Seminar on Tenosr Optimization
Part I: Introduction and Tensor PCA

Lecturer: Bo Jiang

October 9, 2022

上海财经大学 SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS 1917–2017

Outline

Outline

Introduction

Tucker Core

Tensor PCA

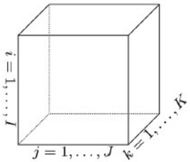
What is A Tensor?

An $n_1 \times n_2$ matrix:

$$A = [a_{i_1 i_2}]_{n_1 \times n_2} \in \mathbb{C}^{n_1 \times n_2}$$

An $n_1 \times n_2 \times \cdots \times n_d$ tensor:

$$\mathcal{A} = [\mathcal{A}_{i_1 i_2 \cdots i_d}] \in \mathbb{C}^{n_1 \times n_2 \times \cdots \times n_d}$$



$\mathcal{A}$ is symmetric if $n_1 = n_2 = \cdots = n_d$ and $\mathcal{A}_{i_1 i_2 \cdots i_d}$ is invariant under any permutation of $\{i_1, i_2, \cdots, i_d\}$.

Space of d-th order super-symmetric tensor is denoted by $\mathbb{S}^{n^d}$.

Tensor Operations I

- The outer product between $\mathcal{A}_1 \in \mathbb{C}^{n_1 \times \cdots \times n_d}$ and $\mathcal{A}_2 \in \mathbb{C}^{n_{d+1} \times \cdots \times n_{d+\ell}}$:

$$(\mathcal{A}_1 \otimes \mathcal{A}_2)_{i_1 i_2 \cdots i_{d+\ell}} = (\mathcal{A}_1)_{i_1 i_2 \cdots i_d} (\mathcal{A}_2)_{i_{d+1} \cdots i_{d+\ell}}.$$

- The outer product among vectors

$$(x^1 \otimes x^2 \otimes \cdots \otimes x^d)_{i_1 i_2 \cdots i_m} = \prod_{k=1}^{d} (x^k)_{i_k} \text{ (Rank-one tensor).}$$

- The inner product between $\mathcal{A}_1$ and $\mathcal{A}_2 \in \mathbb{C}^{n_1 \times n_2 \times \cdots \times n_d}$

$$\mathcal{A}_1 \bullet \mathcal{A}_2 = \sum_{i_1, i_2, \ldots, i_d} (\mathcal{A}_1)_{i_1 i_2 \cdots i_d} (\mathcal{A}_2)_{i_1 i_2 \cdots i_d}.$$

- $$\mathcal{F}(\underbrace{x, \cdots, x}_{2d}) = \sum_{1 \le i_1, \cdots, i_d \le n} \mathcal{F}_{i_1 \cdots i_{2d}} x_{i_1} \cdots x_{i_{2d}} = \mathcal{F} \bullet \underbrace{x \otimes \cdots \otimes x}_{2d}$$

Tensor Operations II

- The Kronecker product of matrices $A \in \mathbb{R}^{I \times J}$ and $B \in \mathbb{R}^{K \times L}$ is denoted by

$$A \circ B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1J}B \\ a_{21}B & a_{22}B & \cdots & a_{2J}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}B & a_{I2}B & \cdots & a_{IJ}B \end{bmatrix}.$$

- The Khatri-Rao is defined by

$$A \odot B = \begin{bmatrix} a_1 \circ b_1 & a_2 \circ b_2 & \cdots & a_K \circ b_K \end{bmatrix}$$

CP Decomposition I

- The CP (Candecomp/Parafac) rank (denoted by $\mathrm{rank}_{CP}(\cdot)$):
  For $\mathcal{F} \in \mathbb{C}^{n_1 \times n_2 \times \cdots \times n_d}$, the smallest integer r such that

$$\mathcal{F} = \sum_{i=1}^{r} a^{1,i} \otimes a^{2,i} \otimes \cdots \otimes a^{d,i}, \quad a^{k,i} \in \mathbb{C}^{n_i}.$$

- The symmetric CP rank (denoted by $\mathrm{rank}_{SCP}(\cdot)$):
  For $\mathcal{F} \in \mathbb{S}^{n^d}$, the smallest integer r such that

$$\mathcal{F} = \sum_{i=1}^{r} \underbrace{a^i \otimes \cdots \otimes a^i}_{d}, \quad a^i \in \mathbb{C}^n.$$

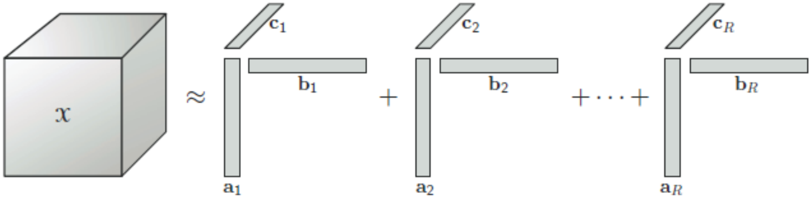
CP DECOMPOSITION II

- The real symmetric CP rank (denoted by $\text{rank}_{RCP}(\cdot)$):

  For $\mathcal{F} \in \mathbf{S}^{n^d}$, the smallest integer r such that

$$\mathcal{F} = \sum_{i=1}^{r} \lambda_i \underbrace{a^i \otimes \cdots \otimes a^i}_{d}, \quad a^i \in \mathbb{R}^n, \lambda_i \in \mathbb{R}^1.$$

CP Decomposition III



Challenge: computing the rank of a tensor

- Determine the CP rank of a specific tensor is already NP-hard in general.
- There is a particular $9 \times 9 \times 9$ tensor, whose rank is only known to be in between 18 and 23.

(1,1,1): 1	(4,2,1): 1	(7,3,1): 1
(1,4,2): 1	(4,5,2): 1	(7,6,2): 1
(1,7,3): 1	(4,8,3): 1	(7,9,3): 1
(2,1,4): 1	(5,2,4): 1	(8,3,4): 1
(2,4,5): 1	(5,5,5): 1	(8,6,5): 1
(2,7,6): 1	(5,8,6): 1	(8,9,6): 1
(3,1,7): 1	(6,2,7): 1	(9,3,7): 1
(3,4,8): 1	(6,5,8): 1	(9,6,8): 1
(3,7,9): 1	(6,8,9): 1	(9,9,9): 1

Uniqueness of Tensor Rank Decomposition I

- Consider a matrix $X \in \mathbb{R}^{I \times J}$ with rank R.

  Then a rank decomposition of X is

$$X = AB^\top = \sum_{r=1}^{R} a_r \otimes b_r.$$

- Let the SVD of X be $X = U\Sigma V^\top$ then we can choose:
  - $A = U\Sigma$ and $B = V$;
  - $A = U\Sigma W$ and $B = VW$, where W is some $R \times R$ orthogonal matrix.
- We can easily construct two completely different sets of R rank-one matrices that sum to the original matrix.

Uniqueness of Tensor Rank Decomposition II

- For a three-way tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ of rank R, suppose $\mathcal{X}$ has a CP decomposition:

$$\mathcal{X} = \sum_{i=1}^{R} a_r \otimes b_r \otimes c_r := [A, B, C],$$

where $A = [a_1, \cdots, a_R]$ and likewise for B and C.

- By uniqueness we mean the decomposition is unique in the sense of scaling and permutation, i.e.

$$\mathcal{X} = [\Pi A, \Pi B, \Pi C] = \sum_{i=1}^{R} (\alpha_r a_r) \otimes (\beta_r b_r) \otimes (\gamma_r c_r),$$

for any permutation matrix Π and $\alpha_r \beta_r \gamma_r = 1$ for r=1,…,R.

Uniqueness of Tensor Rank Decomposition III

- Define the $k-rank$ of a matrix A as k_A, which is the maximum value k such that **any** k columns of A are linearly independent.
- For a N-way tensor $\mathcal{X}$ with rank R and CP-decomposition

$$\mathcal{X} = \sum_{r=1}^{R} a_r^{(1)} \otimes a_r^{(2)} \otimes \cdots \otimes a_r^{(N)} = [A^{(1)}, A^{(2)}, \dots, A^{(N)}].$$

- The sufficient condition for uniqueness is

$$\sum_{n=1}^{N} k_{A^{(n)}} \geq 2R + (N-1).$$

Uniqueness of Tensor Rank Decomposition IV

- The necessary condition for uniqueness is

$$\min_{n=1,\ldots,N} \quad \mathrm{rank}(A^{(1)} \odot \cdots \odot A^{(n-1)} \odot A^{(n+1)} \odot \cdots \odot A^{(N)}) = R,$$

  Further, since $\mathrm{rank}(A \odot B) \leq \mathrm{rank}(A \circ B) \leq \mathrm{rank}(A)\mathrm{rank}(B)$, a simpler necessary condition is

$$\min_{n=1,\ldots,N} \quad \left(\Pi_{m=1,m\neq n}^{N} \mathrm{rank}(A^{(m)})\right) \geq R$$

Discontinuity of Tensor Rank I

- For matrices, the best rank-k approximation is given by the leading k factors of the SVD, i.e. if a rank R matrix A has its SVD

$$A = \sum_{r=1}^{R} \sigma_r u_r \otimes v_r \quad \text{with} \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_R > 0,$$

then a rank-k approximation that minimizes $\|A - B\|$ is given by

$$B = \sum_{r=1}^{k} \sigma_r u_r \otimes v_r.$$

- However, the result does not hold true for higher-order tensors.

Discontinuity of Tensor Rank II

- For example, a rank-three tensor defined by

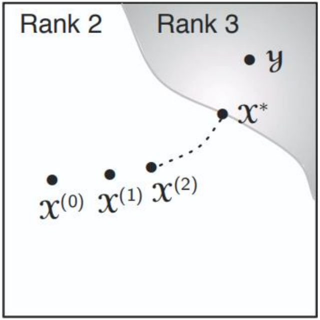
$$\mathcal{X} = a_1 \otimes b_1 \otimes c_2 + a_1 \otimes b_2 \otimes c_1 + a_2 \otimes b_1 \otimes c_1$$

can be approximated arbitrarily closely by a rank-two tensor of the following form:

$$\mathcal{Y} = \alpha(a_1 + \frac{1}{\alpha}a_2) \otimes (b_1 + \frac{1}{\alpha}b_2) \otimes (c_1 + \frac{1}{\alpha}c_2) - \alpha a_1 \otimes b_1 \otimes c_1.$$

Discontinuity of Tensor Rank III

- We say a tensor is **degenerate** if it may be approximated arbitrarily well by a factorization of low rank.
- The following picture shows a sequence $\{\mathcal{X}_k\}$ of rank-two tensors converging to a rank three tensor $\mathcal{X}^*$.



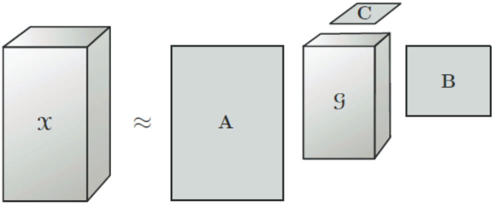
Outline

Introduction

Tucker Core

Tensor PCA

Tucker Decomposition I



$$\mathcal{X} = \mathcal{G} \times_1 A^{(1)} \times_2 \cdots \times_N A^{(N)} \triangleq [\![\mathcal{G}; A^{(1)}, \ldots, A^{(N)}]\!] \qquad (1)$$

Tucker Decomposition II

- $\mathcal{G} \in \mathbb{C}^{J_1 \times J_2 \times \cdots \times J_N}$ is called the *core tensor*.
- $A^{(n)} \in \mathbb{C}^{I_n \times J_n}$ is the n-th *factor matrix* for $n = 1, \ldots, N$, where $\times_n$ is the *mode-n (matrix) product*.
- Tucker decomposition is said to be independent if each of the factor matrix has full column rank; a Tucker decomposition is said to be orthonormal (more standard) if each of the factor matrix has orthonormal columns.

How to Compute Independent Tucker Decomposition I

Orthonormal Tucker Decomposition is named as higher-order SVD (HOSVD) in DeLathauwer, Moor and Vandewalle in 2000. Then how about independent Tucker decomposition.

- For each n one performs a mode-n matricization on $\mathcal{X}$ to get $X_{(1)}, \cdots, X_{(N)}$.
- For each mode-n one computes a matrix factorization such that $X_{(n)} = A^{(n)}C^{(n)}$ and $A^{(n)}$ has full column rank.
- Letting $\mathcal{G} = \mathcal{X} \times_1 (A^{(1)})^H \times_2 \cdots \times_N (A^{(N)})^H$ and $B^{(n)} = A^{(n)}\left((A^{(n)})^H A^{(n)}\right)^{-1}$, $[\![\mathcal{G}; B^{(1)}, \ldots, B^{(N)}]\!]$ is an independent Tucker decomposition of $\mathcal{X}$.

How to Compute Independent Tucker Decomposition II

Algorithm 1 HOSVD

1: **for** $n = 1, 2, \cdots, N$ **do**

2: $\quad A^{(n)} \leftarrow R_n$ leading left singular vectors of $X^{(n)}$

3: **end for**

4: $\mathcal{G} = \mathcal{X} \times_1 (A^{(1)})^H \times_2 \cdots \times_N (A^{(N)})^H$

5: **return** $\mathcal{G}, A^{(1)}, \cdots, A^{(N)}$

Are the two decompositions related ?

- For independent Tucker decomposition $\mathcal{X} = [\![\mathcal{G}; X^{(1)}, \ldots, X^{(N)}]\!]$,
- CP decomposition of the core $\mathcal{G} = \sum_{t=1}^{r} b^{(1,t)} \otimes b^{(2,t)} \otimes \cdots \otimes b^{(N,t)}$.
- Then $\mathcal{X} = \sum_{t=1}^{r} \left(A^{(1)} b^{(1,t)}\right) \otimes \cdots \otimes \left(A^{(N)} b^{(N,t)}\right)$.
- Could be very efficient when the size of the core $\mathcal{G}$ is much smaller than the original tensor $\mathcal{X}$.
- It is known as CANDELINC in the case of orthornormal Tucker decomposition.
  - Based on: "Tensor and Its Tucker Core: the Invariance Relationships", with F. Yang and S. Zhang, Numerical Linear Algebra with Applications, 24(3), e2086, 2017.

Rank Equivalence

- On the other hand, $\mathcal{G} = \mathcal{X} \times_1 B^{(1)} \times_2 \cdots \times_N B^{(N)}$ and $\mathcal{G} \in \mathbb{C}^{R_1 \times R_2 \times \cdots \times R_N}$, where $B^{(n)} = (A^{(n)H} A^{(n)})^{-1} A^{(n)H}$.
- Suppose $\mathcal{X} = \sum_{t=1}^{r} a^{(1,t)} \otimes a^{(2,t)} \otimes \cdots \otimes a^{(N,t)}$, similar argument yields that

$$\mathcal{G} = \sum_{t=1}^{r} \left(B^{(1)} a^{(1,t)}\right) \otimes \cdots \otimes \left(B^{(N)} a^{(N,t)}\right)$$

- Recall that $\mathcal{X} = \sum_{t=1}^{r} \left(A^{(1)} b^{(1,t)}\right) \otimes \cdots \otimes \left(A^{(N)} b^{(N,t)}\right)$.
- We have $\text{rank}_{CP}(\mathcal{X}) = \text{rank}_{CP}(\mathcal{G})$.
- The goal of matrix completion problem:

$$\min_{\mathcal{X}} \quad \sum_{i=1}^{N} \|X^{(i)}\|_* + \frac{\mu}{2} \|\boldsymbol{L}(\mathcal{X}) - b\|_2^2$$

Tensors with a symmetric structure

- *Symmetric CP decomposition*: $\mathcal{X} = \sum_{t=1}^{r} a^{(t)} \otimes \cdots \otimes a^{(t)}$. The *symmetric rank* of $\mathcal{X}$, denoted by $\mathrm{rank}_S(\mathcal{X})$, is the minimal integer r such that a size-r symmetric CP decomposition exists.

- *Symmetric Tucker decomposition*: $\mathcal{X} = [\![\mathcal{G}^s; X, X, \ldots, X]\!]$, and $\mathcal{G}^s$ is also symmetric.

- We still have that $\mathrm{rank}_S(\mathcal{G}^s) = \mathrm{rank}_S(\mathcal{X})$.

- In particular, we can first decompose $\mathcal{G}^s = \sum_{t=1}^{r} \underbrace{b^t \otimes \cdots \otimes b^t}_{N}$ and then

$$\mathcal{X} = \sum_{t=1}^{r} \underbrace{\left(Xb^t\right) \otimes \cdots \otimes \left(Xb^t\right)}_{N}.$$

Frobenius norm

- $\|\mathcal{X}\|_F \triangleq \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$, where $\langle \cdot, \cdot \rangle$ stands for the inner product of two tensors.

- For any given tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ with independent Tucker decomposition $\mathcal{X} = [\![\mathcal{G}; X^{(1)}, \ldots, X^{(N)}]\!]$, we have that

$$\alpha \|\mathcal{G}\|_F \leq \|\mathcal{X}\|_F \leq \beta \|\mathcal{G}\|_F, \tag{2}$$

  where $\beta = \|X^{(1)}\|_2 \|X^{(2)}\|_2 \cdots \|X^{(N)}\|_2$ and $\alpha = \beta / (\prod_{n=1}^{N} \kappa(X^{(n)}))$ with $\kappa(X^{(n)})$ being the condition number of the matrix $X^{(n)}$ for all n.

- For an orthonormal Tucker decomposition, its factor matrix $X^{(n)}$ is orthonormal. Thus $\|X^{(n)}\|_2 = \|(X^{(n)})^{\mathrm{H}}\|_2 = 1$ for $n = 1, 2, \ldots, N$, and we have $\|\mathcal{X}\|_F = \|\mathcal{G}\|_F$.

The quasi-p norm and the tensor nuclear norm I

- For any tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$, the *tensor p-quasi norm* for $0 < p \leq 1$

$$\|\mathcal{X}\|_p \triangleq \min \left\{ \left(\sum_{s=1}^{r} |\lambda_s|^p \right)^{1/p} : \mathcal{X} = \sum_{s=1}^{r} \lambda_s x_s^{(1)} \otimes x_s^{(2)} \otimes \cdots \otimes x_s^{(N)}, \right.$$
$$\left. \|x_s^{(n)}\| = 1, \ \forall s = 1, 2, \ldots, r, \ \forall n = 1, 2, \ldots, N \right\}. \tag{3}$$

- When $p = 1$, the above definition corresponds to the tensor nuclear norm, which was shown to be NP hard (Friedland and Lim).
- Recently, this tensor nuclear norm was applied by Yuan and Zhang (2015) to establish a better statistical properties of Tensor completion problem.

The quasi-p norm and the tensor nuclear norm II

- $\|\mathcal{X}\|_p$ trivially equals to zero for any tensor $\mathcal{X}$ for any $p > 1$ (from Lim).
- For any given tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ with independent Tucker decomposition $\mathcal{X} = [\![\mathcal{G}; X^{(1)}, \ldots, X^{(N)}]\!]$, we have

$$\alpha\|\mathcal{G}\|_p \leq \|\mathcal{X}\|_p \leq \beta\|\mathcal{G}\|_p,$$

  where $\beta = \|X^{(1)}\|_2\|X^{(2)}\|_2 \ldots \|X^{(N)}\|_2$ and $\alpha = \beta/(\prod_{n=1}^{N} \kappa(X^{(n)}))$ with $\kappa(X^{(n)})$ being the condition number of matrix $X^{(n)}$ for all n.

- For an orthonormal Tucker decomposition, we have $\|\mathcal{X}\|_p = \|\mathcal{G}\|_p$.

Invariance of Z-eigenvalue

- A notation:

$$(\mathcal{T}(x^{\circ(N-1)}))_{i_N} \triangleq \sum_{i_1,i_2,\ldots,i_{N-1}=1}^{I} T_{i_1 i_2 \cdots i_{N-1} i_N} \cdot x_{i_1} x_{i_2} \cdots x_{i_{N-1}}.$$

- (Qi, Lim) For a symmetric tensor $\mathcal{T} \in \mathbb{R}^{I\times\cdots\times I}$, exists$\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^I$ such that

$$\mathcal{T}(x^{\circ(N-1)}) = \lambda x,\ x^{\mathrm{T}} x = 1. \tag{4}$$

  Then λ is called the *Z-eigenvalue* of $\mathcal{T}$, and x is called the corresponding *Z-eigenvector*.

- For $\mathcal{T} \in \mathbb{R}^{I\times I\times\cdots\times I}$ with independent symmetric Tucker decomposition $\mathcal{T} = [\![\mathcal{G}^s; X, X, \ldots, X]\!]$, construct

$$\hat{\mathcal{G}}^s = \mathcal{G}^s \times_1 (X^{\mathrm{T}}X)^{1/2} \times_2 \cdots \times_N (X^{\mathrm{T}}X)^{1/2}.$$

Invariance of Z-eigenvalue

- Any Z-eigenvalues of $\hat{\mathcal{G}}^s$ are also Z-eigenvalues of $\mathcal{T}$.
- Any non-zero Z-eigenvalue of $\mathcal{T}$ are also Z-eigenvalues of $\hat{\mathcal{G}}^s$.
- If the Tucker decomposition is orthonormal then $(X^{\mathrm{T}}X)^{1/2} = I$, $\hat{\mathcal{G}}^s = \mathcal{G}^s$, and all the Z-eigenvalues (except zero) of a symmetric tensor equal to the Z-eigenvalues of its core.
- $\mathcal{T} \in \mathbb{R}^{I \times I \times \cdots \times I}$ has an independent symmetric Tucker decomposition $\mathcal{T} = [\![\mathcal{G}^s; X, X, \ldots, X]\!]$ such that $\mathcal{G}^s \in \mathbb{R}^{J \times J \times \cdots \times J}$. If $I > J$, then 0 is an Z-eigenvalue.

Invariance of the M-eigenvalue

- (M-eigenvalue) $\mathcal{T} \in \mathbb{R}^{N\times M\times N\times M}$ partial symmetric, exist two numbers λ and $\mu \in \mathbb{R}$, two nonzero vectors $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$ such that

$$\mathcal{T}(\cdot, y, x, y) = \lambda x,\ x^{\mathrm{T}}x = 1$$
$$\mathcal{T}(x, y, x, \cdot) = \mu y,\ y^{\mathrm{T}}y = 1$$

- $\mathcal{T} = [\![\mathcal{G}^{ps}; A, B, A, B]\!]$, construct

$$\hat{\mathcal{G}}^{ps} = \times_1 (A^{\mathrm{T}}A)^{1/2} \times_2 (B^{\mathrm{T}}B)^{1/2} \times_3 (A^{\mathrm{T}}A)^{1/2} \times_4 (B^{\mathrm{T}}B)^{1/2}.$$

  Then any M-eigenvalues of $\hat{\mathcal{G}}^{ps}$ are also M-eigenvalues of $\mathcal{T}$ while any non-zero M-eigenvalues of $\mathcal{T}$ are also M-eigenvalues of $\hat{\mathcal{G}}^{ps}$.

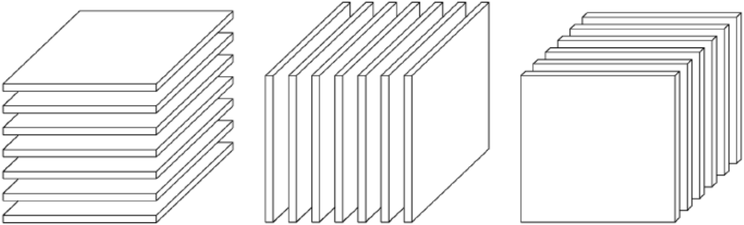
- $\mathcal{T} \in \mathbb{R}^{I_1\times I_2\times I_1\times I_2}$ has an independent partial symmetric Tucker decomposition $\mathcal{T} = [\![\mathcal{G}^{ps}; A, B, A, B]\!]$ such that $\mathcal{G}^{ps} \in \mathbb{R}^{J_1\times J_2\times J_1\times J_2}$. Either $J_1 < I_1$ or $J_2 < I_2$ implies the existence of a zero M-eigenvalue.

Outline

Tensor matricization / Matrix unfolding I

Consider the following tensor $\mathcal{A} \in \mathbb{R}^{3\times4\times2}$ with:

$$\mathcal{A}_{:,:,1} = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix} \quad \mathcal{A}_{:,:,2} = \begin{bmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{bmatrix}$$



(a) Horizontal slices: $\mathbf{X}_{i::}$ (b) Lateral slices: $\mathbf{X}_{:j:}$ (c) Frontal slices: $\mathbf{X}_{::k}$ (or $\mathbf{X}_k$)

Tensor matricization / Matrix unfolding II

- The mode-k matricization of $\mathcal{A} \in \mathbb{C}^{n_1 \times \cdots \times n_d}$ (denoted by $A_{(k)}$):

$$(A_{(k)})_{i_k \ell} = \mathcal{A}_{i_1 \cdots i_d}, \quad \ell = \sum_{j=2, j \neq k}^{d} (i_j - 1) \prod_{q=1, q \neq k}^{j-1} n_q + i_1.$$

- The square unfolding for $\mathcal{F} \in \mathbb{C}^{n_1 \times n_2 \cdots \times n_{2d}}$ (denoted by $\boldsymbol{M}(\mathcal{F})$):

$$\boldsymbol{M}(\mathcal{F})_{k\ell} := \mathcal{F}_{i_1 \cdots i_{2d}},$$

$$\text{where } k = \sum_{j=2}^{d} (i_j - 1) \prod_{q=1}^{j-1} n_q + i_1, \ell = \sum_{j=d+2}^{2d} (i_j - 1) \prod_{q=d+1}^{j-1} n_q + i_{d+1}.$$

Tensor matricization / Matrix unfolding III

- The way of square unfolding is not unique.
- Three options for $\mathcal{F} \in \mathbb{C}^{n_1 \times n_2 \times n_3 \times n_4}$:
  - $\boldsymbol{M}(\mathcal{F}) \in \mathbb{C}^{(n_1 n_2) \times (n_3 n_4)}$;
  - $\boldsymbol{M}(\mathcal{F}_{\pi_1}) \in \mathbb{C}^{(n_1 n_3) \times (n_2 n_4)}$;
  - $\boldsymbol{M}(\mathcal{F}_{\pi_2}) \in \mathbb{C}^{(n_1 n_4) \times (n_2 n_3)}$;

Tensor PCA I

- Find the leading principle component of tensor $\mathcal{F}$:

$$\begin{array}{ll} \max & \mathcal{F}(x, \cdots, x) := \sum_{1 \le i_1, \cdots, i_d \le n} \mathcal{F}_{i_1 \cdots i_d} x_{i_1} \cdots x_{i_d} \\ \text{s.t.} & \|x\| = 1, \end{array}$$

  - Based on: "Tensor Principal Component Analysis via Convex Optimization", with S. Ma and S. Zhang, Mathematical Programming Series A, 150, 423-457, 2015.

Tensor PCA II

- Consider the even order tensor: recall that

$$\mathcal{F}(\underbrace{x, \cdots, x}_{2d}) = \mathcal{F} \bullet \underbrace{x \otimes \cdots \otimes x}_{2d}.$$

By letting

$$\mathcal{X} = \underbrace{x \otimes \cdots \otimes x}_{2d},$$

the tensor PCA problem can be reformulated as:

$$\begin{array}{ll} \max & \mathcal{F} \bullet \mathcal{X} \\ \text{s.t.} & \sum\limits_{k \in \mathbb{K}(n,d)} \frac{d!}{\prod_{j=1}^{n} k_j!} \mathcal{X}_{1^{2k_1} 2^{2k_2} \cdots n^{2k_n}} = 1, \\ & \mathcal{X} \in \mathbf{S}^{n^{2d}}, \ \operatorname{rank}_{RCP}(\mathcal{X}) = 1, \end{array}$$

where the equality due to $\sum\limits_{k \in \mathbb{K}(n,d)} \frac{d!}{\prod_{j=1}^{n} k_j!} \prod_{j=1}^{n} x_j^{2k_j} = \|x\|^{2d} = 1.$

Tensor PCA III

- Denote $X = \boldsymbol{M}(\mathcal{X}), F = \boldsymbol{M}(\mathcal{F})$. Then the objective is $\mathcal{F} \bullet \mathcal{X} = \operatorname{tr}(FX)$.

$$
\begin{array}{ll}
\max & \operatorname{tr}(FX) \\
\text{s.t.} & \operatorname{tr}(X) = 1,\ \boldsymbol{M}^{-1}(X) \in \mathbf{S}^{n^{2d}}, \\
& X \in \mathbf{S}^{n^d \times n^d},\ \operatorname{rank}(X) = 1,
\end{array}
$$

We convert a tensor rank-constrained problem into a matrix rank-constrained problem.

A Nuclear Norm Penalty Approach

- The constraint $\mathrm{rank}(X) = 1$ is still hard.
- $\|X\|_* := \mathrm{tr}\left((X^\top X)^{\frac{1}{2}}\right)$, which is the good convex approximation of $\mathrm{rank}(X)$.
- Nuclear norm penalty formulation

$$\begin{array}{ll} \max & \mathrm{tr}\,(FX) - \rho\|X\|_* \\ \text{s.t.} & \mathrm{tr}\,(X) = 1,\ \boldsymbol{M}^{-1}(X) \in \mathbf{S}^{n^{2d}}, \\ & X \in \mathbf{S}^{n^d \times n^d}, \end{array}$$

- If the solution to the above problem is rank-one, then this solution is also optimal to the original problem.

Rank-One Frequency of Nuclear Norm Penalty Formulation

Test results using cvx for randomly generated instances when $d = 2$ and $\rho = 10$.

Dim	Num	Ratio(%)	AT(Seconds)
3	1000	100	0.1649
4	1000	100	0.3667
5	1000	100	1.0886
6	1000	100	5.7045
7	300	100	37.1881
8	100	100	167.2816

An Alternating Direction Method of Multipliers (ADMM) Implementation

$$\begin{array}{ll} \min & -\mathrm{tr}\,(FY) + \rho\|Y\|_* \\ \text{s.t.} & X - Y = 0, \\ & \mathrm{tr}\,(X) = 1,\ \boldsymbol{M}^{-1}(X) \in \mathbf{S}^{n^{2d}}. \end{array}$$

ADMM procedure:

$$\left\{ \begin{array}{lll} X^{k+1} & = & \mathbf{argmin}_{\mathrm{tr}\,(X)=1,\ \boldsymbol{M}^{-1}(X)\in\mathbf{S}^{n^{2d}}}\ \mathcal{L}(X, Y^k; \Lambda^k) \\ Y^{k+1} & = & \mathrm{argmin}\ \mathcal{L}(X^{k+1}, Y; \Lambda^k) \\ \Lambda^{k+1} & = & \Lambda^k - (X^{k+1} - Y^{k+1})/\mu, \end{array} \right.$$

where the augmented Lagrangian function is

$$\mathcal{L}(X, Y; \Lambda) := -\mathrm{tr}\,(FY) + \rho\|Y\|_* - \mathrm{tr}\,(\Lambda(X - Y)) + \frac{1}{2\mu}\|X - Y\|_F^2.$$

Both two subproblems can be easily (explicitly) solved.

Numerical Performance of ADMM

Performance of CVX and ADMM when $d = 2$, $n = 6$, $\mu = 0.1$, and $\rho = 10$.

Inst. #	Difference		ADMM		CVX
	$\|X_{ADMM} - X_{CVX}\|_F$	$\lvert v_{ADMM} - v_{CVX}\rvert$	Iter #	Time	Time
1	3.949037e-005	2.653805e-006	258	0.109	1.076
2	5.218683e-005	8.584708e-008	252	0.094	0.421
3	2.638832e-005	1.578413e-007	253	0.094	0.406
4	7.813928e-005	2.650758e-006	231	0.094	0.468
5	3.220476e-005	6.379380e-007	229	0.078	0.359
6	9.221593e-005	1.957860e-007	231	0.078	0.359
7	5.094511e-005	3.025078e-006	249	0.078	0.390
8	1.014394e-004	3.282918e-007	248	0.078	0.421
9	1.675034e-005	2.524454e-006	245	0.078	0.406
10	2.557981e-005	4.178700e-007	242	0.078	0.390

SDP Relaxation

What happens when $\rho \to \infty$?

Consider the following SDP relaxation:

$$
\begin{array}{lll}
(SDR) & \max & \mathrm{tr}\,(FX) \\
& \text{s.t.} & \mathrm{tr}\,(X) = 1, \\
& & \boldsymbol{M}^{-1}(X) \in \mathbf{S}^{n^{2d}},\ X \succeq 0.
\end{array}
$$

- Denote X^*_{SDR} to be the solution of (SDR).
- Denote $X^*_{PNP}(\rho)$ to be the solution of the nuclear norm penalty formulation with parameter ρ.

Then it holds that:

$$
\lim_{\rho \to +\infty} \mathrm{tr}\,(FX^*_{PNP}(\rho)) = \mathrm{tr}\,(FX^*_{SDR})
$$