Probabilistic PCA on Tensors

(arXiv:2510.19516)

Piotr Zwiernik

Universitat Pompeu Fabra *Barcelona* — Statistics, Probability and Machine Learning Research Group

joint work with Yaoming Zhen (CUHK, Shenzhen)

OCAMI Workshop 2025

Tensor Basics

Basic terminology

Matrix: $\mathbf{M} = [\mathbf{M}_{i_1 i_2}] \in \mathbb{R}^{n_1 \times n_2}$.

Tensor: $\mathbf{T} = [\mathbf{T}_{i_1 i_2 \cdots i_r}] \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_r}$, $r = 1, 2, 3, \ldots$

Basic terminology

Matrix: $\quad \mathbf{M} = [\mathbf{M}_{i_1 i_2}] \in \mathbb{R}^{n_1 \times n_2}$.

Tensor: $\quad \mathbf{T} = [\mathbf{T}_{i_1 i_2 \cdots i_r}] \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_r}, \quad r = 1, 2, 3, \ldots$

Rank one matrix: $\quad \mathbf{M} = \mathbf{u}_1 \mathbf{u}_2^\top, \qquad \mathbf{u}_1 \in \mathbb{R}^{n_1}, \ \mathbf{u}_2 \in \mathbb{R}^{n_2}$.

$$\mathbf{M}_{i_1 i_2} = (\mathbf{u}_1)_{i_1} (\mathbf{u}_2)_{i_2}$$

Rank one tensor: $\quad \mathbf{T} = \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_r, \quad \mathbf{u}_i \in \mathbb{R}^{n_i} \text{ for } i = 1, \ldots, r$

$$\mathbf{T}_{i_1 \cdots i_r} = (\mathbf{u}_1)_{i_1} \cdots (\mathbf{u}_r)_{i_r}$$

Multilinear action on tensors

Matrix $\mathbf{A} \in \mathbb{R}^{n_1 \times m_1}$ acts on $\mathbf{v} \in \mathbb{R}^{m_1}$: $\mathbf{v} \mapsto \mathbf{A}\mathbf{v} \in \mathbb{R}^{n_1}$.

Multilinear action on tensors

Matrix $\mathbf{A} \in \mathbb{R}^{n_1 \times m_1}$ acts on $\mathbf{v} \in \mathbb{R}^{m_1}$: $\mathbf{v} \mapsto \mathbf{A}\mathbf{v} \in \mathbb{R}^{n_1}$.

Matrices $\mathbf{A} \in \mathbb{R}^{n_1 \times m_1}$, $\mathbf{B} \in \mathbb{R}^{n_2 \times m_2}$ act on $\mathbf{M} \in \mathbb{R}^{m_1 \times m_2}$

$$\mathbf{M} \;\mapsto\; (\mathbf{A}, \mathbf{B}) \cdot \mathbf{M} := \mathbf{A}\mathbf{M}\mathbf{B}^\top \in \mathbb{R}^{n_1 \times n_2}$$

Note: $\quad [\mathbf{A}\mathbf{M}\mathbf{B}^\top]_{i_1 i_2} \;=\; \sum_{j_1, j_2} \mathbf{A}_{i_1 j_1} \mathbf{B}_{i_2 j_2} \mathbf{M}_{j_1 j_2}.$

Multilinear action on tensors

Matrix $\mathbf{A} \in \mathbb{R}^{n_1 \times m_1}$ acts on $\mathbf{v} \in \mathbb{R}^{m_1}$: $\mathbf{v} \mapsto \mathbf{A}\mathbf{v} \in \mathbb{R}^{n_1}$.

Matrices $\mathbf{A} \in \mathbb{R}^{n_1 \times m_1}$, $\mathbf{B} \in \mathbb{R}^{n_2 \times m_2}$ act on $\mathbf{M} \in \mathbb{R}^{m_1 \times m_2}$

$$\mathbf{M} \mapsto (\mathbf{A}, \mathbf{B}) \cdot \mathbf{M} := \mathbf{A}\mathbf{M}\mathbf{B}^\top \in \mathbb{R}^{n_1 \times n_2}$$

Note: $\quad [\mathbf{A}\mathbf{M}\mathbf{B}^\top]_{i_1 i_2} = \sum_{j_1, j_2} \mathbf{A}_{i_1 j_1} \mathbf{B}_{i_2 j_2} \mathbf{M}_{j_1 j_2}$.

Matrices $\mathbf{A}_1 \in \mathbb{R}^{n_1 \times m_1}, \ldots, \mathbf{A}_r \in \mathbb{R}^{n_r \times m_r}$ act on $\mathbf{T} \in \mathbb{R}^{m_1 \times \cdots \times m_r}$

$$\mathbf{T} \mapsto (\mathbf{A}_1, \ldots, \mathbf{A}_r) \cdot \mathbf{T} \in \mathbb{R}^{n_1 \times \cdots \times n_r}$$

where $\quad [(\mathbf{A}_1, \ldots, \mathbf{A}_r) \cdot \mathbf{T}]_{i_1 \cdots i_r} = \sum_{j_1, \ldots, j_r} (\mathbf{A}_1)_{i_1 j_1} \cdots (\mathbf{A}_r)_{i_r j_r} \mathbf{T}_{j_1 \cdots j_r}$.

Tucker decomposition

Singular Value Decomposition: For every matrix $\mathbf{M} \in \mathbb{R}^{\mathbf{n_1} \times \mathbf{n_2}}$ s.t. $\mathrm{rank}(\mathbf{M}) \leqslant \mathbf{s}$ there exist $\mathbf{U} \in \mathbb{R}^{\mathbf{n_1} \times \mathbf{s}}$ and $\mathbf{V} \in \mathbb{R}^{\mathbf{n_2} \times \mathbf{s}}$ such that and

$$\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^{\top} = (\mathbf{U}, \mathbf{V}) \cdot \mathbf{D}$$

with $\mathbf{U}^{\top}\mathbf{U} = \mathbf{V}^{\top}\mathbf{V} = \mathbb{I}_{\mathbf{s}}$ and $\mathbf{D} = \mathrm{diag}(\sigma_{\mathbf{1}}, \ldots, \sigma_{\mathbf{s}}) \in \mathbb{R}^{\mathbf{s} \times \mathbf{s}}$.

Tucker decomposition

Singular Value Decomposition: For every matrix $\mathbf{M} \in \mathbb{R}^{\mathbf{n_1} \times \mathbf{n_2}}$ s.t. $\mathrm{rank}(\mathbf{M}) \leqslant \mathbf{s}$ there exist $\mathbf{U} \in \mathbb{R}^{\mathbf{n_1} \times \mathbf{s}}$ and $\mathbf{V} \in \mathbb{R}^{\mathbf{n_2} \times \mathbf{s}}$ such that and

$$\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^{\top} = (\mathbf{U}, \mathbf{V}) \cdot \mathbf{D}$$

with $\mathbf{U}^{\top}\mathbf{U} = \mathbf{V}^{\top}\mathbf{V} = \mathbb{I}_{\mathbf{s}}$ and $\mathbf{D} = \mathrm{diag}(\sigma_{\mathbf{1}}, \dots, \sigma_{\mathbf{s}}) \in \mathbb{R}^{\mathbf{s} \times \mathbf{s}}$.

A similar statement is true for only very special tensors. Nevertheless...

Tucker decomposition: For $\mathbf{T} \in \mathbb{R}^{\mathbf{n_1} \times \dots \times \mathbf{n_r}}$ and $\mathbf{m_1} \leqslant \mathbf{n_1}, \dots, \mathbf{m_r} \leqslant \mathbf{n_r}$ approximate:

$$\mathbf{T} \approx (\mathbf{A_1}, \dots, \mathbf{A_r}) \cdot \mathbf{S},$$

where $\mathbf{S} \in \mathbb{R}^{\mathbf{m_1} \times \dots \times \mathbf{m_r}}$ is called the core tensor ("multivariate SVD").

Tensor PPCA

PPCA: Probabilistic Principle Component Analysis

PPCA (Tipping & Bishop, 1999): For a random vector $\mathbf{X} \in \mathbb{R}^{\mathbf{n}}$:

$$\mathbf{X} = \mu + \mathbf{A}\mathbf{Z} + \mathbf{E}, \qquad \mu \in \mathbb{R}^{\mathbf{n}}, \; \mathbf{A} \in \mathbb{R}^{\mathbf{n} \times \mathbf{m}}, \; \mathbf{m} \leqslant \mathbf{n},$$

$$\mathbf{Z} \sim \mathbf{N_m}(\mathbf{0}, \mathbb{I_m}), \qquad \mathbf{E} \sim \mathbf{N_n}(\mathbf{0}, \sigma^2 \mathbb{I_n}), \qquad \mathbf{Z} \perp\!\!\!\perp \mathbf{E}.$$

Geometry: samples concentrate around the $\mu + \mathrm{Im}(\mathbf{A})$; $\mathrm{rank}(\mathbf{A}) = \mathbf{m}$.

Note: We can assume $\mathbf{A}^\top \mathbf{A}$ is diagonal (column orthogonal).

- $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, $\mathbf{V} \in \mathbb{R}^{\mathbf{m} \times \mathbf{m}}$ orthogonal, and $\mathbf{V}^\top \mathbf{Z} \overset{\mathbf{d}}{=} \mathbf{Z}$; thus $\mathbf{V} = \mathbb{I_m}$ w.l.o.g.

Tensor PPCA

Suppose now $\mathbf{X} \in \mathbb{V} := \mathbb{R}^{\mathbf{n_1} \times \cdots \times \mathbf{n_r}}$ is a tensor.

Tensor PCA model: $\mathbf{X}$ admits stochastic representation

$$\mathbf{X} = \mu + (\mathbf{A_1}, \ldots, \mathbf{A_r}) \cdot \mathbf{Z} + \mathbf{E},$$

where $\mathbf{Z} \in \mathbb{U} := \mathbb{R}^{\mathbf{m_1} \times \cdots \times \mathbf{m_r}}$ standard normal, $\mathbf{E} \in \mathbb{V}$ isotropic, $\mathbf{Z} \perp\!\!\!\perp \mathbf{E}$.

Dimension reduction assuming: $\mu = \mathbf{0}$ or $\mu = (\mathbf{A_1}, \ldots, \mathbf{A_r}) \cdot \nu$.

Natural questions: identifiability, estimation, learning $\mathbf{m_1}, \ldots, \mathbf{m_r}$.

The distribution of $\mathbf{X}$ is Gaussian

Recall: $\mathbf{X} = (\mathbf{A_1}, \ldots, \mathbf{A_r}) \cdot \mathbf{Z} + \mathbf{E}$ (we assumed zero mean)

Inner product on $\mathbb{V}$: $\quad \langle \mathbf{S}, \mathbf{T} \rangle := \sum_{\mathbf{i_1}, \ldots, \mathbf{i_r}} \mathbf{S_{i_1 \cdots i_r}} \mathbf{T_{i_1 \ldots i_r}}, \quad \|\mathbf{T}\|_{\mathrm{F}} = \sqrt{\langle \mathbf{T}, \mathbf{T} \rangle}.$

$\langle \mathbf{X}, \mathbf{T} \rangle = \langle (\mathbf{A_1}, \ldots, \mathbf{A_r}) \cdot \mathbf{Z}, \mathbf{T} \rangle + \langle \mathbf{E}, \mathbf{T} \rangle = \langle \mathbf{Z}, (\mathbf{A_1^\top}, \ldots, \mathbf{A_r^\top}) \cdot \mathbf{T} \rangle + \langle \mathbf{E}, \mathbf{T} \rangle$

The distribution of $\mathbf{X}$ is Gaussian

Recall: $\mathbf{X} = (\mathbf{A_1}, \ldots, \mathbf{A_r}) \cdot \mathbf{Z} + \mathbf{E}$ (we assumed zero mean)

Inner product on $\mathbb{V}$: $\langle \mathbf{S}, \mathbf{T} \rangle := \sum_{\mathbf{i_1}, \ldots, \mathbf{i_r}} \mathbf{S_{i_1 \cdots i_r}} \mathbf{T_{i_1 \ldots i_r}}, \quad \|\mathbf{T}\|_{\mathrm{F}} = \sqrt{\langle \mathbf{T}, \mathbf{T} \rangle}.$

$\langle \mathbf{X}, \mathbf{T} \rangle = \langle (\mathbf{A_1}, \ldots, \mathbf{A_r}) \cdot \mathbf{Z}, \mathbf{T} \rangle + \langle \mathbf{E}, \mathbf{T} \rangle = \langle \mathbf{Z}, (\mathbf{A_1^\top}, \ldots, \mathbf{A_r^\top}) \cdot \mathbf{T} \rangle + \langle \mathbf{E}, \mathbf{T} \rangle$

We conclude: $\langle \mathbf{X}, \mathbf{T} \rangle \sim \mathbf{N}(\mathbf{0}, \|(\mathbf{A_1^\top}, \ldots, \mathbf{A_r^\top}) \cdot \mathbf{T}\|_{\mathrm{F}}^2 + \|\mathbf{T}\|_{\mathrm{F}}^2).$

The distribution of $\mathbf{X}$ is Gaussian

Recall: $\mathbf{X} = (\mathbf{A_1}, \ldots, \mathbf{A_r}) \cdot \mathbf{Z} + \mathbf{E}$ (we assumed zero mean)

Inner product on $\mathbb{V}$: $\quad \langle \mathbf{S}, \mathbf{T} \rangle := \sum_{\mathbf{i_1}, \ldots, \mathbf{i_r}} \mathbf{S_{i_1 \cdots i_r}} \mathbf{T_{i_1 \cdots i_r}}, \quad \|\mathbf{T}\|_{\mathrm{F}} = \sqrt{\langle \mathbf{T}, \mathbf{T} \rangle}.$

$$\langle \mathbf{X}, \mathbf{T} \rangle = \langle (\mathbf{A_1}, \ldots, \mathbf{A_r}) \cdot \mathbf{Z}, \mathbf{T} \rangle + \langle \mathbf{E}, \mathbf{T} \rangle = \langle \mathbf{Z}, (\mathbf{A_1^\top}, \ldots, \mathbf{A_r^\top}) \cdot \mathbf{T} \rangle + \langle \mathbf{E}, \mathbf{T} \rangle$$

We conclude: $\qquad \langle \mathbf{X}, \mathbf{T} \rangle \sim \mathbf{N}(\mathbf{0}, \|(\mathbf{A_1^\top}, \ldots, \mathbf{A_r^\top}) \cdot \mathbf{T}\|_{\mathrm{F}}^2 + \|\mathbf{T}\|_{\mathrm{F}}^2).$

Covariance operator: $\mathbf{\Sigma} : \mathbb{V} \to \mathbb{V}$ is the linear function

$$\mathbf{\Sigma}(\mathbf{T}) = \sigma^2 \mathbf{T} + (\mathbf{A_1} \mathbf{A_1^\top}, \ldots, \mathbf{A_r} \mathbf{A_r^\top}) \cdot \mathbf{T}$$

Now $\langle \mathbf{X}, \mathbf{T} \rangle \sim \mathbf{N}(\mathbf{0}, \langle \mathbf{\Sigma}(\mathbf{T}), \mathbf{T} \rangle).$ (generalizes multivariate normal)

First results

Spectrum of Σ

Recall: $\Sigma : \mathbb{V} \to \mathbb{V}$ satisfies $\Sigma(\mathbf{T}) = \sigma^2 \mathbf{T} + (\mathbf{A_1 A_1^\top}, \ldots, \mathbf{A_r A_r^\top}) \cdot \mathbf{T}$

Let $\mathbf{u}_i^{(k)}, \lambda_i^{(k)}$ be the eigenvectors and eigenvalues of $\mathbf{B_k = A_k A_k^\top}$.

$$\Sigma(\mathbf{u}_{i_1}^{(1)} \otimes \cdots \otimes \mathbf{u}_{i_r}^{(r)}) = (\sigma^2 + \lambda_{i_1}^{(1)} \cdots \lambda_{i_r}^{(r)}) \cdot \mathbf{u}_{i_1}^{(1)} \otimes \cdots \otimes \mathbf{u}_{i_r}^{(r)}$$

- at most $\mathbf{m_1 \cdots m_r}$ eigenvalues greater than σ^2 $(\mathrm{rank}(\mathbf{B_k}) \leqslant \mathbf{m_k})$,
- remaining $\mathbf{n_1 \cdots n_r - m_1 \cdots m_r}$ eigenvalues are all equal to σ^2.

Σ has rich structure both in eigenvalues and eigenvectors!

Model identifiability

Recall: $\mathbf{X} = (\mathbf{A_1}, \ldots, \mathbf{A_r}) \cdot \mathbf{Z} + \mathbf{E}$

Assumption 1: Each $\mathbf{A_k}$ is column orthogonal.

Assumption 2: $\|\mathbf{B_1}\|_{\mathrm{F}} = \ldots = \|\mathbf{B_r}\|_{\mathrm{F}}$, where $\mathbf{B_k} := \mathbf{A_k}\mathbf{A_k}^{\top}$.

Assumption 3: Either $\mathbf{n_k} > \mathbf{m_k}$ for at least one $\mathbf{k}$ or at least two of the $\mathbf{B_k}$ are not scalar matrices.

Theorem: Under Assumptions 1-3 the parameters $\mathbf{A} = (\mathbf{A_1}, \ldots, \mathbf{A_r})$ and σ^2 are identifiable.

Existence of the MLE

Observation: The MLE for standard PPCA model exists (with probability one) if and only if the sample size N satisfies $N > m$.

Theorem: In TPCA suppose $n_k > m_k$ for some k. Then the MLE exists with probability 1 even for $N = 1$.

To find the MLE we perform the EM algorithm.

This is a computationally expensive procedure.

Next, we study a **computationally efficient estimator**.

Alternative estimator

Pair reshaping

Recall: $\mathbf{\Sigma} : \mathbb{V} \to \mathbb{V}$, $\mathbb{V} = \mathbb{R}^{n_1 \times \cdots \times n_r}$; slightly abusing notation $\mathbf{\Sigma} \in \mathbb{R}^{n_1 \times \cdots \times n_r \times n_1 \times \cdots \times n_r}$

The **pair-reshaping** operator $\mathrm{PAIR}(\cdot)$ that maps $\mathbf{\Sigma}$ to $\mathbb{R}^{n_1^2 \times \cdots \times n_r^2}$:

$$\mathrm{PAIR}(\mathbf{\Sigma}) = \mathrm{VEC}(\mathbf{B_1}) \otimes \cdots \otimes \mathrm{VEC}(\mathbf{B_r}) \; + \; \sigma^2 \, \mathrm{VEC}(\mathbb{I}_{n_1}) \otimes \cdots \otimes \mathrm{VEC}(\mathbb{I}_{n_r}).$$

Pair reshaping

Recall: $\mathbf{\Sigma} : \mathbb{V} \to \mathbb{V}$, $\mathbb{V} = \mathbb{R}^{\mathbf{n_1} \times \cdots \times \mathbf{n_r}}$; slightly abusing notation $\mathbf{\Sigma} \in \mathbb{R}^{\mathbf{n_1} \times \cdots \times \mathbf{n_r} \times \mathbf{n_1} \times \cdots \times \mathbf{n_r}}$

The **pair-reshaping** operator $\mathrm{PAIR}(\cdot)$ that maps $\mathbf{\Sigma}$ to $\mathbb{R}^{\mathbf{n_1^2} \times \cdots \times \mathbf{n_r^2}}$:

$$\mathrm{PAIR}(\mathbf{\Sigma}) = \mathrm{VEC}(\mathbf{B_1}) \otimes \cdots \otimes \mathrm{VEC}(\mathbf{B_r}) \; + \; \sigma^2 \, \mathrm{VEC}(\mathbb{I}_{\mathbf{n_1}}) \otimes \cdots \otimes \mathrm{VEC}(\mathbb{I}_{\mathbf{n_r}}).$$

In the Frobenius norm, the first term massively dominates.

So $\mathrm{PAIR}(\mathbf{\Sigma})$ is **approximately rank one**. In addition:

$$\left\langle \frac{\mathbf{B_k}}{\|\mathbf{B_k}\|_{\mathrm{F}}}, \frac{\mathbb{I}_{\mathbf{n_k}}}{\|\mathbb{I}_{\mathbf{n_k}}\|_{\mathrm{F}}} \right\rangle = \frac{\mathrm{tr}(\mathbf{B_k})}{\|\mathbf{B_k}\|_{\mathrm{F}} \sqrt{\mathbf{n_k}}} \leqslant \sqrt{\frac{\mathbf{m_k}}{\mathbf{n_k}}}$$

Finding a rank one approximation to a given tensor is a very classical problem in tensors and is solved using the power method.

Power method for matrices

Eckart-Young: For a general matrix $\mathbf{M}$, the best rank-1 approximation $\mathbf{M} \approx \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top$ is obtained from its principal left/right singular vectors.

Alternating power iteration: $\quad \mathbf{u}_{t+1} = \dfrac{\mathbf{M}\mathbf{v}_t}{\|\mathbf{M}\mathbf{v}_t\|}, \quad \mathbf{v}_{t+1} = \dfrac{\mathbf{M}^\top \mathbf{u}_{t+1}}{\|\mathbf{M}^\top \mathbf{u}_{t+1}\|}.$

In Tucker notation: $\mathbf{M}\mathbf{v} = (\mathbb{I}, \mathbf{v}) \cdot \mathbf{M}, \quad \mathbf{u}^\top \mathbf{M} = (\mathbf{u}, \mathbb{I}) \cdot \mathbf{M}.$

Our tensor algorithm uses exactly this idea: a rank-1 power method extended from **2** modes to **r** modes; e.g. $\mathbf{u}_1 \leftarrow (\mathbb{I}, \mathbf{u}_2, \ldots, \mathbf{u}_r) \cdot \mathbf{T} / \| \cdot \|$.

Rank-1 power iteration

Recall: $\text{PAIR}(\boldsymbol{\Sigma}) = \text{VEC}(\mathbf{B_1}) \otimes \cdots \otimes \text{VEC}(\mathbf{B_r}) + \sigma^2 \text{VEC}(\mathbb{I}_{\mathbf{n_1}}) \otimes \cdots \otimes \text{VEC}(\mathbb{I}_{\mathbf{n_r}}).$

$$\text{PAIR}(\boldsymbol{\Sigma}) = \mathbf{b_1} \otimes \cdots \otimes \mathbf{b_r} + \text{small perturbation}$$

By Assumption 2, the norms of $\mathbf{b_1}, \ldots, \mathbf{b_r}$ are all equal; say $\omega^{1/r}$.

$$\text{PAIR}(\boldsymbol{\Sigma}) = \omega \, \mathbf{b_1} \otimes \cdots \otimes \mathbf{b_r} + \text{small perturbation}, \quad \|\mathbf{b_1}\| = \cdots = \|\mathbf{b_r}\| = 1$$

Rank-1 power iteration

Recall: $\quad \text{PAIR}(\boldsymbol{\Sigma}) = \text{VEC}(\mathbf{B_1}) \otimes \cdots \otimes \text{VEC}(\mathbf{B_r}) + \sigma^2 \text{VEC}(\mathbb{I}_{\mathbf{n_1}}) \otimes \cdots \otimes \text{VEC}(\mathbb{I}_{\mathbf{n_r}}).$

$$\text{PAIR}(\boldsymbol{\Sigma}) = \mathbf{b_1} \otimes \cdots \otimes \mathbf{b_r} + \text{small perturbation}$$

By Assumption 2, the norms of $\mathbf{b_1}, \ldots, \mathbf{b_r}$ are all equal; say $\omega^{\mathbf{1/r}}$.

$$\text{PAIR}(\boldsymbol{\Sigma}) = \omega\, \mathbf{b_1} \otimes \cdots \otimes \mathbf{b_r} + \text{small perturbation}, \quad \|\mathbf{b_1}\| = \cdots = \|\mathbf{b_r}\| = \mathbf{1}$$

Let $\text{PAIR}(\mathbf{S_N})$ be a sample estimator of $\text{PAIR}(\boldsymbol{\Sigma})$. Our procedure finds:

$$\text{PAIR}(\mathbf{S_N}) \quad \approx \quad \hat{\omega}\, \hat{\mathbf{b}}_{\mathbf{1}} \otimes \cdots \otimes \hat{\mathbf{b}}_{\mathbf{r}}$$

Additional regularization: Each $\hat{\mathbf{b}}_{\mathbf{i}}$ is a vectorization of a low-rank matrix.

Probabilistic guarantees

The right norm

Recall, $\boldsymbol{\Sigma} : \mathbb{V} \to \mathbb{V}$ given by $\boldsymbol{\Sigma}(\mathbf{T}) = \sigma^2 \mathbf{T} + (\mathbf{A_1}\mathbf{A_1}^\top, \ldots, \mathbf{A_r}\mathbf{A_r}^\top) \cdot \mathbf{T}$.

"Standard" operator norm of $\mathbf{F} : \mathbb{V} \to \mathbb{V}$:
$$\|\mathbf{F}\|_{\mathrm{op}} = \max_{\|\mathbf{T}\|_{\mathrm{F}} = 1} \|\mathbf{F}(\mathbf{T})\|_{\mathrm{F}} = \max_{\|\mathbf{S}\|_{\mathrm{F}} = \|\mathbf{T}\|_{\mathrm{F}} = 1} \langle \mathbf{F}(\mathbf{T}), \mathbf{S} \rangle.$$

We work with a weaker norm.
$$\mathbb{B} := \{\mathbf{u} = \mathbf{u_1} \otimes \cdots \otimes \mathbf{u_r} : \|\mathbf{u_1}\| = \cdots = \|\mathbf{u_r}\| = 1\} \subseteq \mathbb{V}$$

$$\|\mathbf{F}\| := \max_{\mathbf{u},\mathbf{v} \in \mathbb{B}} \langle \mathbf{F}(\mathbf{u}), \mathbf{v} \rangle.$$

Concentration

With **N** samples, with high probability,

$$\|\mathbf{S_N} - \boldsymbol{\Sigma}\| \lesssim \underbrace{\sigma^2\,\phi_1(\tfrac{\sum n_k}{N})}_{\text{noise}} + \underbrace{\|A\|^2\,\phi_2(\tfrac{\sum m_k}{N})}_{\text{signal}} + \underbrace{\sigma\|A\|\,\phi_3(\tfrac{\sum n_k + m_k}{N})}_{\text{cross}}.$$

Here:

- $\phi_{\mathbf{i}}$ are dimension-dependent rates specified in the paper.
- $\|\mathbf{A}\| := \|\mathbf{A_1}\| \cdots \|\mathbf{A_r}\|$

$\dfrac{\|\mathbf{S_N}-\boldsymbol{\Sigma}\|}{\|\mathbf{A}\|^2} \to \mathbf{0}$ in probability as long as $\max\{\tfrac{\mathbf{n_{max}}}{\|\mathbf{A}\|^2}, \mathbf{m_{max}}, \tfrac{\mathbf{n_{max}}}{\|\mathbf{A}\|}\}\tfrac{\log \mathbf{N}}{\mathbf{N}} \to \mathbf{0}$.

One-step consistency

Recall: $\text{PAIR}(\mathbf{S_N}) \approx \omega \, \mathbf{b_1} \otimes \cdots \otimes \mathbf{b_r}, \|\mathbf{b_k}\| = 1$. Use power iteration.

ψ = the RHS in the previous slide (small); $\omega = \prod_k \|\mathbf{B_k}\|_F$ (large).

$\theta_k^{(t)}$ = the angle between $\mathbf{b_k}$ and $\mathbf{b_k^{(t)}}$ for $t = 0, 1, 2, \ldots$

$c := \cos^2(\theta_k^{(0)}) \in (0, 1)$. (we need it to be bounded away from 0)

Theorem: Under mild assumptions: $\sin \theta_k^{(1)} \leqslant \frac{\psi}{\omega} \cdot \frac{\sqrt{m_1 \cdots m_r}}{c^{(r-1)/2}}$.

Moreover, $\frac{|\omega - \hat{\omega}^{(1)}|}{\omega} \leqslant \left(\frac{\psi}{\omega}\right)^2 \frac{16 \, r \, m_1 \cdots m_r}{c^{r-1}}$.

Discussion

Multilinear SVD give **deterministic** low-rank practical approximations.

But they lack:

- a probabilistic model for the noise or likelihood for inference,
- uncertainty quantification,
- principled model selection for the ranks m_k.

Our contribution: a probabilistic multilinear model with identifiability, MLE existence, efficient estimation, and probabilistic guarantees.

Discussion

Multilinear SVD give **deterministic** low-rank practical approximations.

But they lack:

- a probabilistic model for the noise or likelihood for inference,
- uncertainty quantification,
- principled model selection for the ranks m_k.

Our contribution: a probabilistic multilinear model with identifiability, MLE existence, efficient estimation, and probabilistic guarantees.

Thank you!

The paper, `arXiv:2510.19516`.

`https://pzwiernik.github.io/slides/tpca-ocami-2025.pdf`