

📄 Decurninge, A., Land, I., and Guillaud, M. (2020).
Tensor-based modulation for unsourced massive random access.
IEEE Wireless Communications Letters, 10(3):552–556.

📄 Dudeja, R. and Hsu, D. (2021).

Statistical query lower bounds for tensor pca.
Journal of Machine Learning Research, 22(83):1–51.

📄 Evnin, O. (2020).
Melonic dominance and the largest eigenvalue of a large random tensor.
arXiv preprint arXiv:2003.11220.

📄 Gurau, R. (2020).
On the generalization of the wigner semicircle law to real symmetric tensors.
arXiv preprint arXiv:2004.02660.

📄 Hastings, M. B. (2020).
Classical and Quantum Algorithms for Tensor Principal Component Analysis.
Quantum, 4:237.

📄 Hopkins, S. (2018).
Statistical inference and the sum of squares method.
PhD thesis, Cornell University.

📄 Jagannath, A., Lopatto, P., and Miolane, L. (2020).
Statistical thresholds for tensor pca.
The Annals of Applied Probability, 30(4):1910–1933.

📄 Janzamin, M., Sedghi, H., and Anandkumar, A. (2015).
Beating the perils of non-convexity: Guaranteed training of neural
networks using tensor methods.
arXiv preprint arXiv:1506.08473.

📄 Kossaifi, J., Panagakis, Y., Anandkumar, A., and Pantic, M. (2016).
Tensorly: Tensor learning in python.
arXiv preprint arXiv:1610.09555.

📄 Kunisky, D., Wein, A. S., and Bandeira, A. S. (2019).
Notes on computational hardness of hypothesis testing: Predictions
using the low-degree likelihood ratio.
arXiv preprint arXiv:1907.11636.

📄 Lesieur, T., Miolane, L., Lelarge, M., Krzakala, F., and Zdeborová, L.
(2017).

Statistical and computational phase transitions in spiked tensor estimation.
In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 511–515. IEEE.

📄 Lin, T. and Bourennane, S. (2013).
Survey of hyperspectral image denoising methods based on tensor decompositions.
EURASIP journal on Advances in Signal Processing, 2013(1):1–11.

📄 Liu, X., Bourennane, S., and Fossati, C. (2012).
Denoising of hyperspectral images using the parafac model and statistical performance analysis.
IEEE Transactions on Geoscience and Remote Sensing, 50(10):3717–3724.

📄 Luo, Y. and Zhang, A. R. (2020).
Open problem: Average-case hardness of hypergraphic planted clique detection.
In *Conference on Learning Theory*, pages 3852–3856. PMLR.

📄 Mannelli, S. S., Biroli, G., Cammarota, C., Krzakala, F., Urbani, P., and Zdeborová, L. (2020).
Marvels and pitfalls of the langevin algorithm in noisy high-dimensional inference.
Physical Review X, 10(1):011057.

📄 Mannelli, S. S., Biroli, G., Cammarota, C., Krzakala, F., and Zdeborová, L. (2019a).
Who is afraid of big bad minima? analysis of gradient-flow in a spiked matrix-tensor model.
arXiv preprint arXiv:1907.08226.

📄 Mannelli, S. S., Krzakala, F., Urbani, P., and Zdeborova, L. (2019b).
Passed & spurious: Descent algorithms and local minima in spiked matrix-tensor models.
In *international conference on machine learning*, pages 4333–4342. PMLR.

Miller, C., Green, R., Thompson, D., Thorpe, A., Eastwood, M., Mccubbin, I., Olson-Duvall, W., Bernas, M., Sarture, C., Nolte, S., et al. (2018).
Above: Hyperspectral imagery from aviris-ng, alaskan and canadian arctic, 2017–2018.
ORNL DAAC.

Montanari, A. and Richard, E. (2015).
Non-negative principal component analysis: Message passing algorithms and sharp asymptotics.
IEEE Transactions on Information Theory, 62(3):1458–1484.

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

Ros, V. (2020).
Distribution of rare saddles in the p-spin energy landscape.
Journal of Physics A: Mathematical and Theoretical, 53(12):125002.

📄 Ros, V., Arous, G. B., Biroli, G., and Cammarota, C. (2019).
Complex energy landscapes in spiked-tensor and simple glassy models:
Ruggedness, arrangements of local minima, and phase transitions.
Physical Review X, 9(1):011003.

📄 Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K.,
Papalexakis, E. E., and Faloutsos, C. (2017).
Tensor decomposition for signal processing and machine learning.
IEEE Transactions on Signal Processing, 65(13):3551–3582.

📄 Wein, A. S., El Alaoui, A., and Moore, C. (2019).
The kikuchi hierarchy and tensor pca.
In *2019 IEEE 60th Annual Symposium on Foundations of Computer
Science (FOCS)*, pages 1446–1468. IEEE.