



HAL
open science

A Novel Non-Iterative Algorithm for Low-Multilinear-Rank Tensor Approximation

José Henrique de Morais Goulart, Pierre Comon

► **To cite this version:**

José Henrique de Morais Goulart, Pierre Comon. A Novel Non-Iterative Algorithm for Low-Multilinear-Rank Tensor Approximation. EUSIPCO 2017 - 25th European Signal Processing Conference, Aug 2017, Kos Island, Greece. hal-01531231

HAL Id: hal-01531231

<https://hal.science/hal-01531231v1>

Submitted on 1 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Novel Non-Iterative Algorithm for Low-Multilinear-Rank Tensor Approximation

José Henrique de Morais Goulart and Pierre Comon
Univ. Grenoble Alpes, CNRS, Gipsa-Lab, F-38000 Grenoble
Email: jose-henrique.de-morais-goulart@gipsa-lab.fr, pierre.comon@gipsa-lab.fr

Abstract—Low-rank tensor approximation algorithms are building blocks in tensor methods for signal processing. In particular, approximations of low multilinear rank (mrank) are of central importance in tensor subspace analysis. This paper proposes a novel non-iterative algorithm for computing a low-mrank approximation, termed sequential low-rank approximation and projection (SeLRAP). Our algorithm generalizes sequential rank-one approximation and projection (SeROAP), which aims at the rank-one case. For third-order mrank-(1,R,R) approximations, SeLRAP’s outputs are always at least as accurate as those of previously proposed methods. Our simulation results suggest that this is actually the case for the overwhelmingly majority of random third- and fourth-order tensors and several different mrank. Though the accuracy improvement is often small, we show it can make a large difference when repeatedly computing approximations, as happens, *e.g.*, in an iterative hard thresholding algorithm for tensor completion.

Index Terms—Multilinear rank, low-rank approximation, tensor, tensor completion.

I. INTRODUCTION

Tensor methods have rapidly become widespread in the signal processing community. A key task in these methods is the computation of low-rank tensors approximations. For instance, low-rank canonical polyadic decompositions of real-world data tensors are always approximative, due to noise and model mismatches [1]. Another recurrent example concerns approximations having low *multilinear rank* (mrank), a concept which is closely related to the Tucker decomposition. Because low-mrank approximation (LMA) is useful for subspace analysis and dimensionality reduction, applications abound, ranging from multidimensional harmonic retrieval [2] to biomedical signal classification [3]; see [4] for more examples. Both low-rank and low-mrank approximations reduce to the same one for matrices, in which case the truncated singular value decomposition (SVD) provides the optimal solution in the least-squares (LS) sense. In the case of higher-order tensors, they are equivalent only for rank one and are NP-hard [5].

Low-mrank approximation methods fall into two classes. The first class encompasses iterative algorithms which typically address a (nonlinear) LS problem [6], [4], [7]. Methods of the second class are non-iterative and focus on finding a reasonable, but suboptimal, low-mrank approximation within a finite number of steps. They are therefore more suitable

when some error is tolerated and/or computing cost is a limiting factor. In particular, they are useful for initializing methods of the first class and also for plugging into iterative algorithms which repeatedly resort to LMAs, such as iterative hard thresholding (IHT) schemes for tensor completion (TC) [8], [9]. The best-known method of this class is the truncated higher-order singular value decomposition (THOSVD), which projects each modal unfolding of the tensor onto its dominant subspace computed via an SVD [10]. Though it is suboptimal, its error is bounded by a multiple of the optimal value. The alternative proposed in [11], which we refer to as sequentially optimal modal projections (SeMP), proceeds similarly but computes the modal projectors in a sequential fashion. This leads to the same error bound, but at a smaller cost due to the reduced size of the SVDs. In particular, for mrank-(1, R, R) approximations it was shown to perform at least as well as THOSVD. Simulation results performed in [11] suggest that this superiority holds in most cases. For the particular case of rank-one approximations, [12] has come up with a two-stage algorithm called sequential rank-one approximation and projection (SeROAP), which first reduces dimensionality just like SeMP and then performs a sequence of projections for refining the approximation. For third-order tensors, [12] proved it performs at least as well as SeMP and, consequently, as THOSVD too.

This paper proposes a generalization of SeROAP, called sequential low-rank approximation and projection (SeLRAP), which handles arbitrary mrank. We show SeLRAP performs at least as well as SeMP for mrank-(1, R, R) approximations. Such a superiority was also observed in the overwhelming majority of our simulations for several other cases.

Notation: Vectors, matrices and tensors are denoted by bold lowercase, bold uppercase and calligraphic uppercase letters, respectively (*e.g.*, \mathbf{x} , \mathbf{X} , \mathcal{X}). The symbols \bullet_n , \otimes and \boxtimes stand for the mode- n , tensor and Kronecker products, respectively. Vector inequalities $\mathbf{x} \leq \mathbf{y}$ are meant entry-wise. $\mathbf{X}_{(n)} = (\mathcal{X})_{(n)}$ denotes the mode- n (flat) matrix unfolding of \mathcal{X} , and $\mathbf{X}_{:,1:R}$ denotes the submatrix of \mathbf{X} composed by its first R columns. \mathbf{I}_M stands for the $M \times M$ identity matrix.

II. STATE OF THE ART

The mrank of an N th-order tensor $\mathcal{Y} \in \bigotimes_{n=1}^N \mathbb{C}^{I_n}$ can be defined as the tuple $\mathbf{r} = (R_1, \dots, R_N)$ such that $R_n =$

This work is supported by the European Research Council under the European Programme FP7/2007-2013, Grant AdG-2013-320594 DECODA.

$\text{rank}(\mathbf{Y}_{\langle n \rangle})$. The numbers R_n are the smallest ones satisfying $\mathcal{Y} \in \bigotimes_{n=1}^N \mathcal{U}_n$ with $\dim(\mathcal{U}_n) = R_n$. Given a tensor $\mathcal{X} \in \bigotimes_{n=1}^N \mathbb{C}^{I_n}$, we consider the approximation problem:

$$\min_{\hat{\mathcal{X}} \in \bigotimes_{n=1}^N \mathcal{U}_n} \|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 \text{ subj. to } \begin{cases} \mathcal{U}_n \subseteq \mathbb{C}^{I_n} \\ \dim(\mathcal{U}_n) = R_n. \end{cases} \quad (1)$$

Below, we recapitulate the existing non-iterative approximate solutions to (1).

A. Truncated HOSVD

An equivalent formulation of (1) is

$$\min_{\hat{\mathbf{P}}^{(n)} \in \mathbb{C}^{I_n \times I_n}} \left\| \mathcal{X} - \mathcal{X} \underset{n=1}{\bullet}^N \hat{\mathbf{P}}^{(n)} \right\|_F^2,$$

where $\hat{\mathbf{P}}^{(n)}$ is an orthogonal projection onto an R_n -dimensional subspace \mathcal{U}_n . Introducing a telescoping sum inside the norm, this problem can be rewritten as [11]

$$\min_{\hat{\mathbf{P}}^{(n)} \in \mathbb{C}^{I_n \times I_n}} \sum_{n=1}^N \left\| \mathcal{X} \underset{m=1}{\bullet}^{n-1} \hat{\mathbf{P}}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2, \quad (2)$$

where $\hat{\mathbf{P}}_{\perp}^{(n)}$ projects onto the orthogonal complement of \mathcal{U}_n . As orthogonal projections are nonexpansive, the cost function of (2) is upper bounded by

$$\sum_{n=1}^N \left\| \mathcal{X} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2 = \sum_{n=1}^N \left\| \hat{\mathbf{P}}_{\perp}^{(n)} \mathbf{X}_{\langle n \rangle} \right\|_F^2. \quad (3)$$

It follows from the Eckart–Young theorem that (3) is minimized by projectors satisfying $\hat{\mathbf{P}}^{(n)} = \mathbf{U}_{:,1:R_n}^{(n)} (\mathbf{U}_{:,1:R_n}^{(n)})^H$, where $\mathbf{U}^{(n)}$ is the matrix of left singular vectors of $\mathbf{X}_{\langle n \rangle}$. One can thus approximate the solution of (2) by these projectors. This is equivalent to computing the HOSVD of \mathcal{X} , given by $\mathcal{X} = \mathcal{S} \bullet_{n=1}^N \mathbf{U}^{(n)}$, and then *truncating* each factor $\mathbf{U}^{(n)}$ at the R_n th column (and \mathcal{S} accordingly). This is the idea behind the THOSVD approach [10].

Now, because this choice gives projectors which are optimal when considered separately (but not jointly), any solution $\hat{\mathcal{X}}^*$ of (1) satisfies $\|\hat{\mathbf{P}}_{\perp}^{(n)} \mathbf{X}_{\langle n \rangle}\|_F^2 \leq \|\mathbf{X}_{\langle n \rangle} - \mathbf{X}_{\langle n \rangle}^*\|_F^2$. Plugging this expression into (3) shows the THOSVD yields LS error no greater than N times the optimal one.

Assuming the SVD of an $I \times M$ matrix requires $\phi(I, M) = \mathcal{O}(IM \min\{I, M\})$ operations, then THOSVD has cost¹ $\sum_{n=1}^N \phi(I_n, M_n) + \sum_{n=1}^N \mathcal{O}(L_n R_n I_n) + \sum_{n=1}^N \mathcal{O}(J_n R_n I_n)$, where $M_n \triangleq \prod_{m \neq n} I_m$, $L_n \triangleq R_1 \dots R_{n-1} I_{n+1} \dots I_N$ and $J_n \triangleq I_1 \dots I_{n-1} R_{n+1} \dots R_N$. The second and third summations correspond to the calculation of \mathcal{S} and $\hat{\mathcal{X}}$, respectively.

B. Sequentially optimal modal projections

Another way of approximately solving (2) is by seeking for its optimizers in a sequential fashion. This leads to the SeMP solution [11], which chooses projectors satisfying

$$\hat{\mathbf{P}}^{(n)} = \arg \min_{\hat{\mathbf{P}}^{(n)} \in \mathbb{C}^{I_n \times I_n}} \left\| \mathcal{X} \underset{m=1}{\bullet}^{n-1} \hat{\mathbf{P}}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2.$$

¹We assume that the contractions needed to calculate \mathcal{S} and $\hat{\mathcal{X}}$ are performed in the order $n = 1, \dots, N$. This simplifies the comparison with the other algorithms.

This is accomplished by computing SVDs of the unfoldings $\mathbf{W}_{\langle n \rangle}^{(n)}$ of tensors $\mathcal{W}^{(n)} \in \left(\bigotimes_{m=1}^{n-1} \mathbb{C}^{R_m} \right) \otimes \left(\bigotimes_{m=n}^N \mathbb{C}^{I_m} \right)$ given by $\mathcal{W}^{(1)} = \mathcal{X}$ and $\mathcal{W}^{(n)} = \mathcal{W}^{(n-1)} \bullet_{n-1} \bar{\mathbf{U}}^{(n-1)H}$ for $n = 2, \dots, N$, where $\bar{\mathbf{U}}^{(n)}$ holds the first R_n left singular vectors of $\mathbf{W}_{\langle n \rangle}^{(n)}$. Then, $\hat{\mathcal{X}} = \bar{\mathcal{W}}^{(N)} \bullet_{n=1}^N \bar{\mathbf{U}}^{(n)}$, where $\bar{\mathcal{W}}^{(N)} = \mathcal{W}^{(N)} \bullet_N \bar{\mathbf{U}}^{(N)H}$.

These SVD problems have smaller size than in the THOSVD, because the dimensions of the tensors $\mathcal{W}^{(n)}$ are gradually reduced. Thus, SeMP has smaller cost, given by

$$\sum_{n=1}^N [\phi(I_n, L_n) + \mathcal{O}(L_n R_n I_n)] + \sum_{n=1}^N \mathcal{O}(J_n R_n I_n).$$

Furthermore, its error is still subject to the same upper bound as the THOSVD [11].

It should be noted that the modes can be of course processed in any other order, which in general leads to different results.

C. Sequential rank-one approximation and projection

The SeROAP algorithm [12] handles the case $R_1 = \dots = R_N = 1$ by proceeding as follows:

1. *Order reduction stage:* Let $\mathbf{w}^{(1)} \triangleq \text{vec}(\mathcal{X}) \in \mathbb{C}^{I_N \dots I_1}$. For $n = 2, \dots, N$, recursively compute $\mathbf{w}^{(n)} \in \mathbb{C}^{I_N \dots I_n}$ via

$$\hat{\lambda} \in \mathbb{R}, \quad \hat{\mathbf{u}}^{(n)} \in \mathbb{C}^{I_{n-1}}, \quad \hat{\mathbf{w}}^{(n)} \in \mathbb{C}^{I_N \dots I_n} \quad \left\| \mathbf{w}^{(n-1)} - \hat{\lambda} \left(\hat{\mathbf{w}}^{(n)} \right)^* \boxtimes \hat{\mathbf{u}}^{(n)} \right\|_2^2,$$

whose solution is the dominant singular triplet of an $I_{n-1} \times I_N \dots I_n$ matrix obtained by “unvectorizing” $\mathbf{w}^{(n-1)}$.

2. *Projection stage:* Let $\mathbf{z}^{(N-1)} \triangleq \mathbf{w}^{(N-1)*} \boxtimes \mathbf{u}^{(N)} \in \mathbb{C}^{I_N \dots I_{N-1}}$. For $n = N-2, \dots, 1$, compute the orthogonal projection

$$\mathbf{z}^{(n)} = \mathbf{P}^{(n)} \mathbf{w}^{(n)} \in \mathbb{C}^{I_N \dots I_n}, \quad (4)$$

where $\mathbf{P}^{(n)} \triangleq \|\mathbf{z}^{(n+1)}\|_2^{-2} \left[\mathbf{z}^{(n+1)*} \mathbf{z}^{(n+1)T} \right] \boxtimes \mathbf{I}_{I_n}$.

Finally, the result $\hat{\mathcal{X}}$ is such that $\text{vec}(\hat{\mathcal{X}}) = \mathbf{z}^{(1)} \in \mathbb{C}^{I_N \dots I_1}$.

III. SEQUENTIAL LOW-MULTILINEAR-RANK APPROXIMATION AND PROJECTION

A. Algorithm

A generalization of SeROAP to any multilinear rank $\mathbf{r} = (R_1, \dots, R_N)$ goes along the following lines:

1. *Dimension reduction stage:* In this stage, a sequence of tensors $\mathcal{W}^{(n)} \in \left(\bigotimes_{m=1}^{n-1} \mathbb{C}^{R_m} \right) \otimes \left(\bigotimes_{m=n}^N \mathbb{C}^{I_m} \right)$ is computed in exactly the same way as in SeMP.

2. *Projection stage:* Here, one recursively obtains tensors $\mathcal{Z}^{(n)}$ of same dimensions as $\mathcal{W}^{(n)}$ by performing a sequence of orthogonal projections. Specifically, this calculation begins with $\mathcal{Z}^{(N)} = \bar{\mathcal{W}}^{(N)} \bullet_N \bar{\mathbf{U}}^{(N)}$, and for $n = N-1, \dots, 1$,

$$\mathcal{Z}_{\langle n \rangle}^{(n)} = \mathbf{W}_{\langle n \rangle}^{(n)} \mathcal{Z}_{\langle n \rangle}^{(n+1)H} \left(\mathbf{Z}_{\langle n \rangle}^{(n+1)} \mathbf{Z}_{\langle n \rangle}^{(n+1)H} \right)^{-1} \mathbf{Z}_{\langle n \rangle}^{(n+1)}. \quad (5)$$

The desired mrank- (R_1, \dots, R_N) approximation is $\hat{\mathcal{X}} = \mathcal{Z}^{(1)}$.

The projection stage of SeLRAP can be seen as an attempt of refining the approximation computed by SeMP. Also, it is not hard to check that the above procedure reduces to SeROAP

when $R_1 = \dots = R_N = 1$. In this particular case, the vectors $\mathbf{w}^{(n)}$ and $\mathbf{z}^{(n)}$ of Sec. II-C are vectorized versions of $\mathbf{W}^{(n)}$ and $\mathbf{Z}^{(n)}$, respectively, and (4) is equivalent to (5).

In practice, (5) can be performed with the aid of an orthonormal basis for the row space of $\mathbf{Z}_{(n)}^{(n+1)}$, obtained from a QR decomposition or an SVD. The constructed projector must have the same rank as $\mathbf{Z}_{(n)}^{(n+1)}$, which can be smaller than R_n . This corresponds to replacing the inverse matrix of (5) by the Moore-Penrose pseudo-inverse. Computing the basis costs $\mathcal{O}(L_n R_n^2)$ flops (assuming $L_n \geq R_n$), since $\mathbf{Z}_{(n)}^{(n+1)}$ has dimensions $R_n \times L_n$, while performing the projection costs $\mathcal{O}(L_n I_n R_n)$. Thus, SeLRAP has the overall complexity

$$\sum_{n=1}^N [\phi(I_n, L_n) + \mathcal{O}(L_n I_n R_n)] + \mathcal{O}(L_N R_N I_N) + \phi_{\text{proj}},$$

where $\phi_{\text{proj}} = \sum_{n=1}^{N-1} [\mathcal{O}(L_n R_n^2) + \mathcal{O}(L_n I_n R_n)]$.

B. Fulfillment of rank constraint

We now show that the approximation delivered by SeLRAP actually meets the desired mrank constraint.

Lemma 1. Let $\mathcal{X} \in \bigotimes_{m=1}^N \mathbb{C}^{K_m}$ and define the tensor $\mathcal{P} \in \left(\bigotimes_{m=1}^{n-1} \mathbb{C}^{K_m} \right) \otimes \mathbb{C}^{R_n} \otimes \left(\bigotimes_{m=n+1}^N \mathbb{C}^{K_m} \right)$, for some $n \in \mathbb{N}_N$. If $\text{mrank}(\mathcal{P}) = (R_1, \dots, R_N)$, then

$$\mathbf{Y}_{\langle n \rangle} = \mathbf{X}_{\langle n \rangle} \mathbf{P}_{\langle n \rangle}^H (\mathbf{P}_{\langle n \rangle} \mathbf{P}_{\langle n \rangle}^H)^{-1} \mathbf{P}_{\langle n \rangle}$$

can be assimilated to the mode- n unfolding of $\mathcal{Y} \in \bigotimes_{m=1}^N \mathbb{C}^{K_m}$ satisfying $\text{mrank}(\mathcal{Y}) \leq (R_1, \dots, R_N)$.

Proof. Since $\text{mrank}(\mathcal{P}) = (R_1, \dots, R_N)$, there exist $\mathcal{G} \in \bigotimes_{m=1}^N \mathbb{C}^{R_m}$ and $\mathbf{U}^{(m)} \in \mathbb{C}^{K_m \times R_m}$ for $m \in \mathbb{N}_N \setminus \{n\}$ such that $\mathbf{U}^{(m)}$ has orthonormal columns and $\mathbf{P}_{\langle n \rangle} = \mathbf{G}_{\langle n \rangle} \mathbf{U}^T$, where $\mathbf{U} \triangleq \mathbf{U}^{(N)} \boxtimes \dots \boxtimes \mathbf{U}^{(n+1)} \boxtimes \mathbf{U}^{(n-1)} \boxtimes \dots \boxtimes \mathbf{U}^{(1)}$. Hence, $\mathbf{Y}_{\langle n \rangle} = \mathbf{X}_{\langle n \rangle} \mathbf{U}^* \mathbf{G}_{\langle n \rangle}^H (\mathbf{G}_{\langle n \rangle} \mathbf{G}_{\langle n \rangle}^H)^{-1} \mathbf{G}_{\langle n \rangle} \mathbf{U}^T$, which implies $\text{rank}(\mathbf{Y}_{\langle n \rangle}) \leq R_n$. Defining $\mathcal{C} \in \left(\bigotimes_{m=1}^{n-1} \mathbb{C}^{R_m} \right) \otimes \mathbb{C}^{K_n} \otimes \left(\bigotimes_{m=n+1}^N \mathbb{C}^{R_m} \right)$ as $\mathbf{C}_{\langle n \rangle} = \mathbf{X}_{\langle n \rangle} \mathbf{U}^* \mathbf{G}_{\langle n \rangle}^H (\mathbf{G}_{\langle n \rangle} \mathbf{G}_{\langle n \rangle}^H)^{-1} \mathbf{G}_{\langle n \rangle}$, it follows that $\mathcal{Y} = \mathcal{C} \bullet_{m \neq n} \mathbf{U}^{(m)}$, implying $\text{mrank}(\mathcal{Y}) \leq (R_1, \dots, R_N)$. \square

Proposition 2. The approximation $\hat{\mathcal{X}}$ produced by SeLRAP satisfies $\text{mrank}(\hat{\mathcal{X}}) \leq \mathbf{r} = (R_1, \dots, R_N)$.

Proof. The definition of $\mathcal{Z}^{(N)}$ implies $\text{mrank}(\mathcal{Z}^{(N)}) \leq \mathbf{r}$. From (5), it thus follows that $\text{mrank}(\mathcal{Z}^{(n)}) \leq \mathbf{r}$ for all n , due to Lemma 1. As $\hat{\mathcal{X}} = \mathcal{Z}^{(1)}$, we have $\text{mrank}(\hat{\mathcal{X}}) \leq \mathbf{r}$. \square

C. Comparison with SeMP

In the following, we analytically compare the quadratic errors incurred by SeMP and SeLRAP for third-order tensors. Let $\mathbf{r} = (R_1, R_2, R_3)$ and denote by $\hat{\mathcal{X}}_{\text{SeMP}} = \mathcal{S} \bullet_{n=1}^3 \bar{\mathbf{U}}^{(n)}$ the approximation delivered by SeMP, where $\bar{\mathbf{U}}^{(n)}$ is as defined in Sec. II-B. Because $\mathcal{S} = \mathcal{X} \bullet_{n=1}^3 \bar{\mathbf{U}}^{(n)H}$, the resulting quadratic error can be written as

$$\begin{aligned} \varepsilon_{\text{SeMP}} &\triangleq \|\mathcal{X} - \hat{\mathcal{X}}_{\text{SeMP}}\|_F^2 = \|\mathcal{X}\|_F^2 - \left\| \mathcal{X} \bullet_{n=1}^3 \bar{\mathbf{U}}^{(n)H} \right\|_F^2 \\ &= \|\mathcal{X}\|_F^2 - \left\| \bar{\mathbf{U}}^{(1)H} \mathbf{X}_{\langle 1 \rangle} \left(\bar{\mathbf{U}}^{(3)*} \boxtimes \bar{\mathbf{U}}^{(2)*} \right) \right\|_F^2. \end{aligned}$$

Since $\bar{\mathbf{U}}^{(1)}$ holds the first R_1 left singular vectors of $\mathbf{X}_{\langle 1 \rangle}$,

$$\varepsilon_{\text{SeMP}} = \|\mathcal{X}\|_F^2 - \left\| \bar{\Sigma}^{(1)} \bar{\mathbf{V}}^{(1)H} \left(\bar{\mathbf{U}}^{(3)*} \boxtimes \bar{\mathbf{U}}^{(2)*} \right) \right\|_F^2,$$

where the columns of $\bar{\mathbf{V}}^{(n)}$ are the first R_n right singular vectors of $\mathbf{X}_{\langle 1 \rangle} = \mathbf{W}_{\langle 1 \rangle}^{(1)}$, while $\bar{\Sigma}^{(1)}$ contains the corresponding singular values in its diagonal. This is a direct generalization of the expression derived in [12] for the case $\mathbf{r} = (1, 1, 1)$.

A similar expression can be derived for SeLRAP. First, define the orthogonal projector

$$\mathbf{P} \triangleq \mathbf{Z}_{\langle 1 \rangle}^{(2)H} \left(\mathbf{Z}_{\langle 1 \rangle}^{(2)} \mathbf{Z}_{\langle 1 \rangle}^{(2)H} \right)^{-1} \mathbf{Z}_{\langle 1 \rangle}^{(2)}.$$

Using this definition along with (5) and the identities $(\hat{\mathcal{X}}_{\text{SeLRAP}})_{\langle 1 \rangle} = \mathbf{Z}_{\langle 1 \rangle}^{(1)}$ and $\mathbf{X}_{\langle 1 \rangle} = \mathbf{W}_{\langle 1 \rangle}^{(1)}$, we derive

$$\varepsilon_{\text{SeLRAP}} \triangleq \|\mathcal{X} - \hat{\mathcal{X}}_{\text{SeLRAP}}\|_F^2 = \|\mathcal{X}\|_F^2 - \left\| \mathbf{W}_{\langle 1 \rangle}^{(1)} \mathbf{P} \right\|_F^2. \quad (6)$$

Writing $\mathbf{W}_{\langle 1 \rangle}^{(n)} = \bar{\mathbf{U}}^{(n)} \bar{\Sigma}^{(n)} \bar{\mathbf{V}}^{(n)H} + \mathbf{E}^{(n)}$, the second norm in (6) can be rewritten as

$$\begin{aligned} \left\| \mathbf{W}_{\langle 1 \rangle}^{(1)} \mathbf{P} \right\|_F^2 &= \text{Tr} \left\{ \mathbf{P} \mathbf{W}_{\langle 1 \rangle}^{(1)H} \mathbf{W}_{\langle 1 \rangle}^{(1)} \mathbf{P} \right\} \\ &= \left\| \bar{\Sigma}^{(1)} \bar{\mathbf{V}}^{(1)H} \mathbf{P} \right\|_F^2 + \left\| \mathbf{E}^{(1)} \mathbf{P} \right\|_F^2. \end{aligned}$$

Plugging the result into (6), we have

$$\begin{aligned} \varepsilon_{\text{SeLRAP}} &= \|\mathcal{X}\|_F^2 - \left\| \bar{\Sigma}^{(1)} \bar{\mathbf{V}}^{(1)H} \mathbf{P} \right\|_F^2 - \left\| \mathbf{E}^{(1)} \mathbf{P} \right\|_F^2 \\ &\leq \|\mathcal{X}\|_F^2 - \left\| \bar{\Sigma}^{(1)} \bar{\mathbf{V}}^{(1)H} \mathbf{P} \right\|_F^2. \end{aligned}$$

Thus, a sufficient condition for having $\varepsilon_{\text{SeLRAP}} \leq \varepsilon_{\text{SeMP}}$ is

$$\left\| \bar{\Sigma}^{(1)} \bar{\mathbf{V}}^{(1)H} \mathbf{P} \right\|_F^2 \geq \left\| \bar{\Sigma}^{(1)} \bar{\mathbf{V}}^{(1)H} \left(\bar{\mathbf{U}}^{(3)*} \boxtimes \bar{\mathbf{U}}^{(2)*} \right) \right\|_F^2. \quad (7)$$

It turns out, though, that a general explicit expression for \mathbf{P} is quite complicated. We thus focus on the case where $R_1 = 1$, which implies $R_2 = R_3 = R$. This can be easily seen from the mode-2 and mode-3 unfoldings of an mrank- $(1, R_2, R_3)$ Tucker model. For this case, the following result holds.

Theorem 3. Let $\mathcal{X} \in \bigotimes_{n=1}^3 \mathbb{C}^{I_n}$ and denote by $\hat{\mathcal{X}}_{\text{SeLRAP}}$ and $\hat{\mathcal{X}}_{\text{SeMP}}$ the mrank- $(1, R, R)$ approximations of \mathcal{X} produced by SeLRAP and SeMP, respectively, by processing the modes in the natural order $(1, 2, 3)$. Then,

$$\left\| \mathcal{X} - \hat{\mathcal{X}}_{\text{SeLRAP}} \right\|_F^2 \leq \left\| \mathcal{X} - \hat{\mathcal{X}}_{\text{SeMP}} \right\|_F^2. \quad (8)$$

Proof. First, SeLRAP computes the SVDs

$$\mathbf{W}_{\langle 1 \rangle}^{(1)} = \bar{\sigma}^{(1)} \bar{\mathbf{u}}^{(1)} \bar{\mathbf{v}}^{(1)H} + \mathbf{E}^{(1)} \in \mathbb{C}^{I_1 \times I_3 I_2},$$

$$\mathbf{W}_{\langle 2 \rangle}^{(2)} = \bar{\mathbf{U}}^{(2)} \bar{\Sigma}^{(2)} \bar{\mathbf{V}}^{(2)H} + \mathbf{E}^{(2)} \in \mathbb{C}^{I_2 \times I_3},$$

where $\mathbf{W}_{(1)}^{(2)} = \bar{\sigma}^{(1)} \bar{\mathbf{v}}^{(1)H}$. Observe that, for $R_1 = 1$,

$$\text{vec} \left(\mathbf{W}_{(1)}^{(2)} \right) = \bar{\sigma}^{(1)} \bar{\mathbf{v}}^{(1)*} = \text{vec} \left(\mathbf{W}_{(2)}^{(2)} \right) \quad (9)$$

$$= \text{vec} \left(\bar{\mathbf{U}}^{(2)} \bar{\Sigma}^{(2)} \bar{\mathbf{V}}^{(2)H} + \mathbf{E}^{(2)} \right) \quad (10)$$

and $\mathbf{W}_{(3)}^{(3)} = \mathbf{W}_{(2)}^{(3)T} = \mathbf{W}_{(2)}^{(2)T} \bar{\mathbf{U}}^{(2)*}$. Hence, the SVD of $\mathbf{W}_{(3)}^{(3)}$ comes “for free,” being given by $\mathbf{W}_{(3)}^{(3)} = \bar{\mathbf{V}}^{(2)*} \bar{\Sigma}^{(2)} = \bar{\mathbf{U}}^{(3)} \bar{\Sigma}^{(3)}$, i.e., $\bar{\mathbf{U}}^{(3)} = \bar{\mathbf{V}}^{(2)*}$. Now, in the projection stage,

$$\mathbf{Z}_{(3)}^{(3)} = \mathbf{W}_{(3)}^{(3)} = \bar{\mathbf{U}}^{(3)} \bar{\Sigma}^{(3)} \in \mathbb{C}^{I_3 \times R}, \quad (11)$$

because $\text{rank}(\mathbf{W}_{(3)}^{(3)}) \leq R$. Furthermore, $\mathbf{Z}_{(2)}^{(3)} = \mathbf{Z}_{(3)}^{(3)T}$. Thus, plugging (11) into (5) for $n = 2$ we obtain

$$\mathbf{Z}_{(2)}^{(2)} = \mathbf{W}_{(2)}^{(2)} \bar{\mathbf{V}}^{(2)} \bar{\mathbf{V}}^{(2)H} = \bar{\mathbf{U}}^{(2)} \bar{\Sigma}^{(2)} \bar{\mathbf{V}}^{(2)H} \in \mathbb{C}^{I_2 \times I_3}.$$

Since $\mathbf{Z}_{(1)}^{(2)} = \text{vec} \left(\mathbf{Z}_{(2)}^{(2)} \right)^T$, using the property $\text{vec}(\mathbf{ABC}^T) = (\mathbf{C} \boxtimes \mathbf{A}) \text{vec}(\mathbf{B})$ we have

$$\mathbf{Z}_{(1)}^{(2)} = \text{vec} \left(\bar{\Sigma}^{(2)} \right)^T \left(\bar{\mathbf{U}}^{(3)} \boxtimes \bar{\mathbf{U}}^{(2)} \right)^T \in \mathbb{C}^{1 \times I_3 I_2},$$

which implies $\mathbf{P} = \|\bar{\sigma}^{(2)}\|_2^{-2} \mathbf{U}^* \bar{\sigma}^{(2)} \bar{\sigma}^{(2)T} \mathbf{U}^T$, where $\bar{\sigma}^{(2)} \triangleq \text{vec} \left(\bar{\Sigma}^{(2)} \right)$ and $\mathbf{U} \triangleq \bar{\mathbf{U}}^{(3)} \boxtimes \bar{\mathbf{U}}^{(2)}$. Applying these definitions to (9)–(10), we have also

$$\bar{\sigma}^{(1)} \bar{\mathbf{v}}^{(1)*} = \mathbf{U} \bar{\sigma}^{(2)} + \text{vec} \left(\mathbf{E}^{(2)} \right). \quad (12)$$

Finally, in view of the derived expressions, computing the left-hand side of (7) for $\mathbf{r} = (1, R, R)$ yields

$$\begin{aligned} \left\| \bar{\sigma}^{(1)} \bar{\mathbf{v}}^{(1)H} \mathbf{P} \right\|_2^2 &= \left(\bar{\sigma}^{(1)} \right)^2 \bar{\mathbf{v}}^{(1)H} \mathbf{P} \bar{\mathbf{v}}^{(1)} \\ &= \left(\bar{\sigma}^{(1)} \right)^2 \|\bar{\sigma}^{(2)}\|_2^{-2} \left| \bar{\sigma}^{(2)T} \mathbf{U}^T \bar{\mathbf{v}}^{(1)} \right|^2 \end{aligned} \quad (13)$$

Due to (12), $\mathbf{U}^T \bar{\mathbf{v}}^{(1)} = \left(\bar{\sigma}^{(1)} \right)^{-1} \left[\bar{\sigma}^{(2)} + \mathbf{U}^T \text{vec} \left(\mathbf{E}^{(2)} \right)^* \right]$. But, by definition of the SVD, the column space of $\mathbf{E}^{(2)}$ is orthogonal to $\bar{\mathbf{U}}^{(2)}$ while its row space is orthogonal to $\bar{\mathbf{V}}^{(2)*} = \bar{\mathbf{U}}^{(3)}$. Thus, it turns out that $\mathbf{U}^T \text{vec} \left(\mathbf{E}^{(2)} \right)^* = \mathbf{0}$, leading to $\mathbf{U}^T \bar{\mathbf{v}}^{(1)} = \left(\bar{\sigma}^{(1)} \right)^{-1} \bar{\sigma}^{(2)}$. Substituting this expression into (13) yields $\left\| \bar{\sigma}^{(1)} \bar{\mathbf{v}}^{(1)H} \mathbf{P} \right\|_2^2 = \|\bar{\sigma}^{(2)}\|_2^2$. On the other hand, for $R_1 = 1$ the right-hand side of (7) is given by $\left(\bar{\sigma}^{(1)} \right)^2 \left\| \bar{\mathbf{v}}^{(1)H} \mathbf{U}^* \right\|_2^2 = \|\bar{\sigma}^{(2)}\|_2^2$. Therefore, (7) holds with equality, implying (8). \square

Note that, for $\mathbf{r} = (1, R, R)$, $\mathbf{W}_{(2)}^{(2)}$ is an $I_2 \times I_3$ matrix whose *best* rank- R approximation is computed by both SeMP and SeLRAP. Since this (vectorized) approximation might be correlated with the rows of $\mathbf{E}^{(1)}$, the projection stage of SeLRAP might improve the approximation accuracy. At worst, there is no such correlation and the error stays the same.

Theorem 3 generalizes Theorem 1 of [12]. Furthermore, together with Theorem 7.2 of [11], it implies the following.

TABLE I

STATISTICS OF Δ AND VALUES OF Ψ FOR TENSORS OF ORDER 3 AND 4

Scenario	THOSVD			SeMP		
	μ_Δ	σ_Δ	Ψ	μ_Δ	σ_Δ	Ψ
1)	3.37e-02	5.83e-03	0.21	1.34e-03	7.74e-04	0.88
2)	4.98e-02	5.87e-03	0.50	1.43e-03	7.14e-04	0.92
3)	2.80e-02	7.07e-03	0.71	1.60e-02	5.43e-03	1.03
4)	4.41e-02	6.19e-03	1.00	3.48e-02	4.74e-03	1.20
5)	2.65e-02	1.78e-03	0.82	6.53e-03	4.78e-04	1.11
6)	1.29e-02	7.40e-04	0.20	8.93e-04	1.51e-04	0.74
7)	2.11e-02	1.43e-03	0.45	1.49e-03	1.80e-04	0.86
8)	9.56e-03	8.19e-04	0.69	6.37e-03	5.37e-04	1.18
9)	1.33e-02	1.48e-03	0.97	6.80e-03	5.03e-04	1.29
10)	5.72e-03	3.27e-04	0.71	6.63e-04	3.91e-05	1.12

Corollary 4. Let $\mathcal{X} \in \otimes_{n=1}^3 \mathbb{C}^{I_n}$ and denote by $\hat{\mathcal{X}}_{\text{SeLRAP}}$, $\hat{\mathcal{X}}_{\text{SeMP}}$ and $\hat{\mathcal{X}}_{\text{THOSVD}}$ the $\text{mr}\text{rank-}(1, R, R)$ approximations of \mathcal{X} produced by SeLRAP, SeMP and THOSVD, respectively. Suppose that the modes are processed in the natural order (1,2,3) by both SeLRAP and SeMP. Then,

$$\left\| \mathcal{X} - \hat{\mathcal{X}}_{\text{SeLRAP}} \right\|_F^2 \leq \left\| \mathcal{X} - \hat{\mathcal{X}}_{\text{SeMP}} \right\|_F^2 \leq \left\| \mathcal{X} - \hat{\mathcal{X}}_{\text{THOSVD}} \right\|_F^2.$$

The same results evidently apply to the cases $\mathbf{r} = (R, 1, R)$ and $\mathbf{r} = (R, R, 1)$, as long as the mode associated with the component $R_n = 1$ be the first one to be processed.

IV. NUMERICAL RESULTS

In the following experiments, the modes are always processed in the natural order in SeLRAP and SeMP, for simplicity. The reported computing times were measured in Matlab R2013a running on a Intel Xeon ES-2630v2 2.60 GHz with 32 GB RAM 1866 MHz. For conciseness, the notation $\mathbf{i} = (I_1, \dots, I_N)$ specifies the tensor dimensions in each scenario.

A. Performance comparison

Following [12], we compare SeLRAP with THOSVD and SeMP by computing

$$\Delta = 1 - \|\mathcal{X} - \hat{\mathcal{X}}_{\text{SeLRAP}}\|_F \|\mathcal{X} - \hat{\mathcal{X}}_{\text{Alg}}\|_F^{-1},$$

with $\text{Alg} \in \{\text{THOSVD}, \text{SeMP}\}$, for 10^5 realizations of third- and fourth-order tensors having entries whose real and imaginary parts are drawn from $(-1, 1)$. This is done in 10 different scenarios: 1) $\mathbf{i} = (4, 8, 20)$, $\mathbf{r} = (1, 2, 2)$; 2) $\mathbf{i} = (4, 8, 20)$, $\mathbf{r} = (2, 4, 8)$; 3) $\mathbf{i} = (20, 8, 4)$, $\mathbf{r} = (2, 2, 1)$; 4) $\mathbf{i} = (20, 8, 4)$, $\mathbf{r} = (8, 4, 2)$; 5) $\mathbf{i} = (20, 20, 20)$, $\mathbf{r} = (10, 10, 10)$; 6) $\mathbf{i} = (4, 8, 16, 20)$, $\mathbf{r} = (1, 2, 4, 5)$; 7) $\mathbf{i} = (4, 8, 16, 20)$, $\mathbf{r} = (2, 4, 8, 10)$; 8) $\mathbf{i} = (20, 16, 8, 4)$, $\mathbf{r} = (5, 4, 2, 1)$; 9) $\mathbf{i} = (20, 16, 8, 4)$, $\mathbf{r} = (10, 8, 4, 2)$; 10) $\mathbf{i} = (20, 20, 20, 20)$, $\mathbf{r} = (10, 10, 10, 10)$. Table I displays the mean (μ_Δ) and standard deviation (σ_Δ) of Δ , and also the ratio of SeLRAP’s flop count to each other method’s count, denoted by Ψ . In all these scenarios except for 3), Δ was *always* positive for both THOSVD and SeMP. Note that Corollary 4 only guarantees that for scenario 1). Moreover, with respect to THOSVD, Δ was negative for only 0.01% of the realizations in scenario 3), while for SeMP that proportion was of 0.30%. Finally, THOSVD’s flop count is always higher than SeLRAP’s, except

for scenario 4), while this is true for SeMP in the scenarios where $I_n < I_{n+1}$ for all $n \in \mathbb{N}_{N-1}$.

B. Application to low-rank tensor completion

One way of performing tensor completion under a low-rank assumption is by minimizing (1) only over the observed entries of \mathcal{X} . In an IHT scheme, at each iteration one updates the current estimate with a gradient step and then computes a low-mrank approximation of the result. Here, we compare three IHT schemes: (i) TIHT [8] uses THOSVD, (ii) SeMPIHT [9] uses SeMP and (iii) SeLRAPIHT uses SeLRAP. The step size is computed as proposed in [13], with a unitary initial candidate. This comparison concerns two scenarios, as follows.

First, $N_r = 300$ random real tensors \mathcal{X} with dimensions $\mathbf{i} = (20, 20, 20, 20)$ and mrank $\mathbf{r} = (8, 8, 8, 8)$ are generated and normalized so that $\|\mathcal{X}\|_F = 1$. Then, we add a Gaussian noise tensor, yielding $\mathcal{Y} = \mathcal{X} + 10^{-2}\mathcal{N}$, where $\|\mathcal{N}\|_F = 1$. We randomly sample 5% of the entries of \mathcal{Y} (uniformly) and then run 500 iterations of each IHT algorithm to complete \mathcal{X} . Fig. 1(a) plots the average squared error $\text{SE}(\hat{\mathcal{X}}; \mathcal{X}) = \|\mathcal{X} - \hat{\mathcal{X}}\|_F^2$ attained at each iteration, against the average elapsed time until completing it. Clearly, SeLRAPIHT outperforms its competitors significantly, due to SeLRAP's higher precision.

In the second scenario, N_r random real tensors are generated as $\mathcal{X} = \mathcal{G} \bullet_{n=1}^4 (\mathbf{Q}_n \text{Diag}(1, 2^{-\varphi}, \dots, 20^{-\varphi}))$ (and normalized), where \mathcal{G} has dimensions $20 \times 20 \times 20 \times 20$ and standard normal entries, \mathbf{Q}_n is a random orthogonal matrix and $\varphi = 2$ controls the decay of the singular values of the unfoldings of \mathcal{X} . Again, 5% of the entries of \mathcal{X} are randomly sampled. The IHT algorithms are then applied to estimate an mrank-(8,8,8,8) approximation of \mathcal{X} from these samples. To improve performance, we employ the gradual rank increase heuristic of [9], starting with mrank $(1, 1, 1, 1)$ and then increasing all mrank components every 5 iterations. All algorithms are run for 80 iterations. As shown in Fig. 1(b), SeLRAP again brings a significant improvement.

V. CONCLUSION

We have proposed a novel low-mrank approximation algorithm, SeLRAP, by generalizing SeROAP to any target mrank. For third-order tensors, the guaranteed superiority of SeROAP over THOSVD and SeMP still holds if the mrank contains a unitary component whose mode is the first one to be processed. Our numerical results indicate that this is the case also for other ranks and for fourth-order tensors, at least with very high probability. Though the improvement in accuracy is usually small, it can yield an important advantage when several approximations are successively computed, as happens in iterative hard thresholding algorithms for tensor completion. In this application, we have observed a significant superiority of SeLRAP over other alternatives when the number of observed tensor entries is very small, *i.e.*, in challenging TC scenarios.

REFERENCES

[1] P. Comon, "Tensors: a brief introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 44–53, 2014.

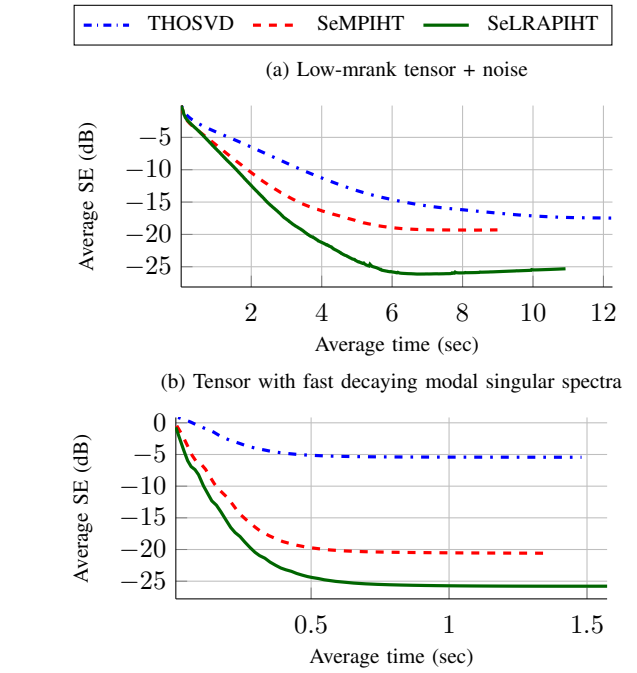


Fig. 1. Performance of IHT algorithms in the completion of a $20 \times 20 \times 20 \times 20$ tensor, with target mrank $\mathbf{r} = (8, 8, 8, 8)$ and 5% sampled entries.

[2] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3198–3213, 2008.

[3] A. Wolczowski and R. Zdunek, "Electromyography and mechanomyography signal recognition: Experimental analysis using multi-way array decomposition methods," *Biocybern. Biomed. Eng.*, vol. 37, no. 1, pp. 103–113, 2017.

[4] M. Ishteva, P.-A. Absil, S. Van Huffel, and L. De Lathauwer, "On the best low multilinear rank approximation of higher-order tensors," in *Recent Adv. Opt. Applicat. Eng.* Springer, 2010, pp. 145–164.

[5] C. J. Hillar and L.-H. Lim, "Most tensor problems are NP-hard," *J. ACM (JACM)*, vol. 60, no. 6, pp. 45:1–45:39, 2013.

[6] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank- (r_1, r_2, \dots, r_n) approximation of higher-order tensors," *SIAM J. Matrix Anal. Applicat.*, vol. 21, no. 4, pp. 1324–1342, 2000.

[7] M. Ishteva, P.-A. Absil, and P. Van Dooren, "Jacobi algorithm for the best low multilinear rank approximation of symmetric tensors," *SIAM J. Matrix Anal. Applicat.*, vol. 34, no. 2, pp. 651–672, 2013.

[8] H. Rauhut, R. Schneider, and Z. Stojanac, "Low rank tensor recovery via iterative hard thresholding," in *Proc. 10th Int. Conf. Sampling Theory Applicat.*, Bremen, Germany, 2013, pp. 21–24.

[9] J. H. de M. Goulart and G. Favier, "Low-rank tensor recovery using sequentially optimal modal projections in iterative hard thresholding (SeMPIHT)," *SIAM J. Sci. Comput.*, vol. 39, no. 3, pp. A860–A889, 2017.

[10] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Applicat.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[11] N. Vannieuwenhoven, R. Vandebril, and K. Meerbergen, "A new truncation strategy for the higher-order singular value decomposition," *SIAM J. Sci. Comput.*, vol. 34, no. 2, pp. A1027–A1052, 2012.

[12] A. P. da Silva, P. Comon, and A. de Almeida, "A finite algorithm to compute rank-1 tensor approximations," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 959–963, 2016.

[13] J. H. de M. Goulart and G. Favier, "An iterative hard thresholding algorithm with improved convergence for low-rank tensor recovery," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Nice, France, Sep. 2015, pp. 1701–1705.