A Projected Newton-type Algorithm for Rank-revealing Nonnegative Block-Term Tensor Decomposition

Eleftherios Kofidis, Paris Giampouras, Athanasios Rontogiannis

A Projected Newton-type Algorithm for Rank-revealing Nonnegative Block-Term Tensor Decomposition

Eleftherios Kofidis
Dept. of Statistics and Insurance Science
University of Piraeus
18534 Piraeus, Greece
kofidis@unipi.gr

Paris V. Giampouras
Mathematical Inst. for Data Science
Johns Hopkins University
Baltimore, MD 21218, USA
parisg@jhu.edu

Athanasios A. Rontogiannis
School of Electr. and Computer Eng.
National Technical University of Athens
15780 Athens, Greece
th_rontogiannis@mail.ntua.gr

***Abstract*—The block-term tensor decomposition (BTD) model has been receiving increasing attention as a quite flexible way to capture the structure of 3-dimensional data that can be naturally viewed as the superposition of R block terms of multilinear rank $(L_r, L_r, 1)$, $r = 1, 2, \ldots, R$. Versions with nonnegativity constraints, especially relevant in applications like blind source separation problems, have only recently been proposed and they all share the need to have an *a-priori* knowledge of the number of block terms, R, and their individual ranks, L_r. Clearly, the latter requirement may severely limit their practical applicability. Building upon earlier work of ours on unconstrained BTD model selection and computation, we develop *for the first time* in this paper a method for nonnegative BTD approximation that is also rank-revealing. The idea is to impose column sparsity jointly on the factors and successively estimate the ranks as the numbers of factor columns of non-negligible magnitude. This is effected with the aid of nonnegative alternating iteratively reweighted least squares, implemented via projected Newton updates for increased convergence rate and accuracy. Simulation results are reported that demonstrate the effectiveness of our method in accurately estimating both the ranks and the factors of the nonnegative least squares BTD approximation.**

***Index Terms*—Block coordinate descent (BCD), block successive upper bound minimization (BSUM), block-term decomposition (BTD), hierachical iterative reweighted least squares (HIRLS), Newton, nonnegative, rank, tensor**

I. Introduction

Nonnegative tensor factorization (NTF) [1], that is, fitting to a nonnegative tensor a decomposition model with some or all of its factors being constrained to only have nonnegative entries, has been well-studied for the classical canonical polyadic decomposition (CPD) and Tucker decomposition (TD) models, and has found successful applications in numerous areas, including hyperspectral image (HSI) unmixing [2], spectroscopy [3], neuroimaging [4], and topic modeling [5], among others. The (almost sure) existence and uniqueness of an NTF approximation has been proved for nonnegative CPD (NCPD) [6, Proposition 10], however the results of [6] provide positive evidence for other models as well [1]. The minimum required number of nonnegative simple tensors that sum up to the given one (referred to with the general term of nonnegative rank) is no smaller than the number of unconstrained terms. Most of the NCPD algorithms are of the block coordinate descent (BCD) type, with nonnegatively constrained subproblems. Such an alternating procedure is guaranteed to be convergent if there is a unique solution per block iteration [7].

However, despite its wide range of applications, NCPD is not always the most appropriate NTF model. Blind HSI unmixing is a notable example, where a nonnegative block term decomposition (BTD) with R rank-$(L_r, L_r, 1)$ terms more naturally represents the mixture of an equal number of materials (endmembers) whose fractions at the image pixels give rise to nonnegative matrices (abundance maps) of rank L_r, $r = 1, 2, \ldots, L_r$ [2]. Moreover, this is only one application example where the knowledge of R is more important than that of the specific values of the L_rs. As demonstrated in [2], the latter can be chosen to be all equal (to L) as the unmixing result is not that sensitive to their choice provided that they are sufficiently large. Of course, in addition to increasing the computational complexity of the BTD computation, setting the L_rs too high may hinder interpretation of the results through letting noise/artifact components interfere with the desired ones [8].

The BTD model, including its rank-$(L_r, L_r, 1)$ version of interest in most applications and in this paper, was first introduced in [9]. For a nearly extensive review of its applications, the reader may refer to [8]. Methods for computing nonnegative BTD (NBTD) models have only recently been proposed (see next subsection) and they all share the need to have an *a-priori* knowledge of the number of (nonnegative) block terms, R, and their individual (nonnegative) ranks, L_r, a requirement that may severely limit the practical applicability of these schemes. Building upon earlier work of ours on unconstrained BTD model selection and computation [8], we develop in this paper *for the first time* a method for NBTD approximation that

The work of E. Kofidis has been partly supported by the University of Piraeus Research Center.

P. V. Giampouras is supported by the European Union under the Horizon 2020 Marie Skłodowska-Curie Global Fellowship program: HyPPOCRATES-H2020-MSCA-IF-2018, Grant Agreement Number: 844290.

is also *rank-revealing*. The idea is, starting from overestimates of R and the L_rs, to impose column sparsity jointly on the factors and successively estimate the ranks as the numbers of factor columns of non-negligible magnitude. This is achieved with the aid of an appropriate regularization of the data fidelity cost function, which respects the hierarchy in the roles of the mode-1, 2 and mode-3 BTD factors, and promotes their column sparsity via mixed 1, 2 norm minimization. Following a block successive upper-bound minimization (BSUM) [10] solution approach, an alternating hierarchical iterative reweighted least squares (HIRLS) procedure results, with the constrained subproblems being inexactly solved in each iteration with the aid of projected Newton updates, for increased convergence rate and accuracy. Besides its rank-revelation ability, the proposed so-called NBTD-HIRLS algorithm also enjoys low complexity, as it only involves matrix inversions of small size. Simulation results are reported that demonstrate the effectiveness of the proposed scheme in both selecting and estimating the model, while exhibiting a considerably faster convergence than the (gradient descent-type) matrix-vector NTF (MV-NTF) method of [2], even when the latter is given the true model order.

A. Background

Existing methods for computing NBTD models include those using alternating nonnegative least squares (ANLS) [11] and nonlinear LS (NLS) [12] with the nonnegativity constraint being implicitly realized through parametrizing the unknowns as, e.g., squares of unconstrained parameters. An equivalent NCPD model could also be computed, followed by a clustering of its terms with similar mode-3 profiles [13]. Alternating optimization (AO) using alternating direction method of multipliers (ADMM) for the inner iterations has been adopted for NBTD in [14] in the context of hyper-/multi-spectral image fusion, and for HSI unmixing in [15], where the abundance maps (not their rank-L_r factors) are directly constrained and regularized to comply with the spatial priors of these images.

To the best of our knowledge, the only rank-revealing constrained BTD method proposed to date appears in [13], where the (Hankel) constraint is imposed only on the low-rank matrices corresponding to the combined 1st and 2nd modes of each of the block terms. The mode-3 factor is left unconstrained instead, while no convergence guarantee is given for this scheme. It is of interest, however, to mention at this point that, in practical applications (such as HSI unmixing), very often it is that matrix that should be kept nonnegative, not necessarily the mode-1 and mode-2 factors. In this paper, we will impose the nonnegativity constraint to all separate modal factors, as it is the case in most of the existing NBTD literature.

The multiplicative update rule (MUR), well-known from the nonnegative matrix factorization (NMF) literature [16] and previously also developed for NCPD [17], was extended to NBTD in [2] in the context of blind HSI unmixing. This so-called MV-NTF method has been very popular for the above application and hence further developed with additional constraints and regularization; see, e.g., [18]–[20] and references therein. However, in addition to requiring an *a-priori* estimate of the model order, this method is also known to be slow in its convergence. This should not be surprising since it can be viewed as a BCD procedure with projected (and appropriately rescaled) gradient-descent updates and it is known that such an algorithm cannot have a convergence rate higher than linear [21]; see also [22].

Employing projected Newton instead not only leads to faster convergence but it can also more accurately reveal the sparse features in the NTF model (since nonnegative tensors are usually sparse) [23]. In this paper, we adopt the projected Newton scheme originally proposed in [24] as it achieves a compromise between plain gradient descent and fully Newton update, thus offering both low computational cost and high convergence speed. The way this is realized is through a partial diagonalization of the Hessian matrix, as dictated by the division of the unknowns into the restricted (on or near the feasible region boundary) and free ones, which amounts to an unconstrained Newton method for the latter. The result is that the update is to a direction similar with that of the negative gradient. When accompanied with a backtracking Armijo rule [24] on the projection arc for the adjustment of the step size, this ensures the decrease of the cost function within the nonnegative orthant. The reader is referred to [25] for an accessible and insightful overview of this so-called two-metric projected Newton algorithm and other variations/simplifications thereof. The present work is also inspired from earlier work of two of the authors on rank-revealing NMF [26].

B. Notation

Lower- and upper-case bold letters are used to denote vectors and matrices, respectively. Higher-order tensors are denoted by upper-case bold calligraphic letters. For a tensor $\boldsymbol{\mathcal{X}}$, $\mathbf{X}_{(n)}$ stands for its mode-n unfolding. $\otimes$ stands for the Kronecker product. The Khatri-Rao product is denoted by $\odot$ in its general (partition-wise) version and by $\odot_{\rm c}$ in its column-wise version. $\circ$ denotes the outer product. The superscript $^{\rm T}$ stands for transposition. The identity matrix of order N and the all-ones column N-vector are respectively denoted by $\mathbf{I}_N$ and $\mathbf{1}_N$. We denote by $\mathrm{vec}(\cdot)$ the row-vectorization operator and by $\mathrm{tr}(\cdot)$ the matrix trace. The Euclidean and Frobenius norms are denoted by $\|\cdot\|_2$ and $\|\cdot\|_{\rm F}$, respectively. $[x]_+ \triangleq \max(x,0)$ is the projection of the real number x onto the set of nonnegative real numbers, $\mathbb{R}_+$. $\mathbf{M} = [\mathbf{N}]_+$ stands for the matrix with nonnegative entries (denoted as $\mathbf{M} \geq \mathbf{0}$) that results from the application of $[\cdot]_+$ in all the entries of the matrix $\mathbf{N}$.

II. Problem Statement — The Proposed Approach

Given an $I \times J \times K$ nonnegative tensor $\boldsymbol{\mathcal{Y}}$, its best (in the LS sense) rank-$(L_r, L_r, 1)$ NBTD approximation is sought for, namely

$$\min_{\mathbf{A}\geq\mathbf{0},\mathbf{B}\geq\mathbf{0},\mathbf{C}\geq\mathbf{0}} \frac{1}{2} \left\| \boldsymbol{\mathcal{Y}} - \sum_{r=1}^{R} \mathbf{A}_r \mathbf{B}_r^{\rm T} \circ \mathbf{c}_r \right\|_{\rm F}^2, \qquad (1)$$

where the matrices $\mathbf{A}_r = \left[\begin{array}{cccc} \mathbf{a}_{r,1} & \mathbf{a}_{r,2} & \cdots & \mathbf{a}_{r,L_r} \end{array}\right] \in \mathbb{R}_+^{I\times L_r}$, $\mathbf{B}_r = \left[\begin{array}{cccc} \mathbf{b}_{r,1} & \mathbf{b}_{r,2} & \cdots & \mathbf{b}_{r,L_r} \end{array}\right] \in \mathbb{R}_+^{J\times L_r}$, $r = 1, 2, \ldots, R$, and $\mathbf{C} = \left[\begin{array}{cccc} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_R \end{array}\right] \in \mathbb{R}_+^{K\times R}$, and the ranks R and L_r, $r = 1, 2, \ldots, R$ are unknown. We define $\mathbf{A} = \left[\begin{array}{cccc} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_R \end{array}\right]$ and similarly for $\mathbf{B}$. In terms of its mode unfoldings, the tensor $\boldsymbol{\mathcal{X}} \triangleq \sum_{r=1}^R \mathbf{A}_r\mathbf{B}_r^{\mathrm{T}} \circ \mathbf{c}_r$ can be written as [9]

$$\mathbf{X}_{(1)}^{\mathrm{T}} = (\mathbf{B}\odot\mathbf{C})\mathbf{A}^{\mathrm{T}} \triangleq \mathbf{P}\mathbf{A}^{\mathrm{T}}, \quad (2)$$

$$\mathbf{X}_{(2)}^{\mathrm{T}} = (\mathbf{C}\odot\mathbf{A})\mathbf{B}^{\mathrm{T}} \triangleq \mathbf{Q}\mathbf{B}^{\mathrm{T}}, \quad (3)$$

$$\mathbf{X}_{(3)}^{\mathrm{T}} = \left[\begin{array}{cccc} (\mathbf{A}_1 \odot_{\mathrm{c}} \mathbf{B}_1)\mathbf{1}_{L_1} & \cdots & (\mathbf{A}_R \odot_{\mathrm{c}} \mathbf{B}_R)\mathbf{1}_{L_R} \end{array}\right]\mathbf{C}^{\mathrm{T}} \triangleq \mathbf{S}\mathbf{C}^{\mathrm{T}} \quad (4)$$

These expressions will be used in alternatingly solving for $\mathbf{A}, \mathbf{B}, \mathbf{C}$, respectively.

In order to make up for the lack of a-priori knowledge of R and $L_r, r = 1, 2, \ldots, R$ in (1), we propose to adapt to the present context the group sparsity-promoting regularization that was proved successful in our earlier work on unconstrained BTD model selection and computation [8], namely rewrite (1) as

$$\min_{\mathbf{A}\geq\mathbf{0},\mathbf{B}\geq\mathbf{0},\mathbf{C}\geq\mathbf{0}} f(\mathbf{A},\mathbf{B},\mathbf{C}) \triangleq \frac{1}{2}\left\|\boldsymbol{\mathcal{Y}} - \sum_{r=1}^R \mathbf{A}_r\mathbf{B}_r^{\mathrm{T}} \circ \mathbf{c}_r\right\|_{\mathrm{F}}^2 + \lambda\sum_{r=1}^R \sqrt{\sum_{l=1}^L \sqrt{\|\mathbf{a}_{r,l}\|_2^2 + \|\mathbf{b}_{r,l}\|_2^2 + \eta^2} + \|\mathbf{c}_r\|_2^2 + \eta^2}, \quad (5)$$

where $\lambda > 0$ is a regularization parameter, η^2 is a very small positive constant that is added to ensure smoothness, and R and L stand for the initial (over)estimates of the model rank parameters. Observe that this is a two-level hierarchical regularizer. At the upper level (outer summation), the sparsity enforcing potential of the $\ell_{1,2}$ norm is used to eliminate whole blocks of $\mathbf{A}$ and $\mathbf{B}$ and the corresponding columns of $\mathbf{C}$. At the lower level (inner summations), it is the $\ell_{1,2}$ norm of the matrix $\left[\begin{array}{c} \mathbf{A}_r \\ \mathbf{B}_r \end{array}\right]$ that induces column sparsity to the "surviving" blocks. This regularization can then reduce the overestimates $R = R_{\mathrm{ini}}$ and $L_r = L_{\mathrm{ini}}$ of the unknown NBTD model ranks to their actual values, with a proper selection of the regularization parameter λ. The problem in (5) can be solved via a constrained BCD approach, detailed in the next section.

III. Proposed Method

The objective function in (5) is multiconvex [22], that is, convex with respect to (w.r.t.) each one of the factors $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ separately but not w.r.t. all of them. It thus makes sense to alternatingly minimize it w.r.t. one of the factors at each time, in a BCD manner. Of course, the method has to ensure that all factors remain within the feasible set and their updates decrease the objective function. The development is simplified if we follow a BSUM approach, that is, successively minimizing over the nonnegative orthant a surrogate (tight upper bound) function for each factor and at each iteration [10]. Since the constrained sub-problems are then to be tackled with the aid of a projected Newton recursion, we choose quadratic functions of the type employed in [24], [25] as surrogates. Thus, at iteration k, the following function

$$g_{\mathbf{A}}(\mathbf{A} \mid \mathbf{A}^k, \mathbf{B}^k, \mathbf{C}^k) = f(\mathbf{A}^k, \mathbf{B}^k, \mathbf{C}^k) + \mathrm{tr}\left[(\mathbf{A}-\mathbf{A}^k)^{\mathrm{T}}\nabla_{\mathbf{A}} f(\mathbf{A}^k, \mathbf{B}^k, \mathbf{C}^k)\right] + \frac{1}{2\alpha_{\mathbf{A}}^k}\mathrm{vec}(\mathbf{A}-\mathbf{A}^k)^{\mathrm{T}}\bar{\mathbf{H}}_{\mathbf{A}^k}\mathrm{vec}(\mathbf{A}-\mathbf{A}^k) \quad (6)$$

upper bounds f at and around $\mathbf{A}_k$, where

$$\nabla_{\mathbf{A}} f(\mathbf{A}^k, \mathbf{B}^k, \mathbf{C}^k) = \mathbf{A}^k\mathbf{H}_{\mathbf{A}^k} - \mathbf{Y}_{(1)}\mathbf{P}^k \quad (7)$$

is the gradient of f w.r.t. $\mathbf{A}$ at the current iteration and

$$\mathbf{H}_{\mathbf{A}^k} = \mathbf{P}^{k\mathrm{T}}\mathbf{P}^k + \lambda\mathbf{D}^k, \quad (8)$$

with $\mathbf{P}^k \triangleq \mathbf{B}^k \odot \mathbf{C}^k$ (cf. (2)) and $\mathbf{D}^k \triangleq (\mathbf{D}_1^k \otimes \mathbf{I}_L)\mathbf{D}_2^k$, where $\mathbf{D}_1^k$ is an $R \times R$ diagonal matrix with diagonal entries

$$\mathbf{D}_1^k(r,r) = \left[\sum_{l=1}^L \sqrt{\|\mathbf{a}_{rl}^k\|_2^2 + \|\mathbf{b}_{rl}^k\|_2^2 + \eta^2} + \|\mathbf{c}_r^k\|_2^2 + \eta^2\right]^{-1/2} \quad (9)$$

and $\mathbf{D}_2^k$ an $RL \times RL$ diagonal matrix, whose $((r-1)L+l)$th diagonal entry is

$$\mathbf{D}_2^k((r-1)L+l, (r-1)L+l) = (\|\mathbf{a}_{rl}^k\|_2^2 + \|\mathbf{b}_{rl}^k\|_2^2 + \eta^2)^{-1/2} \quad (10)$$

$\alpha_{\mathbf{A}}^k$ stands for the sequence of step sizes to be later employed in the projected Newton updates. Moreover, the $ILR \times ILR$ approximate Hessian in (6) is given by

$$\bar{\mathbf{H}}_{\mathbf{A}^k} = \mathbf{I}_I \otimes \mathbf{H}_{\mathbf{A}^k} \quad (11)$$

The corresponding functions and quantities for $\mathbf{B}$ and $\mathbf{C}$ are similarly defined, with $\mathbf{H}_{\mathbf{B}} = \mathbf{Q}^{\mathrm{T}}\mathbf{Q} + \lambda\mathbf{D}$ and $\mathbf{H}_{\mathbf{C}} = \mathbf{S}^{\mathrm{T}}\mathbf{S} + \lambda\mathbf{D}_1$. Our method then works by *inexactly* solving the following constrained sub-problems

$$\mathbf{A}^{k+1} = \arg\min_{\mathbf{A}\geq\mathbf{0}} g_{\mathbf{A}}(\mathbf{A} \mid \mathbf{A}^k, \mathbf{B}^k, \mathbf{C}^k), \quad (12)$$

$$\mathbf{B}^{k+1} = \arg\min_{\mathbf{B}\geq\mathbf{0}} g_{\mathbf{B}}(\mathbf{B} \mid \mathbf{A}^{k+1}, \mathbf{B}^k, \mathbf{C}^k), \quad (13)$$

$$\mathbf{C}^{k+1} = \arg\min_{\mathbf{C}\geq\mathbf{0}} g_{\mathbf{C}}(\mathbf{C} \mid \mathbf{A}^{k+1}, \mathbf{B}^{k+1}, \mathbf{C}^k), \quad (14)$$

in an alternating fashion, using a projected Newton update for each. For the ith row of $\mathbf{A}^k$, the index set of the restricted variables is determined as

$$\mathcal{I}_{\mathbf{A}_{i,:}}^k = \left\{j \mid 0 \leq [\mathbf{A}^k]_{i,j} \leq \epsilon_{\mathbf{A}}^k, [\nabla_{\mathbf{A}} f(\mathbf{A}^k, \mathbf{B}^k, \mathbf{C}^k)]_{i,j} > 0\right\}, \quad (15)$$

where $\epsilon_{\mathbf{A}}^k = \min(\varepsilon, \|\mathbf{A}^k - \nabla_{\mathbf{A}} f(\mathbf{A}^k, \mathbf{B}^k, \mathbf{C}^k)\|_{\mathrm{F}})$ with ε being a very small positive constant, and is then used to partly diagonalize the corresponding Hessian on the block diagonal of (11):

$$[\mathbf{H}_{\mathbf{A}_{i,:}}^{\mathcal{I}_{\mathbf{A}_{i,:}}^k}]_{p,q} \begin{cases} 0, & p \neq q \text{ and } p \in \mathcal{I}_{\mathbf{A}_{i,:}}^k \text{ or } q \in \mathcal{I}_{\mathbf{A}_{i,:}}^k \\ [\mathbf{H}_{\mathbf{A}^k}]_{p,q}, & \text{otherwise} \end{cases} \quad (16)$$

These matrices are obviously always positive definite and hence the following projected Newton update

$$\mathrm{vec}(\mathbf{A}^{k+1}) = [\mathrm{vec}(\mathbf{A}^k) - \alpha_{\mathbf{A}}^k(\bar{\mathbf{H}}_{\mathbf{A}}^{\mathcal{I}_{\mathbf{A}}^k})^{-1}\mathrm{vec}(\nabla_{\mathbf{A}} f(\mathbf{A}^k, \mathbf{B}^k, \mathbf{C}^k))]_+ \quad (17)$$

can be realized, where $(\bar{\mathbf{H}}_{\mathbf{A}}^{\mathcal{I}_{\mathbf{A}}^k})^{-1}$ is the block diagonal matrix with the inverses of the matrices (16) on its block diagonal.

The step size is adjusted via backtracking (Armijo rule) on the projection arc; see [24]–[26] for details. An analogous procedure is followed for the remaining sub-problems. The steps of the proposed algorithm are tabulated as Algorithm 1. Its hierarchical iterative reweighted least squares (HIRLS)

Algorithm 1: NBTD-HIRLS algorithm

Data: $\mathcal{Y}, \lambda, R = R_{\text{ini}}, L = L_{\text{ini}}, \varepsilon > 0$
Result: Best approximation of $\mathcal{Y}$ in the sense of (5)
Initialize: $\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0 \geq \mathbf{0}$;
$k \leftarrow 0$;
repeat
  Compute $\mathbf{D}_1^k, \mathbf{D}_2^k$ from (9) and (10);
  $\mathbf{D}^k \leftarrow (\mathbf{D}_1^k \otimes \mathbf{I}_L)\mathbf{D}_2^k$;
  $\mathbf{P}^k \leftarrow \mathbf{B}^k \odot \mathbf{C}^k$;
  Compute $\nabla_{\mathbf{A}} f(\mathbf{A}^k, \mathbf{B}^k, \mathbf{C}^k)$ from (7);
  Find from (15) the index sets for the restricted variables in $\mathbf{A}^k$;
  Compute inverse Hessian, $(\bar{\mathbf{H}}_{\mathbf{A}}^{\mathcal{I}_{\mathbf{A}}^k})^{-1}$;
  Determine the step size, $\alpha_{\mathbf{A}}^k$, using Armijo backtracking on the feasible direction;
  Update $\mathbf{A}^k$ as in (17) ;
  $\mathbf{Q}^k \leftarrow \mathbf{C}^k \odot \mathbf{A}^{k+1}$;
  Update $\mathbf{B}^k$ in an analogous manner;
  $\mathbf{S}^k \leftarrow$
  $\left[\; (\mathbf{A}_1^{k+1} \odot_{\text{c}} \mathbf{B}_1^{k+1})\mathbf{1}_L \quad \cdots \quad (\mathbf{A}_R^{k+1} \odot_{\text{c}} \mathbf{B}_R^{k+1})\mathbf{1}_L \; \right]$;
  Update $\mathbf{C}^{k+1}$ in an analogous manner;
  $k \leftarrow k + 1$;
until *convergence*;

character lies in the reweighting effect of $\mathbf{D}_1$, which imposes *jointly* block sparsity on $\mathbf{A}$ and $\mathbf{B}$ and column sparsity on $\mathbf{C}$, hence helping in estimating R, and that of $\mathbf{D}_2$, which promotes column sparsity *jointly* to the corresponding blocks of $\mathbf{A}$ and $\mathbf{B}$, thus allowing the estimation of the L_r's. The algorithm enjoys the properties of low complexity (involving only small-sized matrix inversions) and fast convergence. Moreover, one may update the rank estimates in the course of the iterations and drop the redundant blocks and columns accordingly (pruning).

IV. Numerical Examples

In this section, the model selection and computation performance of the proposed algorithm is evaluated, using synthetic data, and in comparison with the popular MV-NTF method from [2]. We simulate $18 \times 18 \times 10$ nonnegative BTD tensors $\boldsymbol{\mathcal{X}}$ of $R = 3$ block terms with ranks 8, 6, 4, contaminated by additive noise, that is, $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} + \sigma\boldsymbol{\mathcal{N}}$, where $\boldsymbol{\mathcal{N}}$ contains independent and identically distributed (i.i.d) half-normal entries of zero location and unit scale parameters and σ is set so as to get a given signal-to-noise ratio (SNR), with SNR in dB defined as $\text{SNR} = 10\log_{10}(\|\boldsymbol{\mathcal{X}}\|_{\text{F}}^2/(\sigma^2\|\boldsymbol{\mathcal{N}}\|_{\text{F}}^2))$. The entries of the matrices $\mathbf{A}_r$ and $\mathbf{B}_r$ and the vectors $\mathbf{c}_r$ have been randomly chosen in a similar manner. Our figure of merit for the tensor approximation accuracy is the normalized mean squared error (NMSE) over the blocks, defined as $\text{NMSE}(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}) = \frac{1}{R}\sum_{r=1}^{R} \frac{\|\mathbf{A}_r\mathbf{B}_r^{\text{T}} \circ \mathbf{c}_r - \hat{\mathbf{A}}_r\hat{\mathbf{B}}_r^{\text{T}} \circ \hat{\mathbf{c}}_r\|_{\text{F}}^2}{\|\mathbf{A}_r\mathbf{B}_r^{\text{T}} \circ \mathbf{c}_r\|_{\text{F}}^2}$, where $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ and $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}})$ denote the true and the estimated factors, respectively. For the stopping criterion, we use the relative difference between two consecutive estimates of the relative reconstruction error (RE), $\|\boldsymbol{\mathcal{Y}} - \hat{\boldsymbol{\mathcal{X}}}\|_{\text{F}}/\|\boldsymbol{\mathcal{Y}}\|_{\text{F}}$. The algorithms stop either when this relative difference becomes less than 10^{-6} or a maximum of 300 iterations is reached. The regularization parameter λ is fine-tuned so that the minimum RE is attained. A useful rule of thumb is to so select it as to be proportional to (a small fraction of) the total number of unknowns and an estimate of σ. For each realization, both algorithms are randomly initialized for 10 times, and the best run among them, in terms of the RE, is kept. For NBTD-HIRLS, $R_{\text{ini}} = 10$ and $L_{\text{ini}} = 10$, while MV-NTF is assumed to know the true R, with all L_rs assumed equal to their maximum value, 8.

TABLE I
NMSE comparison in the presence of noise

	SNR (dB)			
	5	10	15	20
NBTD-HIRLS	0.1186	0.0839	0.0206	0.0061
MV-NTF	0.75	0.286	0.1413	0.0751

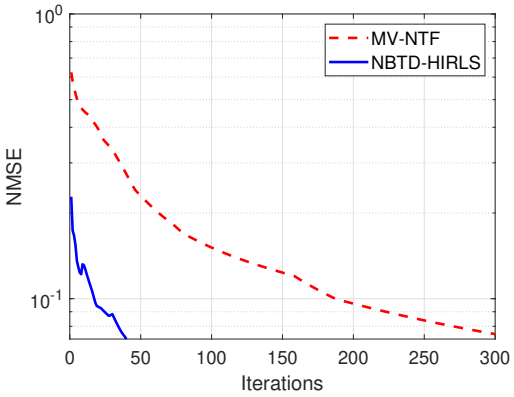

Fig. 1. NMSE of NBTD-HIRLS and MV-NTF vs. iterations for SNR=20 dB.

In Table I, we report the medians of the NMSEs obtained over 20 independent realizations and at various SNR levels. Clearly, and despite the a-priori model order information made available to MV-NTF, the proposed algorithm exhibits a much better estimation performance. This is also true for their convergence rates, as illustrated in Fig. 1 for the example of SNR=20 dB. The frequencies of the model ranks recovered, again at SNR=20 dB and normalized to [0,1], are depicted in Fig. 2. Note that the frequencies for the L_r estimates were only counted for the realizations where R was exactly revealed. Observe that R is recovered almost perfectly, with only a very small overestimate being rarely obtained. In practical applications, this would be much less severe than an underestimation. The L_rs are estimated with a small error, again expected to be harmless in NBTD applications like HSI unmixing. The evolution of the rank estimates with the

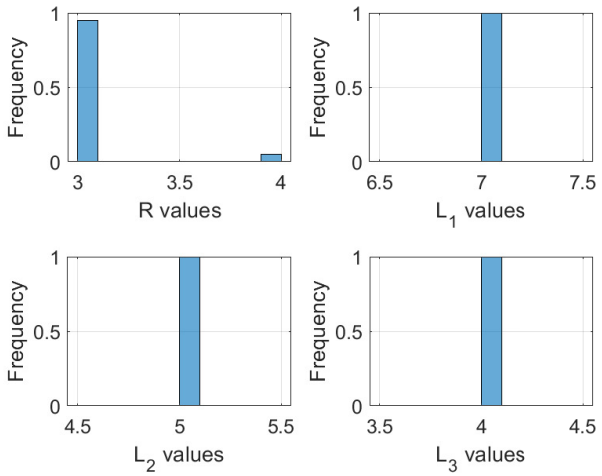

Fig. 2. Frequencies of the R and L_r estimates, at SNR=20 dB.

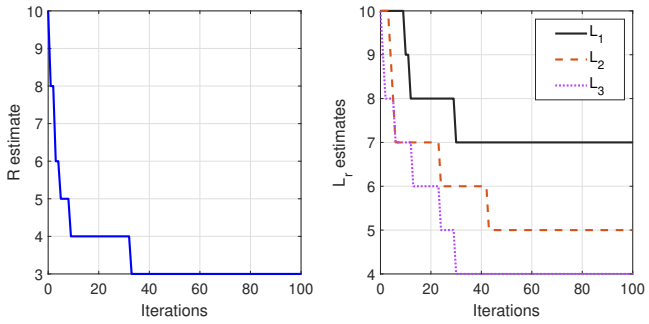

Fig. 3. Estimates of R (left) and the L_rs (right) vs. iterations, for a single realization. SNR=20 dB.

iterations is shown, for a single realization and at the same SNR level, in Fig. 3 and illustrates the possibilities for column and block pruning in the course of the algorithm.

V. Conclusion

The results for NBTD-HIRLS encourage us to further study this idea for tensor completion [22], [27] and extend the method to large-scale NTF [28]. Experimentation with real data, with focus on HSI unmixing (also including sum-to-one and other relevant constraints) and comparisons with the rest (ADMM-based) of the existing methods, is among the next steps to be taken in this context and the results will be reported elsewhere.

Finally, as pointed out in [29] for NCPD, there exists a subtle sign ambiguity issue, which does not seem to have been recognized widely in the NTF literature. We need to revisit our NBTD algorithm also from this perspective, neglected here for the sake of simplicity.

References

[1] G. Zhou *et al.*, "Nonnegative matrix and tensor factorizations," *IEEE Signal Process. Mag.*, May 2014.

[2] Y. Qian *et al.*, "Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, Mar. 2017.

[3] J. Spiegelberg, J. Rusz, and K. Pelckmans, "Tensor decompositions for the analysis of atomic resolution electron energy loss spectra," *Ultramicroscopy*, 2017.

[4] B. Sen and K. K. Parhi, "Constrained tensor decomposition optimization with applications to fMRI data analysis," in *Proc. ACSSC-2018*.

[5] T. Balasubramaniam, R. Nayak, and M. Bashar, "Understanding the spatio-temporal topic dynamics of Covid-19 using nonnegative tensor factorization: A case study," in *Proc. SSCI-2020*.

[6] Y. Qi, P. Comon, and L.-H. Lim, "Uniqueness of nonnegative tensor approximations," *IEEE Trans. Inf. Theory*, Apr. 2016.

[7] J. Kim, Y. He, and H. Park, "Algorithms for nonnegative matrix and tensor factorizations: a unified view based on block coordinate descent framework," *J. Glob. Optim.*, Feb. 2014.

[8] A. A. Rontogiannis, E. Kofidis, and P. V. Giampouras, "Block-term tensor decomposition: Model selection and computation," *IEEE J. Sel. Topics Signal Process.*, Apr. 2021.

[9] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms — Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, pp. 1033–1066, 2008.

[10] M. Hong *et al.*, "A unified algorithmic framework for block-structured optimization involving big data," *IEEE Signal Process. Mag.*, Jan. 2016.

[11] N. Li, "Variants of ALS on tensor decompositions and applications," Ph.D. dissertation, Clarkson University, 2013.

[12] N. Vervliet *et al.*, *Tensorlab user guide — Release 3.0*, Mar. 2016. [Online]. Available: http://www.tensorlab.net/userguide3.pdf

[13] J. H. de M. Goulart *et al.*, "Alternating group lasso for block-term tensor decomposition and application to ECG source separation," *IEEE Trans. Signal Process.*, Apr. 2020.

[14] G. Zhang *et al.*, "Hyperspectral super-resolution: A coupled nonnegative block-term tensor decomposition approach," in *Proc. CAMSAP-2019*.

[15] M. Ding *et al.*, "Factor-regularized nonnegative tensor decomposition for blind hyperspectral unmixing," in *Proc. IGARSS-2021*.

[16] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS-2000*.

[17] M. Welling and M. Weber, "Positive tensor factorization," *Pattern Recogn. Lett.*, 2001.

[18] B. Feng and J. Wang, "Constrained nonnegative tensor factorization for spectral unmixing of hyperspectral images: A case study of urban impervious surface extraction," *IEEE Geosci. Remote Sens. Lett.*, Apr. 2019.

[19] P. Zheng, H. Su, and Q. Du, "Sparse and low-rank constrained tensor factorization for hyperspectral image unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Jan. 2021.

[20] X. Feng, L. Han, and L. Dong, "Weighted group sparsity-constrained tensor factorization for hyperspectral unmixing," *Remote Sens.*, Jan. 2022.

[21] A. Beck and L. Tetruashvili, "On the convergence of block coordinate descent type methods," *SIAM J. Optim.*, 2013.

[22] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, 2013.

[23] S. Hansen, T. Plantenga, and T. G. Kolda, "Newton-based optimization for Kullback–Leibler nonnegative tensor factorizations," *Optim. Meth. Soft.*, 2015.

[24] D. P. Bertsekas, "Projected Newton methods for optimization problems with simple constraints," *SIAM J. Contr. Optim.*, 1982.

[25] M. Schmidt, D. Kim, and S. Sra, "Projected Newton-type methods in machine learning," in *Optimization for Machine Learning*, ser. Neural Information Processing Series, S. Sra, S. Nowozin, and S. J. Wright, Eds. The MIT Press, 2012, ch. 11, pp. 305–329.

[26] P. V. Giampouras, A. A. Rontogiannis, and K. D. Koutroumbas, "A projected Newton-type algorithm for non-negative matrix factorization with model order selection," in *Proc. ICASSP-2019*.

[27] P. V. Giampouras, A. A. Rontogiannis, and E. Kofidis, "Rank-revealing block-term decomposition for tensor completion," in *Proc. ICASSP-2021*.

[28] V. Vigneron *et al.*, "Non-negative sub-tensor ensemble factorization (NsTEF) algorithm. A new incremental tensor factorization for large data sets," *Signal Process.*, 2018.

[29] M. Jouni, M. D. Mura, and P. Comon, "Some issues in computing the CP decomposition of nonnegative tensors," in *Proc. LVA/ICA-2018*.