PALSIR : A new approach to Nonnegative Tensor Factorization

Christos Boutsidis, Efstratios Gallopoulos
Department of Computer Engineering & Informatics, University of Patras, Greece
Peter Zhang, Robert J. Plemmons
Departments of Mathematics and Computer Science, Wake Forest University, USA

ABSTRACT

- PALSIR (Projected Alternating Least Squares with Initialization and Regularization) is designed to decompose a nonnegative tensor $T \in \mathbb{R}_+^{D_1 \times D_2 \times D_3}$ into a sum of k nonnegative rank-1 tensors $T_i \in \mathbb{R}_+^{D_1 \times D_2 \times D_3}$ each of which can be written as the outer product of three nonnegative vectors $x_i \in \mathbb{R}_+^{D_1}$, $y_i \in \mathbb{R}_+^{D_2}$ and $z_i \in \mathbb{R}_+^{D_3}$, $i = 1 : k$.
- PALSIR consists of the following phases: $i)$ initialization of 2 out of the 3 vector groups x_i's, y_i's, z_i's; $ii)$ an iterative tri-alternating procedure where, at each stage, two of the three groups of vectors remain fixed and a nonnegative solution is computed with respect to the third group. Each of these stages requires solving $\{D_1, D_2$ or $D_3\}$ nonnegativity constrained least squares (LS) problems.
- Experiments with $i)$ eye image databases for biometric iris recognition applications for personnel identification and $ii)$ Columbia space shuttle images.

BASIC IDEA MOTIVATION

Basic Question : How to solve the Nonnegative LS subproblems ?

DO NOT: Use nonnegative LS algorithms, perhaps even the fast [4].

DO : Solve LS by projecting onto the nonnegative orthant [7].

Langville et al. [7] using a similar idea to NMF (Nonnegative Matrix Factorization) showed that (LS + projection) is far superior to (nonnegative LS) in terms of runtime while remaining competitive in convergence behavior and approximation accuracy.

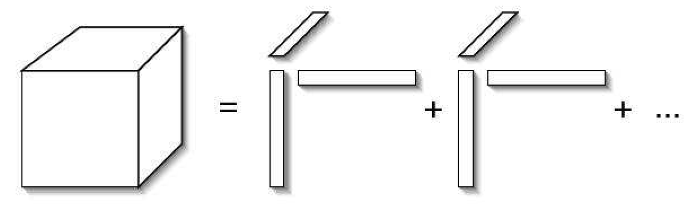
INTRODUCTION TO PARAFAC – NTF MODELS

A multilinear analogue to the SVD matrix decomposition method is the PARAllel FACtor (PARAFAC) Analysis model [5]. Given a tensor $T \in \mathbb{R}^{D_1 \times D_2 \times D_3}$ and a positive integer k, the underlying optimization problem in PARAFAC tensor decomposition is

$$\min_{x_i \in \mathbb{R}^{D_1}, y_i \in \mathbb{R}^{D_2}, z_i \in \mathbb{R}^{D_3}} ||T - \sum_{i=1}^{k} x_i \circ y_i \circ z_i||_F^2 \qquad (1)$$

while in the case of the Nonnegative Tensor Factorization (assume nonnegative T)

$$\min_{x_i \in \mathbb{R}_+^{D_1}, y_i \in \mathbb{R}_+^{D_2}, z_i \in \mathbb{R}_+^{D_3}} ||T - \sum_{i=1}^{k} x_i \circ y_i \circ z_i||_F^2 \qquad (2)$$



- NTF, a multilinear extension of NMF (Nonnegative Matrix Factorization), can be considered as a nonnegative PARAFAC model.
- Utilize PARAFAC algorithms to compute NTF.
- Enforce nonnegativity on PARAFAC.

MOTIVATION – PRACTITIONERS INTERESTED IN A NTF MODEL

Our work is motivated by the need for new multilinear algebra (tensor) based techniques for analysis of high-dimensional data, facilitating (i) reduced dimensionality representation, (ii) object or target identification, (iii) clustering, and (iv) feature extraction from image ensembles or massive data sets

- **T. Kolda** and **B. Bader** studying Higher Order Methods of Web Link Analysis [6] highlight the importance of an efficient nonnegative PARAFAC Decomposition. Their TOPHITS algorithm can naturally be explained only via such a model.
- **R. Bro** in his Ph.D thesis [3] refers to various applications from the field of chemometrics that could be improved by adding nonnegative constraints to PARAFAC model.
- **R. Plemmons** et al., [9] studying Biometric Identification by iris recognition, have emphasized the need for compressing and extracting features from massive databases of tensor arrays of enrolled iris images for personnel identification and verification for security purposes worldwide. Preserving nonnegativity of pixel values is an important aspect of this work.
- **I. Antonellis** and **E. Gallopoulos** study tensor based techniques for Information Retrieval [2] and use PALSIR as an alternative method to HOSVD for local denoising of document's tensors.

PALSIR ALGORITHM

- Assume $T \in \mathbb{R}_+^{D_1 \times D_2 \times D_3}$ and positive integer k, tackle (2)
- Group x_i's, y_i's and z_i's as columns in $X \in \mathbb{R}_+^{D_1 \times k}$, $Y \in \mathbb{R}_+^{D_2 \times k}$ and $Z \in \mathbb{R}_+^{D_3 \times k}$ respectively.
- Initialize $\{X,Y\}$, $\{X,Z\}$ or $\{Z,Y\}$, (suppose $\{X,Y\}$ initialization)
- Iterative Tri-Alternating Minimization

  subproblem 1 : Fix $\{T,X,Y\}$. Fit Z in the LS sense. Project Z onto the nonnegative orthant.

  subproblem 2 : Fix $\{T,X,Z\}$. Fit Y in the LS sense. Project Y onto the nonnegative orthant

  subproblem 3 : Fix $\{T,Z,Y\}$. Fit X in the LS sense. Project X onto the nonnegative orthant

PALSIR SUBPROBLEMS typical – subproblem 1 : Fix $\{T,X,Y\}$

- Fit Z in the LS sense : Solve $(C_z^T C_z) Z^T = C_z^T T_z$ using a regularization scheme, as done for linear ill-posed inverse problems. Here $C_z = X \odot Y \in \mathbb{R}^{D_1 D_2 \times k}$, $T_z \in \mathbb{R}^{D_1 D_2 \times D_3}$ is the unfolding tensor across the D_3-th dimension and $\odot$ stands for the Khatri-Rao product := $X \odot Y$ = [kron(x_1, y_1) kron(x_2, y_2).... kron(x_k, y_k)].
- Projection : Zero out the negative elements of Z.

IDEAS FOR INITIALIZATION STAGE

- random nonnegative initialization
- modify appropriately available PARAFAC initializations [6]
- multilinear extensions of existing NMF initializations [7]

IDEAS FOR REGULARIZATION SCHEME

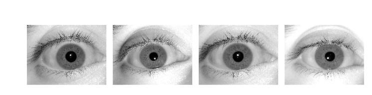
- Deployed to alleviate ill-posedness of $(C_z^T C_z) Z^T = C_z^T T_z$, with $(C_z^T C_z) \in R^{k \times k}$.
- Use Tikhonov regularization and regularization matrix $L \in R^{k \times k}$.
- Choose L as a multiple of the identity matrix $L := \lambda_z I_k$, for appropriately chosen parameter λ_z.
- Solve : $(C_z^T C_z + \lambda_z I_k) Z^T = C_z^T T_z$.

PALSIR COMPLEXITY

Initialization complexity + (number of iterations)*($O((D_1 D_2 + D_2 D_3 + D_1 D_3) k^2 + (D_1 + D_2 + D_3) k^3)$). **Actually this is the same as the basic PARAFAC Complexity**.

PALSIR vs. OTHER NTF ALGORITHMS

- We compare PALSIR with the approach of solving nonnegative LS systems [1] as well as with two other novel NTF approaches [8,10].
- We use a set of thirty (120×160) eye images put together to form the tensor $T \in \mathbb{R}_+^{120 \times 160 \times 30}$ and choose $k = 30$. For the case of PALSIR we initialize $\{X, Y\}$ with ones and choose the regularization parameters equal to 0.1. Define the relative error as : $RE = ||T - \sum_{i=1}^{k} x_i \circ y_i \circ z_i||_F^2 / ||T||_F^2$. Four of the eye images :



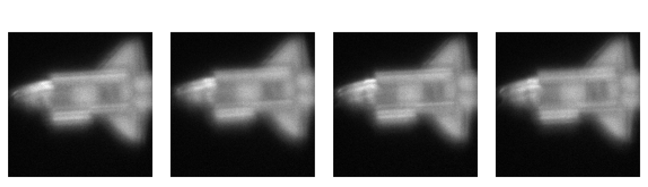
Algorithm	iteration runtime	stopping criterion	iterations
Nonnegative LS[1]	5.15 sec	$RE \leq 2 * 10^{-2}$	2
NTF approach of [10]	0.63 sec	$RE \leq 2 * 10^{-2}$	88
SNTF2D [8]	1.62 sec	$RE \leq 2 * 10^{-2}$	38
PALSIR	0.17 sec	$RE \leq 2 * 10^{-2}$	6

Pentium 2.0 GHz computer with 1 GB RAM, Matlab Environment.

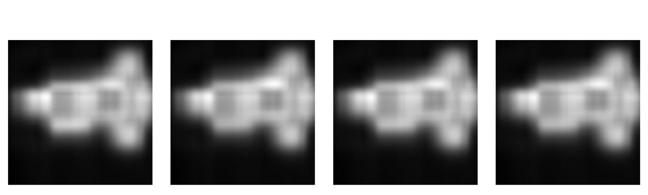
- **Clearly** : PALSIR is faster than the other approaches (column 2) while remaining competitive in convergence behavior (column 4) and approximation accuracy (column 3).
- PALSIR succeeds in compressing eye image databases for biometric iris recognition applications for personnel identification and verification (achieved compression ratio $2 : 100$).

EXPERIMENTS

- A set of sixteen (128×128) space shuttle images put together to form $T \in \mathbb{R}_+^{128 \times 128 \times 16}$. Images are from the space shuttle Columbia on its final orbit before disintegration upon re-entry in February 2003 and were taken in conjunction with our work at the Air Force Maui Space Center. See below four of them:

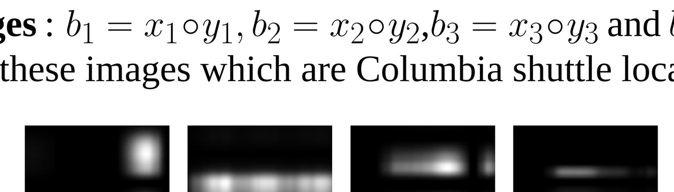


- By decomposing T into sum of rank one tensors, each image is expressed as linear combination of basis images. Due to the nonnegativity constraints of the basis images as well as the weights of these combinations, the decomposition obtained is an **additive** linear combination. This property allows us to have the basis images represent local parts of the Columbia shuttle objects.
- We run PALSIR with $k = 4$, initialize $\{X, Y\}$ as random nonnegative matrices and choose the regularization parameters equal to one. After computing the Factors $X \in \mathbb{R}_+^{128 \times 4}$, $Y \in \mathbb{R}_+^{128 \times 4}$ and $Z \in \mathbb{R}_+^{16 \times 4}$ we reconstruct T as shown below :



LEARNING THE PARTS OF OBJECTS BY NTF

- **Objects of our interest** : Columbia space shuttle on final approach to re-entry and disintegration.
- **Basis images** : $b_1 = x_1 \circ y_1, b_2 = x_2 \circ y_2, b_3 = x_3 \circ y_3$ and $b_4 = x_4 \circ y_4$. See below these images which are Columbia shuttle local parts



- **Additive representation of each Columbia shuttle** :

$$T(:,:,i) \approx z_1(i) * b_1 + z_2(i) * b_2 + z_3(i) * b_3 + z_4(i) * b_4, \; i = 1 : 16.$$

CURRENT WORK

- Further refinement of PALSIR regarding issues of initialization and regularization.
- Incorporate sparsity constraints to the nonnegative Factors by changing regularization parameters, as observed in [7].
- Further experiments with biometric identification data.
- Tests on image arrays of similar but different objects to identify a particular object from the set of images.
- Shed some light on the field of nonnegative tensor rank.

References

[1] C. Andersson and R. Bro. The N-way Toolbox for MATLAB. Chemom. Intell. Lab. Syst. 52 (1):1-4, 2000.

[2] I. Antonellis and E. Gallopoulos. Tensor Based Text Representation : A new Dimension in IR. MMDS 2006, poster.

[3] R. Bro. Multi-way Analysis in the Food Industry. Models, Algorithms and Applications. Ph.D Thesis, U of Amsterdam, 1998.

[4] R. Bro and S.D Jong. A fast non-negativity constrained linear least squares algorithm, Journal of Chemometrics, 11:393401, 1997.

[5] R.A. Harshman. Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis. UCLA working papers in phonetics, **16** (1970), no. 16, 1–84.

[6] T. Kolda and B. Bader. The TOPHITS Model for Higher-Order Web Link Analysis. SIAM Data Mining conference, April 2006.

[7] A. Langville et al. Algorithms, Initializations, and Convergence for the Nonnegative Matrix Factorization, preprint.

[8] M Mørup and M.N. Schmidt. Sparse Non-negative Tensor Factor Double Deconvolution (SNTF2D) for multi channel time-frequency analysis, 2006, Technical Report

[9] R. Plemmons, et al. Computational imaging systems for iris recognition. Proc. SPIE 49th Annual Meeting, Vol. 5559, pp. 346-357, August 2004.

[10] A. Shashua and T. Hazan. 2005. Non-negative tensor factorization with applications to statistics and computer vision. Proceedings of the 22^{nd} International Conf. on Machine Learning, Germany.