Higher-Order Nonnegative CANDECOMP/PARAFAC Tensor Decomposition Using Proximal Algorithm

Deqing Wang[⋆†] Fengyu Cong[⋆†] Tapani Ristaniemi[†]

[⋆]School of Biomedical Engineering, Dalian University of Technology, Dalian, China
[†]Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

Introduction

- We proposed a new nonnegative CANDECOM/PARAFAC (NCP) model using proximal algorithm [1].
- The block principal pivoting (BPP) method in alternating nonnegative least squares (ANLS) framework [2] was employed to minimize the objective function.
- Our method can guarantee the convergence and accelerate the computation.

Mathematical model

The NCP using proximal algorithm is

$$\min_{\boldsymbol{A}^{(1)},\dots,\boldsymbol{A}^{(N)}} \left\{ \frac{1}{2}\left\| \mathcal{X} - [\![\boldsymbol{A}^{(1)},\dots,\boldsymbol{A}^{(N)}]\!] \right\|_F^2 + \sum_{n=1}^{N} \frac{\alpha_n}{2} \left\| \widetilde{\boldsymbol{A}}^{(n)} - \boldsymbol{A}^{(n)} \right\|_F^2 \right\} \quad \text{s.t. } \boldsymbol{A}^{(n)} \geqslant 0 \text{ for } n=1,\dots,N, \tag{1}$$

where $\widetilde{\boldsymbol{A}}^{(n)} \in \mathbb{R}^{I_n \times R}$ is the former version of $\boldsymbol{A}^{(n)}$ in previous iteration. $\boldsymbol{A}^{(n)}$ in the kth iteration can be updated alternatively by the following subproblem:

$$\boldsymbol{A}_{k+1}^{(n)} = \arg\min_{\boldsymbol{A}^{(n)}\geqslant 0} \left\{ \frac{1}{2}\left\| \boldsymbol{X}_{(n)} - \boldsymbol{A}^{(n)}\left(\boldsymbol{B}_k^{(n)}\right)^T \right\|_F^2 + \frac{\alpha_n}{2}\left\| \boldsymbol{A}_k^{(n)} - \boldsymbol{A}^{(n)} \right\|_F^2 \right\}, \tag{2}$$

where $\boldsymbol{B}_k^{(n)} = \left(\boldsymbol{A}_k^{(N)} \odot \cdots \odot \boldsymbol{A}_k^{(n+1)} \odot \boldsymbol{A}_{k+1}^{(n-1)} \odot \cdots \odot \boldsymbol{A}_{k+1}^{(1)} \right)$.

Nonnegative Least Squares

The objective function in (2) can be equivalently rewritten as

$$\mathcal{F}_2 = \frac{1}{2}\left\| \begin{pmatrix} \boldsymbol{X}_{(n)}^T \\ \sqrt{\alpha_n}\left(\boldsymbol{A}_k^{(n)}\right)^T \end{pmatrix} - \begin{pmatrix} \boldsymbol{B}_k^{(n)} \\ \sqrt{\alpha_n}\boldsymbol{I}_R \end{pmatrix} \left(\boldsymbol{A}^{(n)}\right)^T \right\|_F^2 .$$

Obviously, (2) is still a nonnegative least squares (NNLS) problem. Therefore, we employ the block principal pivoting (BPP) method [2] to solve the subproblem in (2).

Experiments & Results

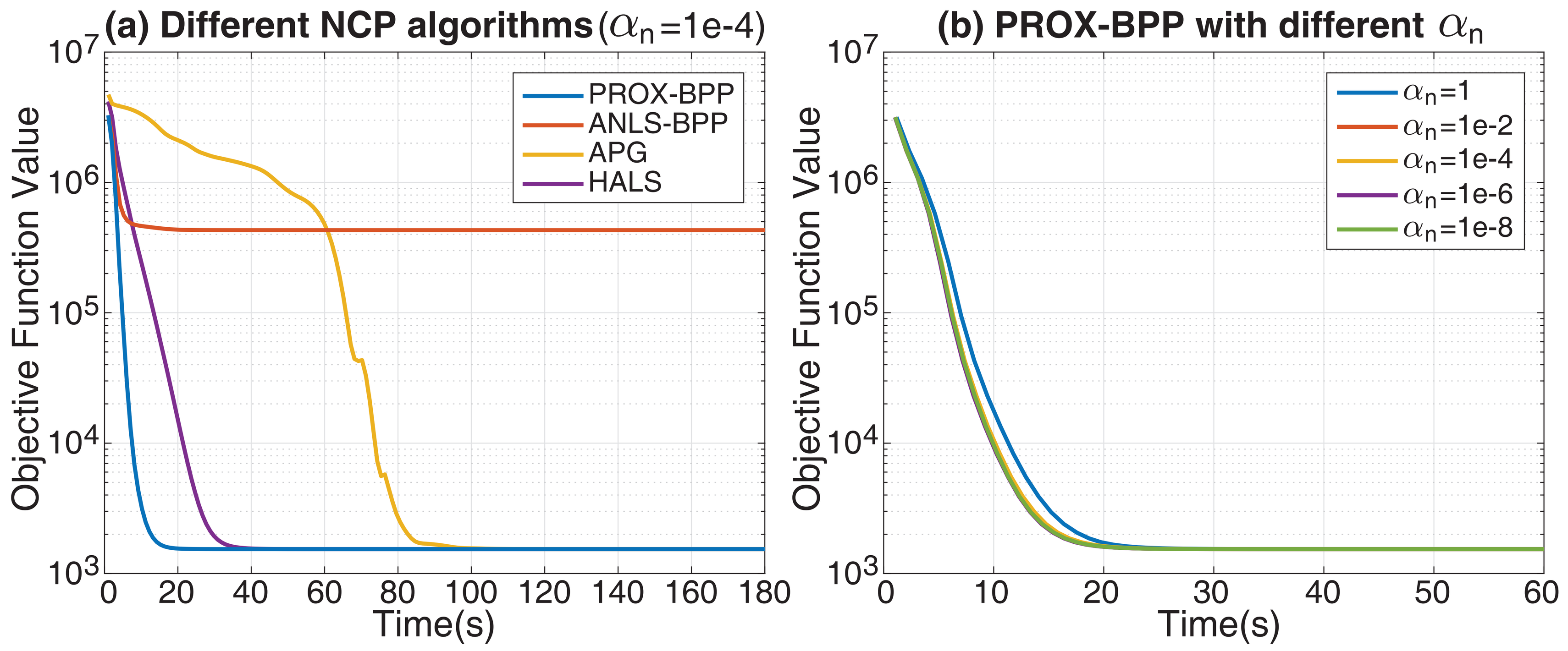

Fig. 1. Convergence of NCP algorithms on the synthetic data.

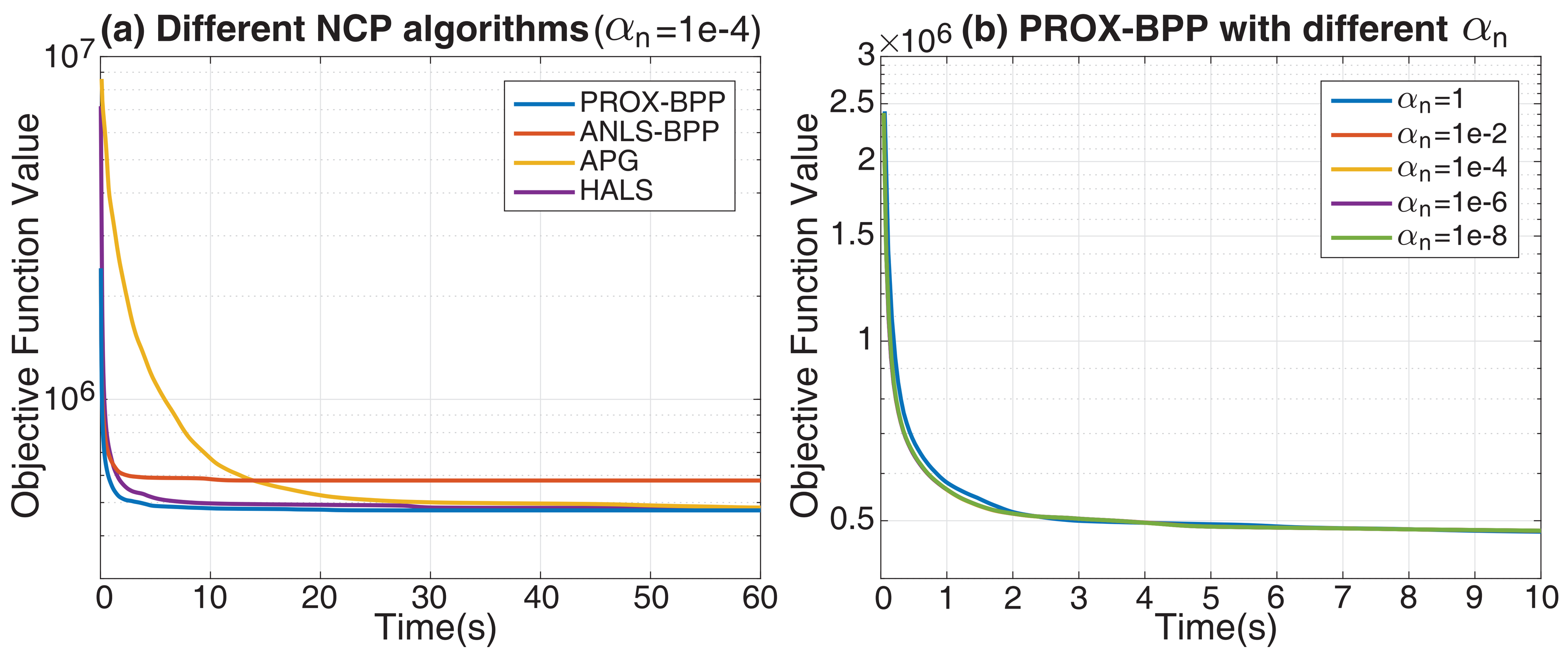

Fig. 2. Convergence of NCP algorithms on the real ERP data.

Data description:

- We synthesized a fourth-order nonnegative tensor by $\mathcal{X}_{\text{Syn}} = [\![\boldsymbol{S}^{(1)}, \boldsymbol{A}^{(2)}, \boldsymbol{A}^{(3)}, \boldsymbol{A}^{(4)}]\!] \in \mathbb{R}^{1000\times100\times100\times5}$, in which $\boldsymbol{S}^{(1)} \in \mathbb{R}^{1000\times7}$ is the signal matrix, and $\boldsymbol{A}^{(2)}, \boldsymbol{A}^{(3)} \in \mathbb{R}^{100\times7}$, $\boldsymbol{A}^{(4)} \in \mathbb{R}^{5\times7}$ are random matrices in uniform distribution.
- We utilized a set of real preprocessed fourth-order event-related potential (ERP) tensor data $\mathcal{X}_{\text{ERP}} \in \mathbb{R}^{9\times71\times60\times42}$ (channel × frequency × time × subject-group = 9 × 71 × 60 × 42).

Experimental setting:

- The factor matrices were initialized using random numbers by command `max(0,randn(`I_n, R`))`.
- The number of components are 7 and 40 for the synthetic and real data.
- The values of 1, 1e-2, 1e-4, 1e-6, and 1e-8 were tested for α_n.

Algorithm

Algorithm 1: NCP using proximal algorithm

Input : $\mathcal{X}$, R, $\boldsymbol{\alpha}$
Output: $\boldsymbol{A}^{(n)}$, $n = 1,\dots,N$

Initialize $\boldsymbol{A}^{(n)} \in \mathbb{R}^{I_n \times R}$, $n = 1,\dots,N$, using random numbers;
repeat
  for $n = 1$ **to** N **do**
    Make mode-n unfolding of $\mathcal{X}$ as $\boldsymbol{X}_{(n)}$;
    Compute MTTKRP $\boldsymbol{X}_{(n)}\boldsymbol{B}_k^{(n)}$ and $\left(\boldsymbol{B}_k^{(n)}\right)^T\boldsymbol{B}_k^{(n)}$;
    $\left(\boldsymbol{B}_k^{(n)}\right)^T\boldsymbol{B}_k^{(n)} \leftarrow \left(\boldsymbol{B}_k^{(n)}\right)^T\boldsymbol{B}_k^{(n)} + \alpha_n \boldsymbol{I}_R$;
    $\boldsymbol{X}_{(n)}\boldsymbol{B}_k^{(n)} \leftarrow \boldsymbol{X}_{(n)}\boldsymbol{B}_k^{(n)} + \alpha_n \boldsymbol{A}_k^{(n)}$;
    Update factor $\boldsymbol{A}^{(n)}$ based on (2) using BPP:
    $\boldsymbol{A}_{k+1}^{(n)} = \operatorname{argmin}_{\boldsymbol{A}^{(n)}\geqslant 0} \mathcal{F}_2\left(\boldsymbol{A}^{(n)}\right)$
    $= \texttt{NNLS\_BPP}\left(\boldsymbol{X}_{(n)}\boldsymbol{B}_k^{(n)}, \left(\boldsymbol{B}_k^{(n)}\right)^T\boldsymbol{B}_k^{(n)}\right)$.
  end
until *some termination criterion is reached*;
return $\boldsymbol{A}^{(n)}$, $n = 1,\dots,N$.

Remarks

- The choice of parameter α_n for PROX-BPP is said to be related the noise level in the data. Surprisingly, our method is very robust with different α_n values. We suggest to select 1e-2 $\leqslant \alpha_n \leqslant$ 1e-4.
- Our proposed NCP method using proximal algorithm can be combined with many other NNLS algorithms.

Contact Information

Deqing Wang, deqing.wang@foxmail.com, http://deqing.me/
Fengyu Cong, cong@dlut.edu.cn
Tapani Ristaniemi, tapani.e.ristaniemi@jyu.fi

References

[1] D. P. Bertsekas, *Nonlinear Programming, Third Edition*. Belmont, Massachusetts: Athena Scientific, 2016.

[2] J. Kim and H. Park, "Fast nonnegative matrix factorization: An active-set-like method and comparisons," *SIAM Journal on Scientific Computing*, vol. 33, no. 6, pp. 3261–3281, Jan. 2011.