

# Algorithmic Strategies for Tensorial Principal Components

Mohamed Ouerfelli

CEA List  
IJCLab, Paris-Saclay University

In collaboration with  
Vincent Rivasseau & Mohamed Tamaazousti

Random Geometry in Heidelberg, May 16th, 2022

- 1 Introduction
- 2 First strategy : Random Tensor Theory
- 3 Statistical-computational gap
- 4 Second strategy : SMPI
- 5 Conclusion

- 1 Introduction
- 2 First strategy : Random Tensor Theory
- 3 Statistical-computational gap
- 4 Second strategy : SMPI
- 5 Conclusion

# Introduction

Given a data matrix  $X$ , Principal Component Analysis (PCA) can be regarded as a 'denoising' technique that replaces  $X$  by its closest rank-one approximation.

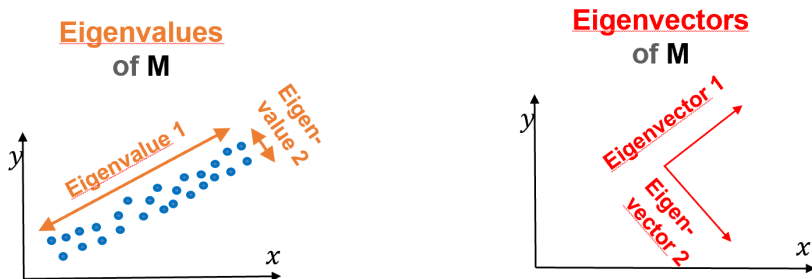
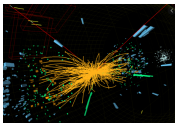
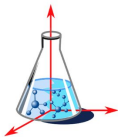


Figure: Matrix PCA

## Existent applications of matrix PCA



High energy physics [Huang et al., 2020]



Chemochemistry [Wold et al., 1987]



Quantitative finance [Pasini, 2017]



Geology [Joreskog et al., 1976]



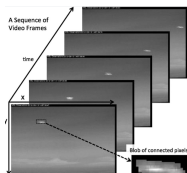
Neuroscience [Cunningham and Byron, 2014]



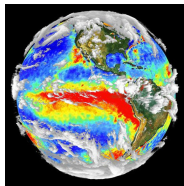
Computer vision [De la Torre and Black, 2001]

# Matrix PCA limit: Tensors

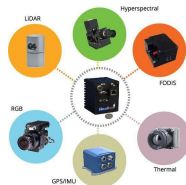
- Powerful computers and acquisition devices have made it possible to capture and store real-world multidimensional data.
- Thus, the generalization of PCA to tensors is motivated by problems in which it is important to exploit higher order moments, where data elements are naturally given more than two indices.



Video data



Weather data



Data from multi-sensors

- Tensor PCA is a statistical model introduced by [Richard and Montanari, 2014], it consists in inferring an unknown unit vector  $\mathbf{v}_0$  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  $Z_{i_1\dots i_k} \sim \mathcal{N}(0, 1)$  and  $\beta$  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.



- 1 Introduction
- 2 First strategy : Random Tensor Theory**
- 3 Statistical-computational gap
- 4 Second strategy : SMPI
- 5 Conclusion

# Eigenvalues and eigenvectors

- An important concept in problems involving matrices is the spectral theory.
- Recovering the complete general information requires computing the eigenvalues and their respective eigenvectors.
- Examples:

|                   | Eigenvalues             | Eigenvectors                       |
|-------------------|-------------------------|------------------------------------|
| Matrix PCA        | Data variability        | Direction of the variability       |
| Signal processing | Intensity of the signal | Direction of the signal associated |
| Quantum physics   | Energy levels           | States associated                  |

# Eigenvalues and eigenvectors

- An important property of the eigenvalues of a  $n$ -dimensional matrix  $\mathbf{M}$  is its invariance under orthogonal transformations.

$$\{\mathbf{M} \rightarrow \mathbf{O}\mathbf{M}\mathbf{O}^{-1}, \mathbf{O} \in \mathcal{O}(n)\}$$

- Indeed, since these transformations essentially just rotate the basis used to define the coordinate system, they must not affect intrinsic information like data variability.
- There are more such invariants than eigenvalues. For example the trace of the powers of  $\mathbf{M}$ :  $\text{Tr}(\mathbf{M}^{2k}), k \in \mathbb{N}$

- However, the concept of eigenvalue and eigenvector is ill defined in the tensor case and not practical: the number of eigenvalues is exponential with the dimension  $n$  ! [Qi, 2005]
- In contrast, we have a very convenient generalization of the traces of the power matrices for the tensors that we call trace invariants.
- They have been extensively studied during the last years in the context of high energy physics and many important properties have been proven ( [Gurau, 2017]).

# Graph invariants

- A tensor  $T$  transforms under the group  $\bigotimes_{a=1}^d O(N)$  as:

$$\mathbf{T}_{a_j^1 \dots a_j^D} \rightarrow O_{a_j^1 b_j^1}^{(1)} \dots O_{a_j^D b_j^D}^{(D)} \mathbf{T}_{b_j^1 \dots b_j^D} \quad \text{for } O^{(i)} \in O(N) \quad \forall i \in [D] \quad (1)$$

- We can build scalars by contracting the indices of  $2k$  copies of the tensor  $T$ .
- The number of these invariants is investigated in [Avohou et al., 2020]. They have a very simple illustration as graphs.

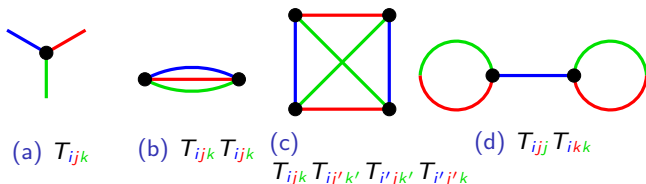


Figure: Example of graphs and their associated invariants

# Matrices associated to a graph

- An invariant should be able to detect a signal. But if our goal is to recover it, we should find mathematical objects that are able to provide a vector.
- To this effect, we introduce a new set of tools in the form of matrices. We denote by  $\mathbf{M}_{\mathcal{G},e}$  the matrix obtained by cutting an edge  $e$  of a graph  $\mathcal{G}$  in two half edges.

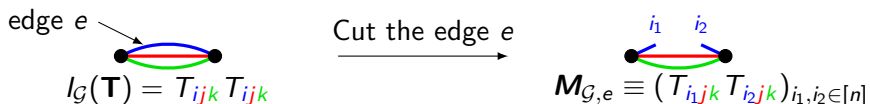


Figure: Obtaining a matrix by cutting the edge of a trace invariant graph  $\mathcal{G}$

# The matrix

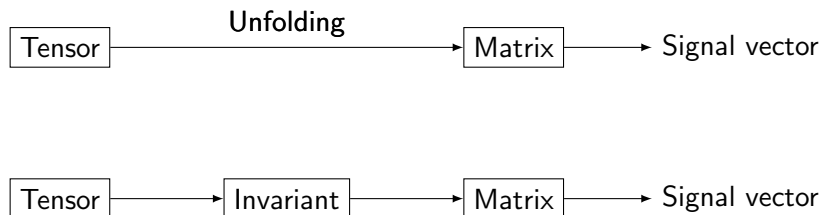


Figure: Using a more pertinent matrix for the PCA

Obtaining a vector requires using a matrix. We want to build a more pertinent one so it is easier to study.

# Decomposition of the tensor

- We can represent the tensor from which we hope to extract the signal represented graphically as:


$$\mathbf{T}_{ij_1 \dots j_{k-1}} = \sqrt{N} \beta \mathbf{v}_i \mathbf{v}_{j_1} \dots \mathbf{v}_{j_{k-1}} + \mathbf{Z}_{ij_1 \dots j_{k-1}}$$


Figure: Graphical decomposition of the tensor  $\mathbf{T}$

- We decompose in a similar way the tensor trace invariant into  $\mathbf{B}^{(0)}$  associated to the pure noise tensor and  $\mathbf{B}^{(1)}$  enclosing the additional contributions from the signal.

$$\mathbf{B} = \mathbf{B}^{(0)} + \mathbf{B}^{(1)}$$



# Matrix decomposition

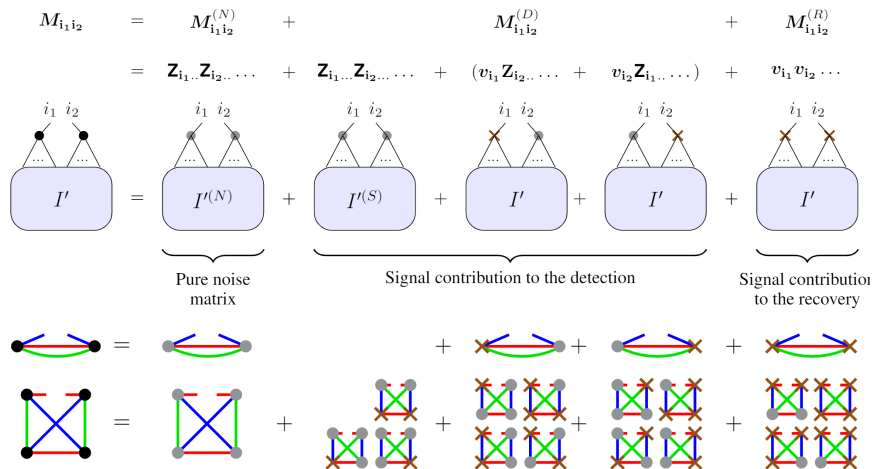
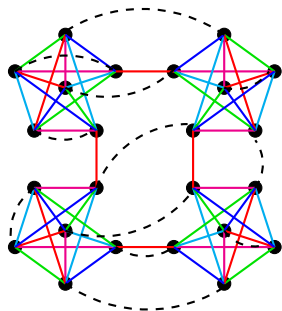


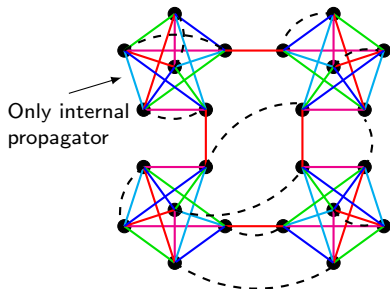
Figure: Graph decomposition of a matrix

# Graph associated to the norm operator

Let  $\mathbf{A} = \mathbf{M}_{\mathcal{G},e}^\top \mathbf{M}_{\mathcal{G},e}$  where  $\mathbf{M}_{\mathcal{G},e}$  is the matrix associated to a graph  $\mathcal{G}$  and an edge  $e$ . The trace of the power  $r$  of  $\mathbf{A}$ ,  $\text{Tr}(\mathbf{A}^r)$  can be represented as the graph gluing the open edges of  $2r \mathcal{G}$ .



(a) Covering graph contributing



(b) Covering graph not contributing to  $\mathbb{E}(\text{Tr}((\mathbf{M} - \mathbb{E}(\mathbf{M}))^2))$

Figure: Two covering graphs for the graph of  $\text{Tr}((\mathbf{M}^\top \mathbf{M})^2)$

---

**Algorithm 1:** Recovery algorithm associated to the graph  $\mathcal{G}$  and edge  $e$

---

**Input:** The tensor  $\mathbf{T} = \beta \mathbf{v}^{\otimes k} + \mathbf{Z}$

**Goal:** Estimate  $\mathbf{v}_0$ .

Calculate the matrix  $\mathbf{M}_{\mathcal{G},e}(\mathbf{T})$

Compute its top eigenvector by matrix power iteration (repeat  $v_i \leftarrow M_{ij} v_j$ ).

**Output:** Obtaining an estimated vector  $\mathbf{v}$

---

**Note:** You can apply Tensor Power iteration ( $\mathbf{v} \leftarrow \frac{\mathbf{T}\mathbf{v}\mathbf{v}}{\|\mathbf{T}\mathbf{v}\mathbf{v}\|}$ ) on the output of the algorithm to slightly increase the precision.

- It appears that the two state of the art (SOTA) methods are equivalent to the algorithms associated to graphs of degree 2.
- This striking fact incites us to investigate the algorithm associated to the tetrahedral graph which is a graph of degree 4 as illustrated in the following Figure.

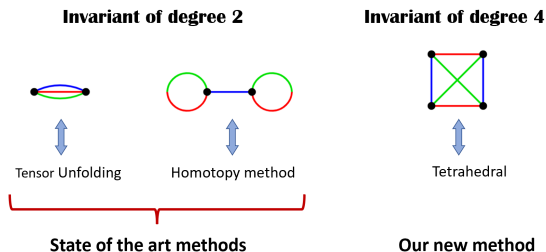


Figure: Methods associated to invariant graphs

## Simple generalization: perfect one factorization

A one factorization graph is said to be perfect if the union of any two of its distinct 1-factors (edges of a given color) is a cycle that visits each vertex exactly once (also called maximally single trace (MST) graphs [Ferrari et al., 2019]).

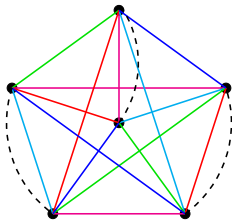
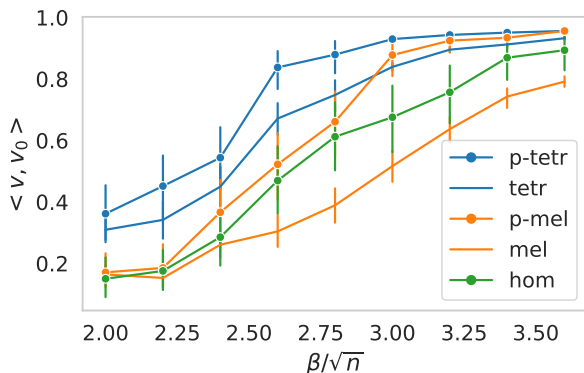


Figure: Complete graph for  $N = 5$

### Theorem

*The algorithm associated to a perfect one-factorization graph is able to recover the signal vector for  $\beta = O(n^{k/4})$ .*

# Numerical experiment



**Figure:** Comparison of different methods for symmetric recovery.  $n=150$ . The prefix "p-" indicates that power iteration is performed on the output of the method.

## Novel theoretical threshold

Let's first consider the more general case where the tensor  $\mathbf{T}$  has axes of different dimensions  $n_i$  ( $\mathbf{T} \in \bigotimes_{i=1}^k \mathbb{R}^{n_i}$ ). We can assume without any loss of generality that  $n_1 \geq n_2 \geq \dots \geq n_k$ .

$$\mathbf{T} = \beta \mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_k + \mathbf{Z} \quad \text{where} \quad \mathbf{v}_i \in \mathbb{R}^{n_i}, \quad n_i \in \mathbb{N}. \quad (2)$$

### Theorem

*Using the melon graph, the threshold for  $\mathbf{v}_1$  is given by  $\max\left(\left(\prod_{i=1}^k n_i\right)^{1/4}, n_1^{1/2}\right)$  while the thresholds for  $\mathbf{v}_j$ ,  $j \geq 2$  are equal to  $\left(\prod_{i=1}^k n_i\right)^{1/4}$ .*

# Algorithmic threshold given by the generalized melon

## Theorem

Let  $k \geq 3$ . It is impossible to detect or recover the signal using a single graph below the threshold  $\beta \leq n^{(k-2)/4}$  which is related to the minimal Gaussian variance of any graph  $\mathcal{G}$ .

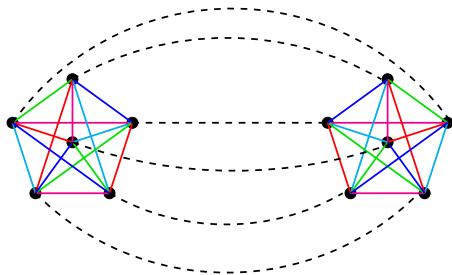


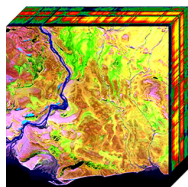
Figure: Generalized melon for the complete graph for  $N = 5$



One of the advantages of this framework is that it is generalizable for the tensor decomposition model (both CP and Tucker decomposition).

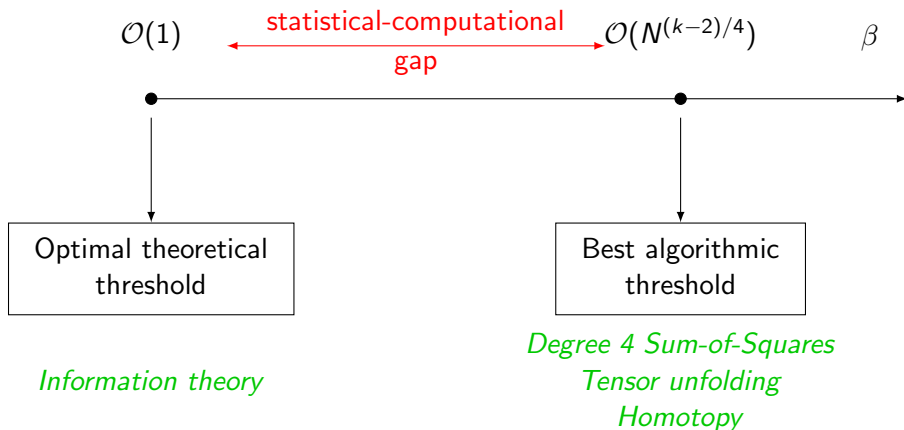
We performed experiments on the problem of denoising Hyperspectral images (HSI) on a real world data: the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) HSI, an airborne hyperspectral system flown by NASA/Jet Propulsion Laboratory (JPL). We showed an improvement on the performance comparing to existent methods.

This first strategy was the object of a paper : Ouerfelli, Tamaazousti and Rivasseau (2022). Random tensor theory for tensor decomposition. *In Proceedings of the AAAI Conference on Artificial Intelligence*



- 1 Introduction
- 2 First strategy : Random Tensor Theory
- 3 Statistical-computational gap**
- 4 Second strategy : SMPI
- 5 Conclusion

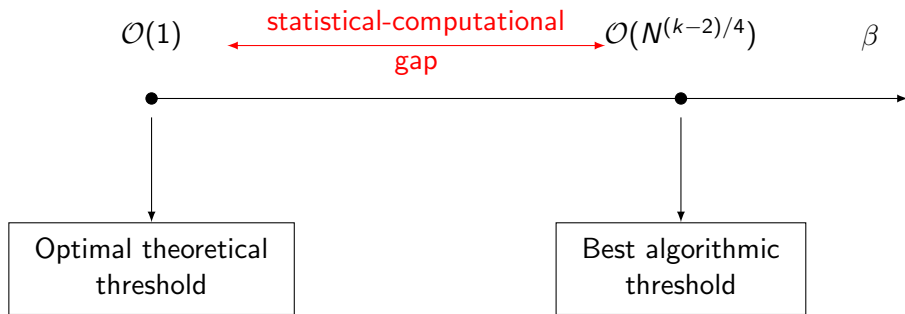
# Existent Algorithms



# 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

# Existent Algorithms



*Tensor resolvent :  
Infinite sum of graphs*

*One graph based method.*

The resolvent  $\omega(w; \mathbf{T}) = \sum_{n \geq 0} \frac{1}{w^{n+1}} \frac{1}{N} \sum_{b \in \mathcal{B}_n} \text{Tr}_b(\mathbf{T})$  where  $\mathcal{B}_n$  denotes the set of connected rooted  $p$ -valent maps with  $n$  unlabelled vertices.

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].

- 1 Introduction
- 2 First strategy : Random Tensor Theory
- 3 Statistical-computational gap
- 4 Second strategy : SMPI**
- 5 Conclusion

# Naive power iteration

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

$$\begin{aligned}\mathbf{T}\mathbf{v}\mathbf{v} &= \mathbf{Z}\mathbf{v}\mathbf{v} + \beta\langle\mathbf{v},\mathbf{v}_0\rangle^2\mathbf{v}_0 \\ &\equiv \mathbf{g}_N + \mathbf{g}_S\end{aligned}$$

- A naive approach is to consider that  $\mathbf{g}_N$  is a random vector at each step. Thus, we have to study in which case we can increase the correlation with  $\mathbf{v}_0$  at each step.
  - If  $\beta \gg n^{(k-2)/2}$ , power iteration will always be successful.
  - If  $\beta \ll n^{(k-2)/2}$ , power iteration with a random initialization will fail with an exponential probability.



# Power iteration and gradient descent

- The gradient of a function  $f$  on a sphere is given by  $\nabla f(\mathbf{v}) - (\nabla f(\mathbf{v}) \cdot \mathbf{v}) \cdot \mathbf{v} = \mathbf{T}\mathbf{v}\mathbf{v} - (\mathbf{T}\mathbf{v}\mathbf{v}\mathbf{v})\mathbf{v}$  (more mathematical details in [Ros et al., 2019])
- In our case, it is equal to  $g = \mathbf{T}\mathbf{v}\mathbf{v} - (\mathbf{T}\mathbf{v}\mathbf{v}\mathbf{v})\mathbf{v}$ .
- The power iteration could be considered as a gradient descent with a step size equal to  $1/(\mathbf{T}\mathbf{v}\mathbf{v}\mathbf{v})$  indeed  
$$\mathbf{T}\mathbf{v}\mathbf{v} = \mathbf{T}\mathbf{v}\mathbf{v} - (\mathbf{T}\mathbf{v}\mathbf{v}\mathbf{v})\mathbf{v} + (\mathbf{T}\mathbf{v}\mathbf{v}\mathbf{v})\mathbf{v} = g + (\mathbf{T}\mathbf{v}\mathbf{v}\mathbf{v})\mathbf{v} = ((\mathbf{T}\mathbf{v}\mathbf{v}\mathbf{v})\mathbf{v})\left(\mathbf{v} + \frac{g}{(\mathbf{T}\mathbf{v}\mathbf{v}\mathbf{v})\mathbf{v}}\right).$$

---

## Algorithm 2: Selective Multiple Power Iteration

---

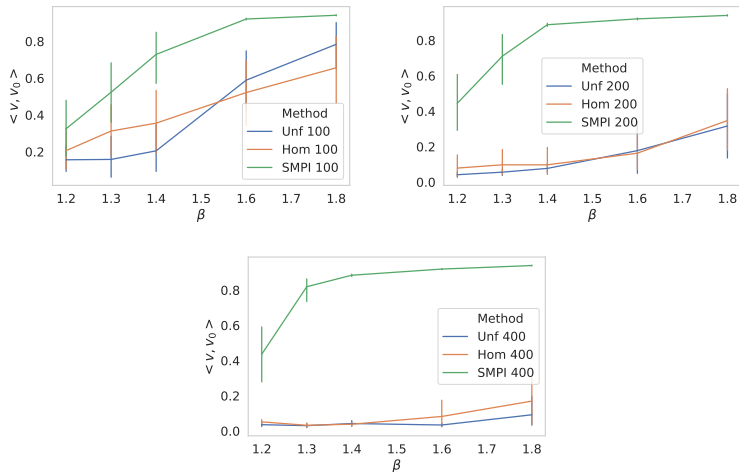
- 1: **Input:** The tensor  $\mathbf{T} = \mathbf{Z} + \beta \mathbf{v}_0^{\otimes k}$ ,  $m_{\text{init}} > 10n$ ,  $m_{\text{iter}} > 10n, \Lambda$
  - 2: **Goal:** Estimate  $\mathbf{v}_0$ .
    - 3:     **for**  $i=0$  to  $m_{\text{init}}$  **do**
      - 4:         Generate a random vector  $\mathbf{v}_{i,0}$  **for**  $j=0$  to  $m_{\text{iter}}$  **do**
        - 5:             
$$\mathbf{v}_{i,j+1} = \frac{\mathbf{T}(:, \mathbf{v}_{i,j}, \mathbf{v}_{i,j})}{\|\mathbf{T}(:, \mathbf{v}_{i,j}, \mathbf{v}_{i,j})\|}$$
**if**  $j > \Lambda$  and  $|\langle \mathbf{v}_{i,j-\Lambda}, \mathbf{v}_{i,j} \rangle| \geq 1 - \varepsilon$  **then**
          - 6:                  $\mathbf{v}_{i,m_{\text{iter}}} = \mathbf{v}_{i,j}$
  - 7: Select the vector  $\mathbf{v} = \arg \max_{1 \leq i \leq m_{\text{init}}} \mathbf{T}(\mathbf{v}_{i,m_{\text{iter}}}, \mathbf{v}_{i,m_{\text{iter}}}, \mathbf{v}_{i,m_{\text{iter}}})$
  - 8: **Output:** the estimated vector  $\mathbf{v}$
-

# Comparison Features

**Table:** The five essential features of SMPI compared to other power iteration based studies

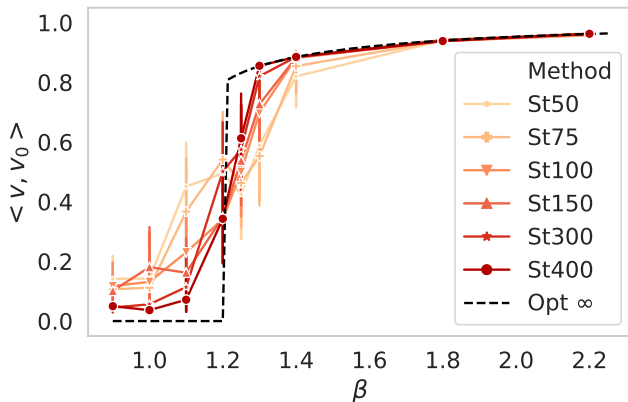
| Algorithm             | Symmetry | Discreet step size | Poly. nb of initialisat. | Poly. nb of iterations | No stop condition |
|-----------------------|----------|--------------------|--------------------------|------------------------|-------------------|
| Wang et al. 2017      | Yes      | Yes                | No                       | No                     | No                |
| Huang et al. 2020     | No       | Yes                | Yes                      | Yes                    | Yes               |
| Ben Arous et al. 2020 | Yes      | No                 | Yes                      | Yes                    | Yes               |
| Dudeja et al. 2022    | Yes      | Yes                | Yes                      | No                     | Yes               |
| SMPI 2021             | Yes      | Yes                | Yes                      | Yes                    | Yes               |

# Empirical comparison



**Figure:** Comparison of the results of SMPI with the unfolding method (Unf) and Homotopy-based method (Hom)

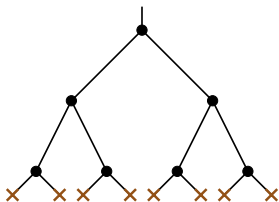
# Statistical algorithmic gap



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

# Possible ideas for a theoretical proof

- Adapting the method of Oleg Evnin that previously studied power iteration in order to investigate the largest eigenvalue of a random tensor in [Evnin, 2020].



**Figure:** The graph associated to the power iteration method with 3 iterations for an initialization  $\mathbf{v}$ . The cross represents the vector  $\mathbf{v}$  and the black dot the tensor  $\mathbf{T}$ .

- Using advanced probability tools to prove the success of SMPI (with Ben Arous).





- 1 Introduction
- 2 First strategy : Random Tensor Theory
- 3 Statistical-computational gap
- 4 Second strategy : SMPI
- 5 Conclusion**

# Conclusion

- The results obtained and the new insights opens the way to explore further questions:
  - Spin glass phenomena
  - Gradient descent dynamics in machine learning
  - Concrete applications (compression of neural network, telecommunication, etc.).
  - New mathematical tools
  - Quantum gravity
- Possible approaches to improve our understanding:
  - Investigate what choice of graphs to sum.
  - Adapt existing probability methods for a theoretical proof of SMPI.
  - Renormalization group approach.

⇒ New group at CEA Paris Saclay to explore these multiple directions.



-  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.
-  Avohou, R. C., Ben Geloun, J., and Dub, N. (2020).  
On the counting of  $O(N)$  tensor invariants.  
*Adv. Theor. Math. Phys.*, 24(4):821–878.
-  Cunningham, J. P. and Byron, M. Y. (2014).  
Dimensionality reduction for large-scale neural recordings.  
*Nature neuroscience*, 17(11):1500–1509.
-  De la Torre, F. and Black, M. J. (2001).  
Robust principal component analysis for computer vision.  
In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 1, pages 362–369. IEEE.



Evnin, O. (2020).

Melonic dominance and the largest eigenvalue of a large random tensor.

*arXiv preprint arXiv:2003.11220.*



Ferrari, F., Rivasseau, V., and Valette, G. (2019).

A New Large  $N$  Expansion for General Matrix–Tensor Models.

*Commun. Math. Phys.*, 370(2):403–448.



Gurau, R. (2017).

*Random Tensors.*

Oxford University Press.



Huang, R., Armengaud, E., Augier, C., Barabash, A., Bellini, F., Benato, G., Benoît, A., Beretta, M., Bergé, L., Billard, J., et al. (2020).

Pulse shape discrimination in cupid-mo using principal component analysis.

*arXiv preprint arXiv:2010.04033.*



Jagannath, A., Lopatto, P., and Miolane, L. (2020).

Statistical thresholds for tensor pca.

*The Annals of Applied Probability*, 30(4):1910–1933.



Joreskog, K. G., Klovan, J. E., and Reyment, R. A. (1976).

*Geological factor analysis*.

Elsevier Scientific Pub. Co.



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.



Pasini, G. (2017).

Principal component analysis for stock portfolio management.

*International Journal of Pure and Applied Mathematics*, 115(1):153–167.



Qi, L. (2005).

Eigenvalues of a real supersymmetric tensor.

*Journal of Symbolic Computation*, 40(6):1302 – 1324.



Richard, E. and Montanari, A. (2014).

A statistical model for tensor pca.

In *Advances in Neural Information Processing Systems*, pages 2897–2905.



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.



Wold, S., Esbensen, K., and Geladi, P. (1987).

Principal component analysis.

*Chemometrics and intelligent laboratory systems*, 2(1-3):37–52.