CVPR 2017

A New Tensor Algebra - Tutorial

Lior Horesh
lhoresh@us.ibm.com

Misha Kilmer
misha.kilmer@tufts.edu

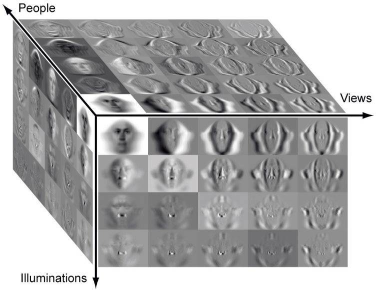
July 26, 2017

Outline

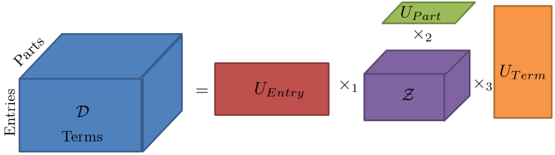
- Motivation
- Background and notation
- New t-product and associated algebraic framework
- Implementation considerations
- The t-SVD and optimality
  - Application in Facial Recognition
- Proper Orthogonal Decomposition
  - - Dynamic Model Redcution
- A tensor Nuclear Norm from the t-SVD
  - Applications in video completion

Tensor Applications:

- **Machine vision**: understanding the world in 3D, enable understanding phenomena such as perspective, occlusions, illumination



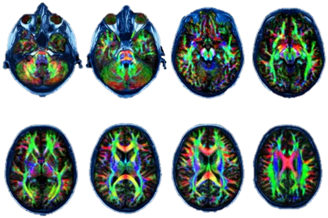
- **Latent semantic tensor indexing**: common terms vs. entries vs. parts, co-occurrence of terms



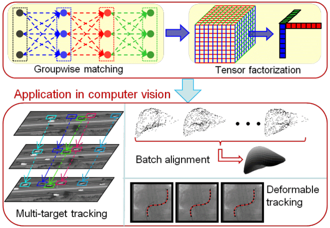
Tensor subspace Analysis for Viewpoint Recognition, T. Ivanov, L. Mathies, M.A.O. Vasilescu, ICCV, 2nd IEEE International Workshop on Subspace Methods, September, 2009

Tensor Applications:

- **Medical imaging**: naturally involves 3D (spatio) and 4D (spatio-temporal) correlations



- **Video surveillance and Motion signature**: 2D images + 3^{rd} dimension of time, 3D/4D motion trajectory

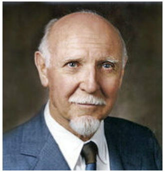


Multi-target Tracking with Motion Context in Tenor Power Iteration X. Shi, H. Ling, W. Hu, C. Yuan, and J. Xing IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Columbus OH, 2014

Tensors: Historical Review

- 1927 F.L. Hitchcock: "The expression of a tensor or a polyadic as a sum of products" (Journal of Mathematics and Physics)
- 1944 R.B. Cattell introduced a multiway model: "Parallel proportional profiles and other principles for determining the choice of factors by rotation" (Psychometrika)
- 1960 L.R. Tucker: "Some mathematical notes on three-mode factor analysis" (Psychometrika)
- 1981 tensor decomposition was first used in chemometrics
- Past decade, computer vision, image processing, data mining, graph analysis, etc.

F.L. Hitchcock

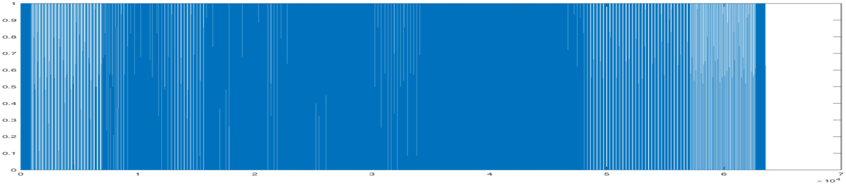


R.B. Cattell

L.R. Tucker

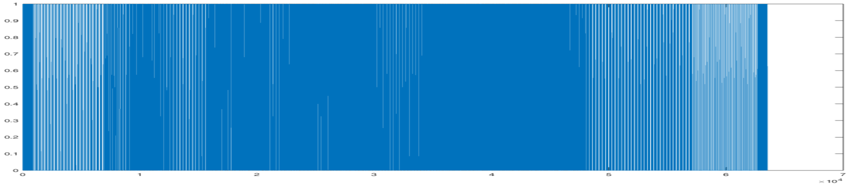
The Power of Proper Representation

- What is that ?



The Power of Proper Representation

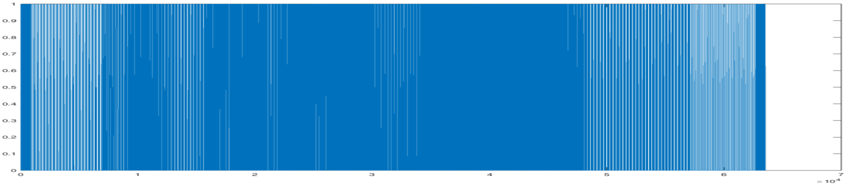
- What is that ?



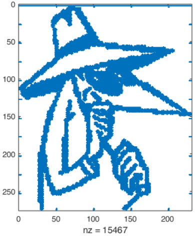
- Let's observe the same data but in a different (matrix rather than vector) representation

The Power of Proper Representation

- What is that ?



- Let's observe the same data but in a different (matrix rather than vector) representation



- **Representation matters!** some correlations can only be realized in appropriate representation

Motivation

- Much real-world **data** is inherently multidimensional
  - color video data – 4 way
  - 3D medical image, evolving in time (4 way); multiple patients (5 way)
- Many **operators** and **models** are also multi-way
- Traditional matrix-based methods based on data vectorization (e.g. matrix PCA) generally **agnostic** to possible high dimensional correlations

Can we **uncover hidden patterns** in tensor data by computing an appropriate tensor decomposition/approximation?

Need to decide on the tensor decomposition – **application dependent**!

What do we mean by **‘decompose’**?

Tensors: Background and Notation

- **Notation** : $\mathcal{A}^{n_1 \times n_2 \ldots, \times n_j}$ - j^{th} order tensor
- Examples

Tensors: Background and Notation

- **Notation** : $\mathcal{A}^{n_1 \times n_2 \ldots, \times n_j}$ - j^{th} order tensor
- Examples
  - 0^{th} order tensor

Tensors: Background and Notation

- **Notation** : $\mathcal{A}^{n_1 \times n_2 \ldots, \times n_j}$ - j^{th} order tensor
- Examples
  - 0^{th} order tensor - scalar

Tensors: Background and Notation

- **Notation** : $\mathcal{A}^{n_1 \times n_2 \ldots, \times n_j}$ - j^{th} order tensor
- Examples
  - 0^{th} order tensor - scalar

  - 1^{st} order tensor

Tensors: Background and Notation

- **Notation** : $\mathcal{A}^{n_1 \times n_2 \ldots, \times n_j}$ - j^{th} order tensor
- Examples
  - 0^{th} order tensor - scalar

  - 1^{st} order tensor - vector

Tensors: Background and Notation

- **Notation** : $\mathcal{A}^{n_1 \times n_2 \ldots, \times n_j}$ - j^{th} order tensor
- Examples
  - 0^{th} order tensor - scalar

  - 1^{st} order tensor - vector

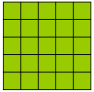
  - 2^{nd} order tensor

Tensors: Background and Notation

- **Notation** : $\mathcal{A}^{n_1 \times n_2 \ldots, \times n_j}$ - j^{th} order tensor
- Examples
  - 0^{th} order tensor - scalar

  - 1^{st} order tensor - vector

  - 2^{nd} order tensor - matrix

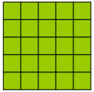


Tensors: Background and Notation

- **Notation** : $\mathcal{A}^{n_1 \times n_2 \ldots, \times n_j}$ - j^{th} order tensor
- Examples
  - 0^{th} order tensor - scalar

  - 1^{st} order tensor - vector

  - 2^{nd} order tensor - matrix



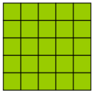
  - 3^{rd} order tensor ...

Tensors: Background and Notation

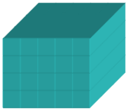
- **Notation** : $\mathcal{A}^{n_1 \times n_2 \ldots, \times n_j}$ - j^{th} order tensor
- Examples
  - 0^{th} order tensor - scalar

  - 1^{st} order tensor - vector

  - 2^{nd} order tensor - matrix

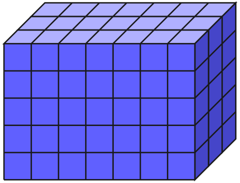


  - 3^{rd} order tensor ...



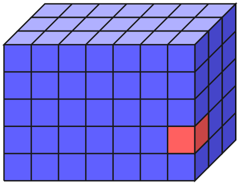
Notation

$\mathcal{A}_{i,j,k}$ = element of $\mathcal{A}$ in row i, column j, tube k



Notation

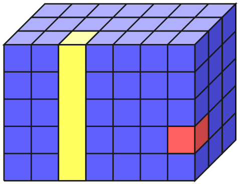
$\mathcal{A}_{i,j,k}$ = element of $\mathcal{A}$ in row i, column j, tube k



$\leftarrow \mathcal{A}_{4,7,1}$

Notation

$$\mathcal{A}_{i,j,k} = \text{element of } \mathcal{A} \text{ in row } i, \text{ column } j, \text{ tube } k$$

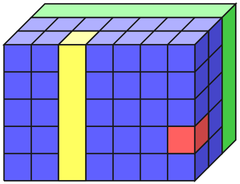


$\leftarrow \mathcal{A}_{4,7,1}$

$\leftarrow \mathcal{A}_{:,3,1}$

Notation

$$\mathcal{A}_{i,j,k} = \text{element of } \mathcal{A} \text{ in row } i, \text{ column } j, \text{ tube } k$$



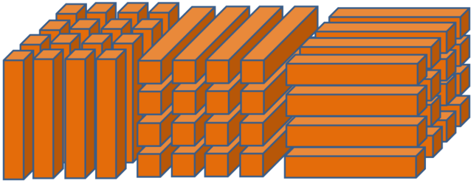
$\leftarrow \mathcal{A}_{4,7,1}$

$\leftarrow \mathcal{A}_{:,3,1}$

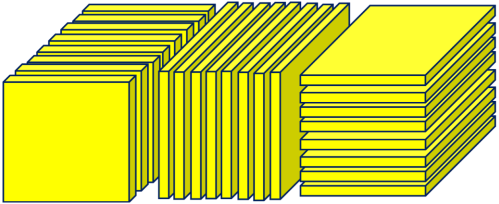
$\leftarrow \mathcal{A}_{:,:,3}$

Tensors: Background and Notation

- *Fiber* - a vector defined by fixing all but one index while varying the rest



- *Slice* - a matrix defined by fixing all but two indices while varying the rest



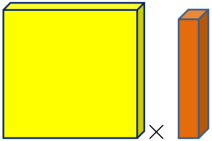
Tensor Multiplication

- **Definition** : The k - mode multiplication of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \ldots, \times n_d}$ with a matrix $U \in \mathbb{R}^{J \times n_k}$ is denoted by $\mathcal{X} \times_k U$ and is of size $n_1 \times \cdots \times n_{k-1} \times J \times n_{k+1} \times \cdots \times n_d$

- Element-wise

$$(\mathcal{X} \times_k U)_{i_1 \cdots i_{k-1} j i_{k+1} \cdots i_d} = \sum_{i_k=1}^{n_d} x_{i_1 i_2 \cdots i_d} u_{j i_k}$$

- 1-mode multiplication



Tensor Holy Grail and the Matrix Analogy

Find a way to express a tensor that leads to the possibility for **compressed representation** (near redundancy removed) that **maintains important features** of the original tensor

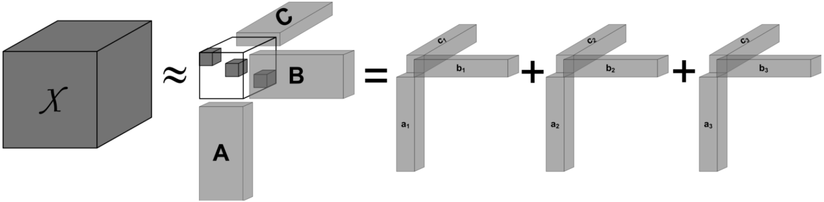
---

$$\min \|\mathbf{A} - \mathbf{B}\|_F \qquad \text{s.t. } \mathbf{B} \text{ has rank } p \leq r$$

$\mathbf{B} = \sum_{i=1}^{p} \sigma_i ({}_V u^{(i)} \circ_V v^{(i)})$ where $\mathbf{A} = \sum_{i=1}^{r} \sigma_i (\mathbf{u}^{(i)} \circ \mathbf{v}^{(i)})$

Tensor Decompositions - CP

- **CP (CANDECOMP-PARAFAC) Decomposition** [1] :



$$\mathcal{X} \approx \sum_{i=1}^{r} a_i \circ b_i \circ c_i$$

- Outer product

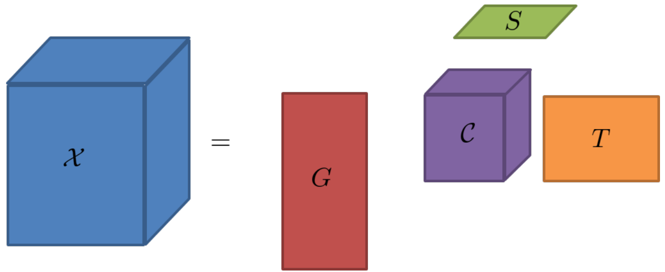
$$\mathcal{T} = u \circ v \circ w \Rightarrow \mathcal{T}_{ijk} = u_i v_j w_k$$

  - Columns of $A = [a_1, \ldots, a_r], B = [b_1, \ldots, b_r], C = [c_1, \ldots, c_r]$ are not orthogonal
  - If r is minimal, then r is called the *rank* of the tensor
  - No perfect procedure for fitting CP for a given number of components [2]

[1] R. Harshman, 1970; J. Carroll and J. Chang, 1970

[2] V. de Silva, L. Lim, *Tensor Rank and the Ill-Posedness of the Best Low-Rank Approximation Problem, 2008*

- **Tucker Decomposition** :



$$\mathcal{X} \approx \mathcal{C} \times_1 G \times_2 T \times_3 S = \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \sum_{k=1}^{r_3} c_{ijk} g_i \circ t_j \circ s_k$$

- $\mathcal{C}$ is the *core* tensor
- G, T, S are the *components* of factors
- Can either have diagonal core or orthogonal columns in components [DeLathauwer et al.]
- Truncated Tucker decomposition is not optimal in approximating the norm of the difference

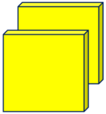
$$\|\mathcal{X} - \mathcal{C} \times_1 G \times_2 T \times_3 S\|$$

- **t-product** : Let $\mathcal{A}$ be $n_1 \times n_2 \times n_3$ and $\mathcal{B}$ be $n_2 \times \ell \times n_3$. Then the *t-product* $\mathcal{A} * \mathcal{B}$ is the $n_1 \times \ell \times n_3$ tensor

$$\mathcal{A} * \mathcal{B} = fold(circ(\mathcal{A}) \cdot vec(\mathcal{B}))$$

- $$circ(\mathcal{A}) \cdot vec(\mathcal{B}) = \begin{pmatrix} \mathcal{A}_1 & \mathcal{A}_{n_3} & \mathcal{A}_{n_3-1} & \cdots & \mathcal{A}_2 \\ \mathcal{A}_2 & \mathcal{A}_1 & \mathcal{A}_{n_3} & \cdots & \mathcal{A}_3 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathcal{A}_{n_3-1} & \mathcal{A}_{n_3-2} & \mathcal{A}_{n_3-3} & \cdots & \mathcal{A}_{n_3} \\ \mathcal{A}_{n_3} & \mathcal{A}_{n_3-1} & \mathcal{A}_{n_3-2} & \cdots & \mathcal{A}_1 \end{pmatrix} \begin{pmatrix} \mathcal{B}_1 \\ \mathcal{B}_2 \\ \mathcal{B}_3 \\ \vdots \\ \mathcal{B}_{n_3} \end{pmatrix}$$

- $fold(vec(\mathcal{B})) = \mathcal{B}$



- $\mathcal{A}_i$, $\mathcal{B}_i$, $i = 1, \ldots, n_3$ are frontal slices of $\mathcal{A}$ and $\mathcal{B}$

M.E. Kilmer and C.D. Martin. Factorization strategies for third-order tensors, *Linear Algebra and its Applications*, Special Issue in Honor of G. W. Stewart's 70^{th} birthday, vol. 435(3):641–658, 2011

Block Circulants

- A block circulant can be block-diagonalized by a (normalized) DFT in the 2^{nd} dimension:

$$(\mathbf{F} \otimes \mathbf{I})\text{circ}\,(\boldsymbol{\mathcal{A}})\,(\mathbf{F}^* \otimes \mathbf{I}) = \begin{bmatrix} \hat{\mathbf{A}}_1 & 0 & \cdots & 0 \\ 0 & \hat{\mathbf{A}}_2 & 0 & \cdots \\ 0 & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & \hat{\mathbf{A}}_n \end{bmatrix}$$

- Here $\otimes$ is a Kronecker product of matrices
- If $\mathbf{F}$ is $n \times n$, and $\mathbf{I}$ is $m \times m$, $(\mathbf{F} \otimes \mathbf{I})$ is the $mn \times mn$ block matrix, of n block rows and columns, each block is $m \times m$, where the ij^{th} block is $f_{i,j}\mathbf{I}$
- But we never implement it this way because an FFT along tube fibers of $\boldsymbol{\mathcal{A}}$ yields a tensor, $\hat{\boldsymbol{\mathcal{A}}}$ whose frontal slices are the $\hat{\mathbf{A}}_i$

t-product Identity

- **Definition**: The $n \times n \times \ell$ *identity* tensor $\mathcal{I}_{nn\ell}$ is the tensor whose frontal face is the $n \times n$ identity matrix, and whose other faces are all zeros

---

t-product Identity

- **Definition**: The $n \times n \times \ell$ *identity* tensor $\mathcal{I}_{nn\ell}$ is the tensor whose frontal face is the $n \times n$ identity matrix, and whose other faces are all zeros

- Class Exercise: Let $\mathcal{A}$ be $n_1 \times n \times n_3$, show that

$$\mathcal{A} * \mathcal{I} = \mathcal{A} \quad \text{and} \quad \mathcal{I} * \mathcal{A} = \mathcal{A}$$

t-product Identity

- **Definition**: The $n \times n \times \ell$ *identity* tensor $\mathcal{I}_{nn\ell}$ is the tensor whose frontal face is the $n \times n$ identity matrix, and whose other faces are all zeros

---

- Class Exercise: Let $\mathcal{A}$ be $n_1 \times n \times n_3$, show that

$$\mathcal{A} * \mathcal{I} = \mathcal{A} \quad \text{and} \quad \mathcal{I} * \mathcal{A} = \mathcal{A}$$

$\mathcal{A} * \mathcal{I}$

t-product Identity

- **Definition**: The $n \times n \times \ell$ *identity* tensor $\mathcal{I}_{nn\ell}$ is the tensor whose frontal face is the $n \times n$ identity matrix, and whose other faces are all zeros

- Class Exercise: Let $\mathcal{A}$ be $n_1 \times n \times n_3$, show that

$$\mathcal{A} * \mathcal{I} = \mathcal{A} \quad \text{and} \quad \mathcal{I} * \mathcal{A} = \mathcal{A}$$

$$\mathcal{A} * \mathcal{I} = fold\left(circ\left(\mathcal{A}\right) \cdot vec\left(\mathcal{I}\right)\right)$$

t-product Identity

- **Definition**: The $n \times n \times \ell$ *identity* tensor $\mathcal{I}_{nn\ell}$ is the tensor whose frontal face is the $n \times n$ identity matrix, and whose other faces are all zeros

---

- Class Exercise: Let $\mathcal{A}$ be $n_1 \times n \times n_3$, show that

$$\mathcal{A} * \mathcal{I} = \mathcal{A} \quad \text{and} \quad \mathcal{I} * \mathcal{A} = \mathcal{A}$$

$$\mathcal{A} * \mathcal{I} = fold\,(circ\,(\mathcal{A}) \cdot vec\,(\mathcal{I})) = \begin{pmatrix} \mathcal{A}_1 & \mathcal{A}_{n_3} & \mathcal{A}_{n_3-1} & \cdots & \mathcal{A}_2 \\ \mathcal{A}_2 & \mathcal{A}_1 & \mathcal{A}_{n_3} & \cdots & \mathcal{A}_3 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathcal{A}_{n_3-1} & \mathcal{A}_{n_3-2} & \mathcal{A}_{n_3-3} & \cdots & \mathcal{A}_{n_3} \\ \mathcal{A}_{n_3} & \mathcal{A}_{n_3-1} & \mathcal{A}_{n_3-2} & \cdots & \mathcal{A}_1 \end{pmatrix} \begin{pmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

t-product Identity

- **Definition**: The $n \times n \times \ell$ *identity* tensor $\mathcal{I}_{nn\ell}$ is the tensor whose frontal face is the $n \times n$ identity matrix, and whose other faces are all zeros

- Class Exercise: Let $\mathcal{A}$ be $n_1 \times n \times n_3$, show that

$$\mathcal{A} * \mathcal{I} = \mathcal{A} \quad \text{and} \quad \mathcal{I} * \mathcal{A} = \mathcal{A}$$

$$\mathcal{A} * \mathcal{I} = fold\,(circ\,(\mathcal{A}) \cdot vec\,(\mathcal{I})) = \begin{pmatrix} \mathcal{A}_1 & \mathcal{A}_{n_3} & \mathcal{A}_{n_3-1} & \cdots & \mathcal{A}_2 \\ \mathcal{A}_2 & \mathcal{A}_1 & \mathcal{A}_{n_3} & \cdots & \mathcal{A}_3 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathcal{A}_{n_3-1} & \mathcal{A}_{n_3-2} & \mathcal{A}_{n_3-3} & \cdots & \mathcal{A}_{n_3} \\ \mathcal{A}_{n_3} & \mathcal{A}_{n_3-1} & \mathcal{A}_{n_3-2} & \cdots & \mathcal{A}_1 \end{pmatrix} \begin{pmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathcal{A}$$

t-product Transpose

- **Definition**: If $\mathcal{A}$ is $n_1 \times n_2 \times n_3$, then $\mathcal{A}^\top$ is the $n_2 \times n_1 \times n_3$ tensor obtained by transposing each of the frontal faces and then reversing the order of transposed faces 2 through n_3

- Example: If $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times 4}$ and its frontal faces are given by the $n_1 \times n_2$ matrices $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4$, then

$$\mathcal{A}^\top = fold\left(\begin{bmatrix} \mathcal{A}_1^\top \\ \mathcal{A}_4^\top \\ \mathcal{A}_3^\top \\ \mathcal{A}_2^\top \end{bmatrix}\right)$$

- **Mimetic property**: when $n = 1$, the $*$ operator collapses to traditional matrix multiplication between two matrices and tranpose becomes matrix transposition

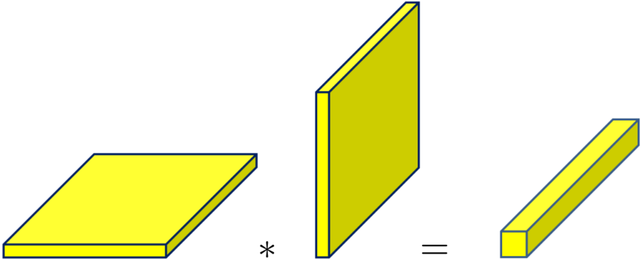
t-product Orthogonality

- **Definition**: An $n \times n \times l$ real-valued tensor $\mathcal{Q}$ is orthogonal if

$$Q^\top * Q = Q * Q^\top = \mathcal{I}$$

- Note that this means that

$$\mathcal{Q}(:,i,:)^\top * \mathcal{Q}(:,j,:) = \begin{cases} e_1 & i = j \\ 0 & i \neq j \end{cases}$$



t-SVD and Trunction Optimality

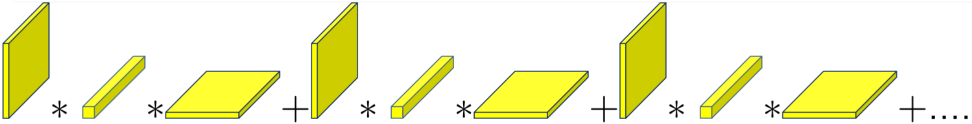
- **Theorem**: Let the $\mathcal{T}$-SVD of $\mathcal{A} \in \mathbb{R}^{\ell \times m \times n}$ be given by $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, with $\ell \times \ell \times n$ orthogonal tensor $\mathcal{U}$, $m \times m \times n$ orthogonal tensor $\mathcal{V}$, and $\ell \times m \times n$ f-diagonal tensor $\mathcal{S}$
  - For $k < \min(l, m)$, define

$$\mathcal{A}_k = \mathcal{U}(:, 1:k, :) * \mathcal{S}(1:k, 1:k, :) * \mathcal{V}^\top(:, 1:k, :) = \sum_{i=1}^{k} \mathcal{U}(:, i, :) * \mathcal{S}(i, i, :) * \mathcal{V}(:, i, :)^\top$$

  - Then

$$\mathcal{A}_k = \underset{\hat{\mathcal{A}} \in M}{\arg\min} \|\mathcal{A} - \hat{\mathcal{A}}\|$$

where $M = \{\mathcal{C} = \mathcal{X} * \mathcal{Y} \mid \mathcal{X} \in \mathbb{R}^{\ell \times k \times n}, \mathcal{Y} \in \mathbb{R}^{k \times m \times n}\}$



t-SVD and Optimality in Truncation

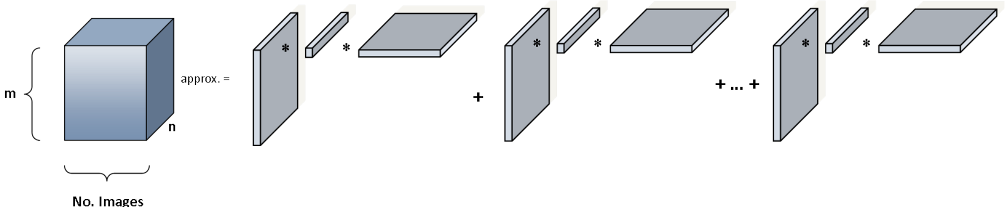
- Let $\mathcal{A} \in \mathbb{R}^{m \times p \times n}$, for $k < \min(m, p)$, define

$$\mathcal{A}_k = \sum_{i=1}^{k} \mathcal{U}(:, i, :) * \mathcal{S}(i, i, :) * \mathcal{V}(:, i, :)^\top$$

- Then

$$\mathcal{A}_k = \underset{\widetilde{\mathcal{A}} \in M}{\arg\min} \|\mathcal{A} - \widetilde{\mathcal{A}}\|$$

where $M = \{\mathcal{C} = \mathcal{X} * \mathcal{Y} \;|\mathcal{X} \in \mathbb{R}^{m \times k \times n}, \mathcal{Y} \in \mathbb{R}^{k \times p \times n}\}$



t-SVD example

- Let $\mathcal{A}$ be $2 \times 2 \times 2$

$$(\mathbf{F} \otimes \mathbf{I})\text{circ}\,(\mathcal{A})\,(\mathbf{F}^* \otimes \mathbf{I}) = \begin{bmatrix} \hat{\mathbf{A}}_1 & 0 \\ 0 & \hat{\mathbf{A}}_2 \end{bmatrix} \in \mathbb{C}^{4\times 4}$$

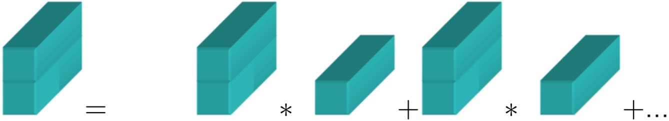
$$\begin{bmatrix} \hat{\mathbf{A}}_1 & 0 \\ 0 & \hat{\mathbf{A}}_2 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{U}}_1 & 0 \\ 0 & \hat{\mathbf{U}}_2 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \hat{\sigma}_{\color{red}1}^{(1)} & 0 \\ 0 & \hat{\sigma}_2^{(1)} \end{bmatrix} & \\ & \begin{bmatrix} \hat{\sigma}_{\color{red}1}^{(2)} & 0 \\ 0 & \hat{\sigma}_2^{(2)} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{V}}_1^* & 0 \\ 0 & \hat{\mathbf{V}}_2^* \end{bmatrix}$$

- The $\mathcal{U}, \mathcal{S}, \mathcal{V}^T$ are formed by putting the hat matrices as frontal slices, then ifft along tubes
- e.g. $\mathcal{S}_{(1,1,:)}$ obtained from ifft of vector $\begin{bmatrix} \hat{\sigma}_{\color{red}1}^{(1)} \\ \hat{\sigma}_{\color{red}1}^{(2)} \end{bmatrix}$ oriented into screen

$\mathcal{T}$-SVD and Multiway PCA

- $\mathcal{X}_j,\ j = 1, 2, \ldots, m$ are the training images
- M is the mean image
- $\mathcal{A}(:, j, :) = \mathcal{X}_j - M$ stores the mean-subtracted images
- $\mathcal{K} = \mathcal{A} * \mathcal{A}^\top = \mathcal{U} * \mathcal{S} * \mathcal{S}^\top * \mathcal{U}^\top$ is the covariance tensor
- Left orthogonal $\mathcal{U}$ contains the principal components with respect to $\mathcal{K}$

$$\mathcal{A}(:, j, :) \approx \mathcal{U}(:, 1:k, :) * \mathcal{U}(:, 1:k, :)^\top * \mathcal{A}(:, j, :) = \sum_{t=1}^{k} \mathcal{U}(:, t, :) * \mathcal{C}(t, j, :)$$



$\mathcal{T}$-QR Decomposition

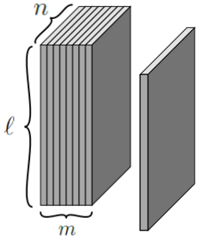
- **Theorem**: Let $\mathcal{A}$ be an $\ell \times m \times n$ real-valued tensor, then $\mathcal{A}$ can be factored as

$$\mathcal{A} * \mathcal{P} = \mathcal{Q} * \mathcal{R}$$

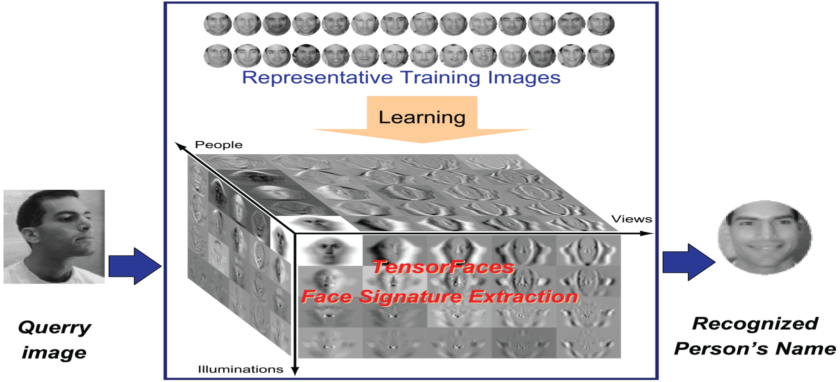
where $\mathcal{Q}$ is orthogonal $\ell \times \ell \times n$, $\mathcal{R}$ is $\ell \times m \times n$ f-upper triangular, and $\mathcal{P}$ is a permutation tensor

---

- Cheaper for updating and downdating



Face Recognition Task



- Multilinear (Tensor) ICA and Dimensionality Reduction", M.A.O. Vasilescu, D. Terzopoulos, Proc. 7th International Conference on Independent Component Analysis and Signal Separation (ICA07), London, UK, September, 2007. In Lecture Notes in Computer Science, 4666, Springer-Verlag, New York, 2007, 818-826

Face Recognition Task

- Experiment 1: randomly selected 15 images of each person as training set and test all remaining images
- Experiment 2: randomly selected 5 images of each person as the training set and test all remaining images
- Preprocessing: decimated the images by a factor of 3 to 64×56 pixels
- 20 trials for each experiment

- The Extended Yale Face Database B, `http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html`

$\mathcal{T}$-SVD vs. PCA

	RR	Storage for $\mathcal{T}$-SVD	Storage for PCA
mean	0.8095	34762	98654
median	0.83	34580	91274
maximum	0.93	37492	132056
minimum	0.61	31668	77680

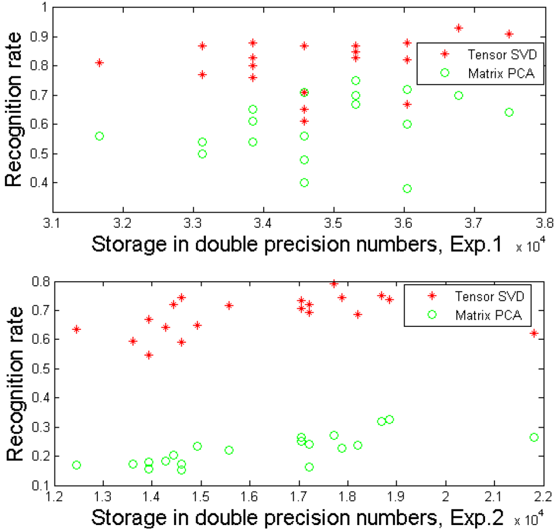
Table: Comparison between Tensor SVD and PCA in Experiment 1.

	RR	Storage for $\mathcal{T}$-SVD	Storage for PCA
mean	0.6845	16203	94310
median	0.7	16318	92100
maximum	0.79	21812	117888
minimum	0.5467	12464	73680

Table: Comparison between Tensor SVD and PCA in Experiment 2.

- N. Hao, M.E. Kilmer, K. Braman, R.C. Hoover, Facial Recognition Using Tensor-Tensor Decompositions, SIAM J. Imaging Sci., 6(1), 437-463

$\mathcal{T}$-SVD vs. PCA



- N. Hao, M.E. Kilmer, K. Braman, R.C. Hoover, Facial Recognition Using Tensor-Tensor Decompositions, SIAM J. Imaging Sci., 6(1), 437-463

$\mathcal{T}$-QR vs. PCA

	RR	Storage for T-PQR	Storage for PCA
mean	0.849	78788	127978
median	0.86	78624	133998
maximum	0.95	85540	147592
minimum	0.72	71708	100984

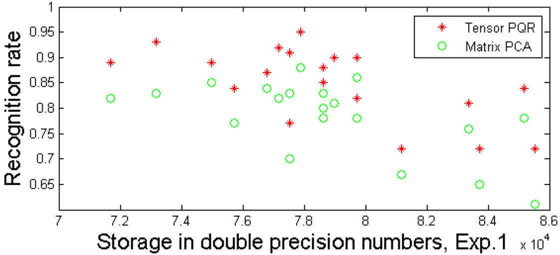
Table: Comparison between Tensor PQR and PCA in Experiment 1.

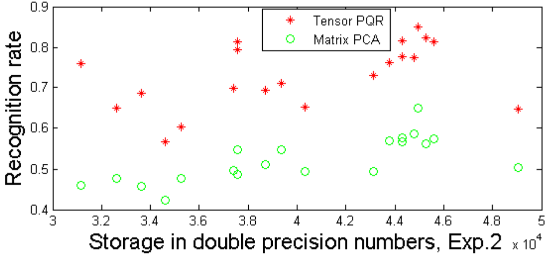
	RR	Storage for T-PQR	Storage for PCA
mean	0.731	40164	121940
median	0.745	39852	116046
maximum	0.85	49036	154728
minimum	0.5667	31160	84732

Table: Comparison between Tensor PQR and PCA in Experiment 2.

- N. Hao, M.E. Kilmer, K. Braman, R.C. Hoover, Facial Recognition Using Tensor-Tensor Decompositions, SIAM J. Imaging Sci., 6(1), 437-463

$\mathcal{T}$-QR vs. PCA





- N. Hao, M.E. Kilmer, K. Braman, R.C. Hoover, Facial Recognition Using Tensor-Tensor Decompositions, SIAM J. Imaging Sci., 6(1), 437-463

Non-Negative Tensor Decompositions - t-product

- Given a nonnegative third-order tensor $\mathcal{T} \in \mathbb{R}^{\ell \times m \times n}$ and a positive integer $k < \min(l, m, n)$
- Find nonnegative $\mathcal{G} \in \mathbb{R}^{\ell \times k \times n}$, $\mathcal{H} \in \mathbb{R}^{k \times m \times n}$ such that

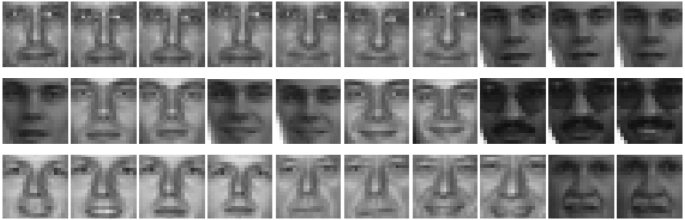
$$\min_{\hat{\mathcal{G}}, \hat{\mathcal{H}}} \|\mathcal{T} - \mathcal{G} * \mathcal{H}\|_F^2$$

Non-Negative Tensor Decompositions - t-product

- Given a nonnegative third-order tensor $\mathcal{T} \in \mathbb{R}^{\ell \times m \times n}$ and a positive integer $k < \min(l, m, n)$
- Find nonnegative $\mathcal{G} \in \mathbb{R}^{\ell \times k \times n}$, $\mathcal{H} \in \mathbb{R}^{k \times m \times n}$ such that

$$\min_{\hat{\mathcal{G}}, \hat{\mathcal{H}}} \|\mathcal{T} - \mathcal{G} * \mathcal{H}\|_F^2$$

- Facial Recognition Example:
  - Dataset: The Center for Biological and Computational Learning (CBCL) Database
  - Training images: 200
  - $k = 10$



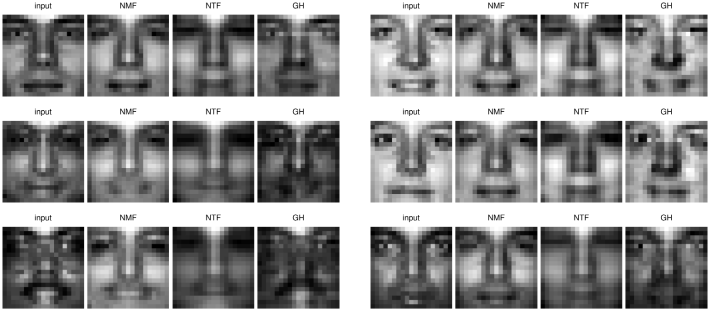
Reconstructed Images Based on NMF, NTF-CP and NTF-GH

NMF: $A(:,j) \approx WH(:,j)$

NTF-CP: $\mathcal{T}(:,j,:) \approx \sum_{i=1}^{K} b^{(i)}(j)(a^{(i)} \circ c^{(i)})$

NTF-GH: $\mathcal{T}(:,j,:) \approx \mathcal{G} * \mathcal{H}(:,j,:)$

Input NMF NTF GH Input NMF NTF GH



N. Hao, L. Horesh, M. Kilmer, Non-negative Tensor Decomposition, *Compressed Sensing & Sparse Filtering*, Springer, 123–148, 2014

Tensor Nuclear Norm

- If A is an $\ell \times m$, $\ell \geq m$ matrix with singular values σ_i, the nuclear norm $\|A\|_{\circledast} = \sum_{i=1}^{m} \sigma_i$.
- However, in the t-SVD, we have singular tubes (the entries of which need not be positive), which sum up to a singular tube!
- The entries in the jth singular tube are the inverse Fourier coefficients of the length-n vector of the jth singular values of $\widehat{\mathcal{A}}_{:,:,i}, i = 1..n$.

Definition

For $\mathcal{A} \in \mathbb{R}^{\ell \times m \times n}$, our tensor nuclear norm is $\|\mathcal{A}\|_{\circledast} = \sum_{i=1}^{\min(\ell,m)} \|\sqrt{n} F_V s_i\|_1 = \sum_{i=1}^{\min(\ell,m)} \sum_{j=1}^{n} \widehat{\mathcal{S}}_{i,i,j}$. (Same as the matrix nuclear norm of $\mathrm{circ}(\mathcal{A})$).

Tensor Nuclear Norm

Theorem (Semerci,Hao,Kilmer,Miller)

The tensor nuclear norm is a valid norm.

Since the t-SVD extends to higher-order tensors [Martin et al, 2012], the norm does, as well.

Tensor Completion

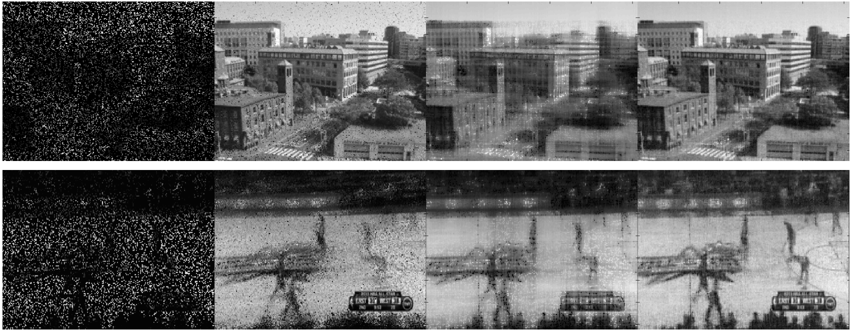
- Given unknown tensor ${}^T\!M$ of size $n_1 \times n_2 \times n_3$, given a subset of entries $\{{}^T\!M_{ijk} : (i,j,k) \in \mathbf{\Omega}\}$ where $\mathbf{\Omega}$ is an indicator tensor of size $n_1 \times n_2 \times n_3$. Recover the entire ${}^T\!M$:

$$\begin{aligned} \text{min} \quad & \|{}^T\!X\|_{\circledast} \\ \text{subject to} \quad & P_{\mathbf{\Omega}}({}^T\!X) = P_{\mathbf{\Omega}}({}^T\!M) \end{aligned}$$

- The $(i,j,k)_{th}$ component of $P_{\mathbf{\Omega}}({}^T\!X)$ is equal to ${}^T\!M_{ijk}$ if $(i,j,k) \in \mathbf{\Omega}$ and zero otherwise.
- Similar to the previous problem, this can be solved by ADMM, with 3 update steps, one which decouples, one that is a shrinkage / thresholding step.

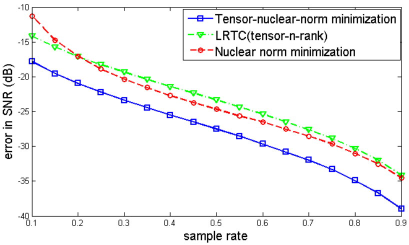
Numerical Results

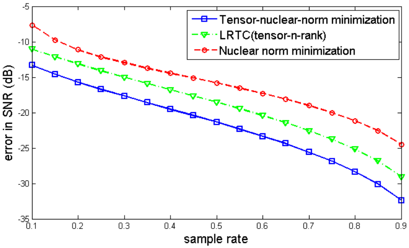
- TNN minimization, Low Rank Tensor Completion (LRTC) [Liu, et al, 2013] based on tensor-n-rank [Gandy, et al, 2011], and the nuclear norm minimization on the vectorized video data [Cai, et al, 2010].
- MERL[3] video, Basketball video



[3] with thanks to A. Agrawal

Numerical Results





Toolbox and References

- Brett W. Bader, Tamara G. Kolda and others. MATLAB Tensor Toolbox Version 2.5, Available online, January 2012, http://www.sandia.gov/~tgkolda/TensorToolbox/
- T. G. Kolda and B. W. Bader, Tensor Decompositions and Applications, *SIAM Review 51(3):455-500, 2009*
- *A. Cichocki, R. Zdunek, A.H. Phan, S.i. Amari, Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation, 2009*
- *M.E. Kilmer and C.D. Martin. Factorization strategies for third-order tensors,* Linear Algebra and its Applications, *Special Issue in Honor of G. W. Stewart's* 70^{th} *birthday, vol. 435(3):641–658, 2011*
- *N. Hao, L. Horesh, M. Kilmer, Non-negative Tensor Decomposition,* Compressed Sensing & Sparse Filtering, *Springer, 123–148, 2014*