Non-iterative low-multilinear-rank tensor approximation with application to decomposition in rank-(1,L,L) terms

José Henrique de Morais Goulart, Pierre Comon

NON-ITERATIVE LOW-MULTILINEAR-RANK TENSOR APPROXIMATION WITH APPLICATION TO DECOMPOSITION IN RANK-$(1, L, L)$ TERMS*

JOSÉ HENRIQUE DE MORAIS GOULART† AND PIERRE COMON†

Abstract. Computing low-rank approximations is one of the most important and well-studied problems involving tensors. In particular, approximations of low multilinear rank (mrank) have long been investigated by virtue of their usefulness for subspace analysis and dimensionality reduction purposes. The first part of this paper introduces a novel algorithm which computes a low-mrank tensor approximation non-iteratively. This algorithm, called sequential low-rank approximation and projection (SeLRAP), generalizes a recently proposed scheme aimed at the rank-one case, SeROAP. We show that SeLRAP is always at least as accurate as existing alternatives in the rank-$(1, L, L)$ approximation of third-order tensors. By means of computer simulations with random tensors, such a superiority was actually observed for a range of different tensor dimensions and mranks. In the second part, we propose an iterative deflationary approach for computing a decomposition of a tensor in low-mrank blocks, termed DBTD. It first extracts an initial estimate of the blocks by employing SeLRAP, and then iteratively refines them by recomputing low-mrank approximations of each block plus the residue. Our numerical results show that, in the rank-$(1, L, L)$ case, this remarkably simple scheme outperforms existing algorithms if the blocks are not too correlated. In particular, it is much less sensitive to discrepancies among the block's norms.

Key words. Multilinear rank, low-rank approximation, block term decomposition, tensor.

AMS subject classifications. 15A69, 15A03, 65F99.

1. Introduction.

Approximating high-order tensors by parsimonious models is a recurrent problem across many engineering and scientific disciplines. In particular, given an Nth-order tensor $\boldsymbol{\mathcal{X}} \in \bigotimes_{n=1}^{N} \mathcal{V}_n \triangleq \mathcal{V}_1 \otimes \cdots \otimes \mathcal{V}_N$, one is often interested in finding subspaces $\mathcal{U}_n \subseteq \mathcal{V}_n$ of *reduced dimension* such that $\boldsymbol{\mathcal{X}}$ is well approximated by a tensor $\hat{\boldsymbol{\mathcal{X}}} \in \bigotimes_{n=1}^{N} \mathcal{U}_n$ according to some relevant criterion. The present work addresses this approximation problem in the finite-dimensional complex setting with a least-squares (LS) criterion:

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

where the target dimensions $R_n \leq I_n$ are given and $\|\cdot\|_F$ denotes the Frobenius norm. This is called *best low-multilinear-rank approximation* (LMA) problem, because the multilinear rank (mrank) of a tensor is defined as the tuple $\mathbf{m} = (R_1, \ldots, R_N)$ containing the minimal numbers such that (1) yields zero, i.e., $\boldsymbol{\mathcal{X}} = \hat{\boldsymbol{\mathcal{X}}}$. Contrarily to the best low-rank approximation problem, which is generally ill-posed for tensors of order higher than two, minimizers of (1) always exist [10].

A direct connection exists between mrank and the so-called Tucker decomposition: every finite-dimensional complex rank-$(R_1, \ldots, R_N)$ tensor[1] $\hat{\boldsymbol{\mathcal{X}}} \in \bigotimes_{n=1}^{N} \mathbb{C}^{I_n}$ can be expressed in the form

$$\hat{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{G}} \bullet_{n=1}^{N} \mathbf{U}^{(n)} \triangleq \boldsymbol{\mathcal{G}} \bullet_1 \mathbf{U}^{(1)} \bullet_2 \ldots \bullet_N \mathbf{U}^{(N)}, \tag{2}$$

*The contents of this work have been partially submitted to the EUSIPCO'2017 conference [13].

†Univ. Grenoble Alpes, CNRS, Gipsa-Lab, F-38000 Grenoble (jose-henrique.de-morais-goulart@gipsa-lab.fr, pierre.comon@gipsa-lab.fr). This work is supported by the European Research Council under the European Programme FP7/2007-2013, Grant AdG-2013-320594 "DECODA."

[1]Though $(R_1, \ldots, R_N)$ is the *mrank* of $\hat{\boldsymbol{\mathcal{X}}}$ rather than its rank, we employ the usual terminology "rank-$(R_1, \ldots, R_N)$" without confusion since we exclusively work with the mrank throughout.

where $\mathcal{G} \in \bigotimes_{n=1}^{N} \mathbb{C}^{R_n}$ is called the (Tucker) *core tensor*, $\mathbf{U}^{(n)} \in \mathbb{C}^{I_n \times R_n}$ is a *matrix factor* and $\bullet_n$ denotes mode-n contraction (also called mode-n product, see subsection 1.3 for a definition). This fundamental relation allows parameterizing problem (1), though in a highly non-unique manner. Furthermore, the factors $\mathbf{U}^{(n)}$ can be constrained to have orthonormal columns without loss of generality.

In applications, problem (1) is tackled for subspace analysis and dimensionality reduction purposes (see, e.g., [15, 27] and the examples given by [17]) by resorting to one of a host of existing iterative and non-iterative algorithms. A widely used iterative one is the higher-order orthogonal iteration (HOOI) [8], which extracts at each iteration an orthonormal basis for the dominant low-dimensional subspace associated with each mode by means of a singular value decompositions (SVD), thereby producing a low-mrank Tucker decomposition. This scheme, which is essentially an alternating least-squares (ALS) one with orthogonality constraints, generalizes the classical orthogonal iteration method [12]. A globally convergent Jacobi algorithm for symmetric tensors is derived in [16], being better suited than HOOI especially for large tensors, as it does not require SVDs. Another approach consists in iteratively performing the minimization over a Riemannian manifold; see [11, 17] and references in [17]. This allows circumventing the non-unique nature of the Tucker decomposition by restricting the search of the factors $\mathbf{U}^{(n)}$ to the product of quotient manifolds. The rank-one case (i.e., $R_n = 1$ for all n) has been studied by [28], which proposes and compares three iterative algorithms.

Non-iterative algorithms constitute a more suitable recourse whenever some error is tolerated or the computing cost must be kept at a low level (or both). The reason is that they try to compute a reasonable but generally suboptimal solution within a finite number of steps. As such, they are useful for initializing iterative LMA algorithms and also for plugging into other iterative algorithms which repeatedly compute LMAs, such as iterative hard thresholding (IHT) schemes for tensor completion (TC) [20, 14, 21]. The first and foremost non-iterative LMA algorithm is known as truncated higher-order SVD (THOSVD) [7]. It consists in projecting $\mathcal{X}$ onto the tensor product of *dominant* low-dimensional modal subspaces, i.e., those which (separately) capture most of the energy of each modal unfolding of $\mathcal{X}$ (see subsection 1.3 for a definition). Computing these subspaces requires N SVDs. Even though this solution is suboptimal in general, its LS error is bounded as

$$\|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 \leq N \|\mathcal{X} - \mathcal{X}^\star\|_F^2, \quad \text{where } \mathcal{X}^\star \text{ is a minimizer of (1).} \tag{3}$$

The alternative proposed in [25], which we refer to as sequentially optimal modal projections (SeMP), is less computationally intensive than the THOSVD, especially for small dimensions R_n. It also computes N SVDs, but they are interleaved with contractions which gradually reduce the tensor dimensions, each one being optimal given the preceding ones. Moreover, the resulting approximation error also obeys the bound in (3). In particular, for rank-$(1, L, L)$ approximations SeMP was shown to perform at least as well as THOSVD. Simulation results presented in [25] with random tensors suggest that this superiority actually holds in most cases. Concerning the special case of rank-one approximations, [3] has come up with a two-stage algorithm called sequential rank-one approximation and projection (SeROAP). It first reduces dimensionality similarly to SeMP, and then performs a sequence of "backward" projections to refine the approximation. For third-order tensors, it has been proven in [3] that SeROAP performs at least as well as SeMP and, consequently, as THOSVD too.

1.1. A generalization of SeROAP. Our first contribution is a generalization of SeROAP to arbitrary target mranks. We call this algorithm sequential low-rank approximation and projection (SeLRAP). In the case of third-order tensors, we show that SeLRAP performs at least as well as SeMP (and, consequently, as THOSVD) for rank-$(1, L, L)$ approximations. Despite the lack of proof for other cases, such a superiority was also observed in the overwhelming majority of our numerical simulations. This situation is similar to that described in [25] concerning the relationship between SeMP and THOSVD.

We point out that the contents of this article are partially reported in the conference paper [13]. In that paper, we give examples of TC scenarios where a SeLRAP-based IHT algorithm converges faster and attains smaller approximation error than SeMP- and THOSVD-based ones. Here, we focus instead on applying SeLRAP to the decomposition of a tensor in rank-$(1, L, L)$ blocks, as discussed next.

1.2. Decomposition in rank-$(1, L, L)$ terms via deflation. The block term decomposition (BTD) problem consists in decomposing a tensor $\mathcal{Y} \in \bigotimes_{n=1}^{N} \mathbb{C}^{I_n}$ as [4]

$$\mathcal{Y} = \sum_{r=1}^{R} \mathcal{Y}_r, \qquad \text{such that } \operatorname{mrank}(\mathcal{Y}_r) \leq \mathbf{m}^{(r)} = \left(R_1^{(r)}, \ldots, R_N^{(r)}\right). \tag{4}$$

In particular, when $N = 3$ and $\mathbf{m}^{(r)} = (1, L, L)$ for all r, it is known as the decomposition of $\mathcal{Y}$ in rank-$(1, L, L)$ terms,[2] and can be written as

$$\mathcal{Y} = \sum_{r=1}^{R} \mathbf{a}_r \otimes \left(\mathbf{B}_r \, \mathbf{C}_r^T\right), \tag{5}$$

where $\mathbf{a}_r \in \mathbb{C}^{I_1}$, $\mathbf{B}_r \in \mathbb{C}^{I_2 \times L}$ and $\mathbf{C}_r \in \mathbb{C}^{I_3 \times L}$. This particularization, which we will denote by BTD-$(1, L, L)$, has received a great deal of attention in the literature due to the various applications it finds—examples include, e.g., blind deconvolution [6], multidimensional harmonic retrieval [19], blind source separation [22] and electron energy loss spectroscopy [24].

On the theoretical side, conditions for the uniqueness of the blocks of a BTD-$(1, L, L)$ have been derived (up to a permutation of their indices) in [4, 5]. Such conditions are of central importance in applications because these blocks are typically computed as a means of estimating some quantities of interest. For the numerical computation of (5), an ALS algorithm has been put forth by [9]. In [18], an enhanced line search (ELS) scheme with exact (complex) step computation is incorporated into this algorithm, greatly improving its convergence speed. More recently, [23] has proposed conjugate gradient, quasi-Newton, Gauss-Newton and Levenberg-Marquardt algorithms for the BTD-$(1, L, L)$ problem.

Our second main contribution is the proposition of a deflation-based approach for the computation of a BTD. It extends the deflation-based canonical polyadic decomposition (DCPD) algorithm proposed in [2], which sequentially extracts rank-one terms from a tensor by computing approximations with SeROAP. Our extension, named deflation-based BTD (DBTD), employs SeLRAP to sequentially extract low-mrank approximations, yielding estimates of the desired blocks. Similarly to the rank-one case, a single application of this procedure does not suffice in general. So, an iterative refinement stage sequentially absorbs each estimated block into the residue and computes a new LMA, which is then subtracted from the residue. We show that the analysis of DCPD presented in [2] carries over to DBTD. In particular, monotonic

[2] The order of the modes and, consequently, of the components in $(1, L, L)$, can of course be permuted without changing the nature of the problem.

decrease of the residue norm is guaranteed under the assumption that optimal LMAs are computed.

Although in principle DBTD applies to the general decomposition (4), we shall focus here on the computation of rank-$(1, L, L)$ blocks. The reason is that even the computation of blocks $\mathcal{Y}_r$ having mranks of the form $(1, L_r, L_r)$ but with possibly different values for L_r is already considerably more difficult, due to the existence of local minima corresponding to wrong matchings of mranks to blocks [23]. When all blocks have mrank $(1, L, L)$ (i.e., $L_r = L$ for all r), our simulation results demonstrate that, despite being remarkably simple, DBTD outperforms competing alternatives, provided correlation among blocks is low.

1.3. Basic definitions and notation. Before proceeding, we introduce some basic definitions and notation. Scalars, vectors, matrices and tensors are denoted by lowercase, bold lowercase, bold uppercase and calligraphic uppercase letters, respectively (e.g., x, $\mathbf{x}$, $\mathbf{X}$, $\mathcal{X}$). The symbols $\otimes$ and $\boxtimes$ stand for the tensor and Kronecker products, respectively. The symbol $\mathcal{O}$ denotes the null tensor. Vector inequalities $\mathbf{x} \leq \mathbf{y}$ are meant entry-wise. Submatrices and subtensors are denoted by MATLAB-like notation, as in $[\mathbf{X}]_{:,1:R}$, which holds the first R columns of $\mathbf{X}$, and $[\mathbf{X}]_{i,:}$, which holds its ith row. The notation $\mathbf{X}_{\langle n \rangle} = (\mathcal{X})_{\langle n \rangle}$ stands for the mode-n (flat) matrix unfolding of $\mathcal{X}$, whose columns are subtensors $[\mathcal{X}]_{i_1,\ldots,i_{n-1},:,i_{n+1},\ldots,i_N}$ sorted in reverse lexicographical order with respect to the fixed indices. Given $\mathcal{X} \in \bigotimes_{n=1}^{N} \mathbb{C}^{I_n}$ and $\mathbf{P} \in \mathbb{C}^{M \times I_n}$, the mode-$n$ contraction (or product) is defined such that $(\mathcal{X} \bullet_n \mathbf{P})_{\langle n \rangle} = \mathbf{P}\mathbf{X}_{\langle n \rangle}$. For brevity, we employ the shorthands $\mathbb{N}_N \triangleq \{1, \ldots, N\}$, $\bar{I} \triangleq \prod_n I_n$ and $\bar{I}_n \triangleq \bar{I}/I_n$. Finally, $\mathbf{I}_M$ stands for the $M \times M$ identity matrix.

1.4. Paper organization. The rest of this work is organized in the following manner. Section 2 provides a brief review of existing non-iterative LMA algorithms, their properties and computational complexity. Then, our proposed approach is described and analyzed in section 3. Following that, section 4 introduces the DBTD algorithm and investigates its properties. Numerical results of computer simulations are presented in section 5, encompassing comparisons of SeLRAP and DBTD with competing alternatives for LMA and DBTD-$(1, L, L)$ computation, respectively. Concluding remarks and perspectives are then drawn in section 6.

2. State of the art.

2.1. Truncated higher-order singular value decomposition. Let us denote by $\mathcal{S}(I, R) = \{\mathbf{U} \in \mathbb{C}^{I \times R} : \mathbf{U}^H\mathbf{U} = \mathbf{I}_R\}$ the Stiefel manifold of column-wise orthonormal matrices and define

$$\mathcal{P}(I, R) = \left\{\mathbf{P} \in \mathbb{C}^{I \times I} : \mathbf{P} = \mathbf{U}\mathbf{U}^H,\, \mathbf{U} \in \mathcal{S}(I, R)\right\}. \tag{6}$$

Observe that $\mathcal{P}(I, R)$ contains all orthogonal projectors onto R-dimensional subspaces of $\mathbb{C}^I$. With this notation, one can equivalently formulate (1) as

$$\min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I_n, R_n)} \left\| \mathcal{X} - \mathcal{X} \overset{N}{\underset{n=1}{\bullet}} \hat{\mathbf{P}}^{(n)} \right\|_F^2. \tag{7}$$

Introducing a telescoping sum inside the norm, one obtains

$$\min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I, R)} \left\| \mathcal{X} - \sum_{n=1}^{N} \mathcal{X} \overset{n}{\underset{m=1}{\bullet}} \hat{\mathbf{P}}^{(m)} + \sum_{n=1}^{N-1} \mathcal{X} \overset{n}{\underset{m=1}{\bullet}} \hat{\mathbf{P}}^{(m)} \right\|_F^2. \tag{8}$$

Regrouping the terms, we have [25]

$$\min_{\hat{\mathbf{P}}^{(n)}\in\mathcal{P}(I,R)} \left\| \sum_{n=1}^{N} \mathcal{X} \overset{n-1}{\underset{m=1}{\bullet}} \hat{\mathbf{P}}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2 = \min_{\hat{\mathbf{P}}^{(n)}\in\mathcal{P}(I,R)} \sum_{n=1}^{N} \left\| \mathcal{X} \overset{n-1}{\underset{m=1}{\bullet}} \hat{\mathbf{P}}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2, \tag{9}$$

where $\hat{\mathbf{P}}_{\perp}^{(n)} \triangleq \mathbf{I}_{I_n} - \hat{\mathbf{P}}^{(n)}$ projects onto the orthogonal complement of $\operatorname{span}(\hat{\mathbf{P}}^{(n)})$ and the equality follows from pairwise orthogonality of the terms in the sum. The nonexpansiveness of orthogonal projections entails

$$\sum_{n=1}^{N} \left\| \mathcal{X} \overset{n-1}{\underset{m=1}{\bullet}} \hat{\mathbf{P}}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2 \leq \sum_{n=1}^{N} \left\| \mathcal{X} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2 = \sum_{n=1}^{N} \left\| \hat{\mathbf{P}}_{\perp}^{(n)} \mathbf{X}_{\langle n \rangle} \right\|_F^2. \tag{10}$$

Hence, it follows from the Eckart–Young theorem that the upper bound in (10) is minimized by projectors $\mathbf{P}^{(n)} = [\mathbf{U}^{(n)}]_{:,1:R_n} ([\mathbf{U}^{(n)}]_{:,1:R_n})^H$, where $\mathbf{U}^{(n)}$ is the matrix of left singular vectors of $\mathbf{X}_{\langle n \rangle}$. It is thus reasonable to approximate the solution of (7) by these projectors. To construct them, one can compute the SVD of each unfolding $\mathbf{X}_{\langle n \rangle}$, and then *truncate* the obtained matrix of left singular vectors $\mathbf{U}^{(n)}$ at the R_nth column. These matrices are the factors of the HOSVD of $\mathcal{X}$ [7]. Then, a truncated core is computed as $\mathcal{S} = \mathcal{X} \bullet_{n=1}^{N} ([\mathbf{U}^{(n)}]_{:,1:R_n})^H$, from which $\hat{\mathcal{X}}$ is obtained via $\hat{\mathcal{X}} = \mathcal{S} \bullet_{n=1}^{N} [\mathbf{U}^{(n)}]_{:,1:R_n}$. Now, because these projectors $\mathbf{P}^{(n)}$ are optimal when considered separately (but not jointly), any solution $\mathcal{X}^\star$ of (1) satisfies

$$\left\| \mathbf{P}_{\perp}^{(n)} \mathbf{X}_{\langle n \rangle} \right\|_F^2 \leq \left\| \mathbf{X}_{\langle n \rangle} - \mathbf{X}_{\langle n \rangle}^\star \right\|_F^2.$$

Plugging this expression into (10) shows the cost function value attained by the THOSVD solution is no greater than N times (1), which proves the bound in (3).

Denoting by $C_{\text{SVD}}(I, M)$ the number of operations required to compute the SVD of an $I \times M$ matrix, THOSVD's cost can be expressed as

$$C_{\text{THOSVD}} = \textstyle\sum_{n=1}^{N} C_{\text{SVD}}(I_n, \bar{I}_n) + \sum_{n=1}^{N} \mathcal{O}(H_n R_n I_n) + \sum_{n=1}^{N} \mathcal{O}(J_n R_n I_n), \tag{11}$$

where $H_n \triangleq R_1 \dots R_{n-1} I_{n+1} \dots I_N$ 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.[3]

If one uses a standard algorithm for computing the full ("economical") SVD prior to truncation, then $C_{\text{SVD}}(I, M) = \mathcal{O}(IM \min\{I, M\})$. Though there exist methods which in principle cost $\mathcal{O}(RIM)$ for obtaining the R dominant singular triplets of an $I \times M$ matrix [1], in practice they often fall behind on computing time, except for very small R.

2.2. Sequentially optimal modal projections. Another way of computing an approximate solution of (9) is by sequentially minimizing the cost function with respect to the projectors. This leads to the SeMP solution [25], defined as:

$$\text{Given } \mathbf{P}^{(1)}, \dots, \mathbf{P}^{(n-1)}, \text{ compute } \mathbf{P}^{(n)} = \underset{\hat{\mathbf{P}}^{(n)}\in\mathcal{P}(I_n,R_n)}{\arg\min} \left\| \mathcal{X} \overset{n-1}{\underset{m=1}{\bullet}} \mathbf{P}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2. \tag{12}$$

For simplicity of exposition, we have considered such a computation in the natural order $(1, ..., N)$, but any other order can be adopted, generally leading to different results. The projectors defined by (12) are computed as follows:

[3] We 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.

1. Let $\mathcal{W}^{(1)} = \mathcal{X}$.
2. For $n = 1, \ldots, N$:
   (i) compute the SVD of $\mathbf{W}^{(n)}_{\langle n \rangle}$ to obtain $\overline{\mathbf{U}}^{(n)} \in \mathbb{C}^{I_n \times R_n}$, which holds its first R_n left singular vectors;
   (ii) compute $\mathcal{W}^{(n+1)} = \mathcal{W}^{(n)} \bullet_n \overline{\mathbf{U}}^{(n)H} \in \left(\bigotimes_{m=1}^{n} \mathbb{C}^{R_n}\right) \otimes \left(\bigotimes_{m=n+1}^{N} \mathbb{C}^{I_n}\right)$.
3. Finally, construct the solution $\hat{\mathcal{X}} = \mathcal{W}^{(N+1)} \bullet_{n=1}^{N} \overline{\mathbf{U}}^{(m)}$.

It is easy to show that the resulting approximation error is subject to the same upper bound as the THOSVD. Indeed, we have [25]

$$\begin{aligned} \|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 &= \sum_{n=1}^{N} \min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I,R)} \left\| \mathcal{X} \overset{n-1}{\underset{m=1}{\bullet}} \mathbf{P}^{(m)} \bullet_n \hat{\mathbf{P}}^{(n)}_{\perp} \right\|_F^2 \\ &\leq \sum_{n=1}^{N} \min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I,R)} \left\| \mathcal{X} \bullet_n \hat{\mathbf{P}}^{(n)}_{\perp} \right\|_F^2 \leq N \|\mathcal{X} - \mathcal{X}^{\star}\|_F^2, \end{aligned} \tag{13}$$

where $\mathcal{X}^{\star}$ is any solution of (1). Furthermore, the SVDs in step 2.(i) for $n > 1$ have smaller size than the corresponding ones computed by THOSVD, due to the dimension reduction performed in step 2.(ii). Thus, the resulting cost

$$C_{\text{SeMP}} = \textstyle\sum_{n=1}^{N} \left[C_{\text{SVD}}(I_n, H_n) + \mathcal{O}(H_n R_n I_n)\right] + \sum_{n=1}^{N} \mathcal{O}(J_n R_n I_n) \tag{14}$$

is always smaller than (11), since $H_n < \bar{I}_n$ must hold for at least one n (otherwise there is no rank reduction). Typically, $H_n < \bar{I}_n$ for all $n > 1$. The smaller the ratios R_n/I_n, the greater the computational advantage with respect to THOSVD. With the goal of reducing the computing effort, a heuristic is described in [25] for choosing the order in which modes are processed. The idea is to sort them according to their dimensions, in ascending order. This is a greedy strategy in the sense that it picks at each step the mode whose unfolding has the least costly SVD.

From our practical experience, the approximations obtained via SeMP are virtually always more accurate than those given by the THOSVD. This is in line with the conclusions reported in [25]. However, a proof of its superiority currently exists only for rank-$(1, L, L)$ approximations, as stated below.

THEOREM 1 (Theorem 7.2 of [25]). *Let $\mathcal{X} \in \bigotimes_{n=1}^{3} \mathbb{C}^{I_n}$ and denote by $\hat{\mathcal{X}}_{SeMP}$ and $\hat{\mathcal{X}}_{THOSVD}$ the rank-$(1, L, L)$ approximations of $\mathcal{X}$ produced by SeMP and THOSVD, respectively, by processing the modes in the natural order (1,2,3). Then,*

$$\left\| \mathcal{X} - \hat{\mathcal{X}}_{SeMP} \right\|_F^2 \leq \left\| \mathcal{X} - \hat{\mathcal{X}}_{THOSVD} \right\|_F^2. \tag{15}$$

The proof given in [25] exploits the facts that (1) the projector $\mathbf{P}^{(1)}$ computed by SeMP is the same as in the THOSVD solution and (2) $\mathcal{W}^{(2)}$ actually reduces to a matrix when $R_1 = 1$. Thus, $\mathbf{P}^{(2)}$ and $\mathbf{P}^{(3)}$ are obtained in SeMP with a single SVD. Because by construction these projectors are optimal given $\mathbf{P}^{(1)}$, THOSVD's outcome cannot be more accurate.

2.3. Sequential rank-one approximation and projection. When $R_1 = \cdots = R_N = 1$, problem (1) reduces to the best rank-one approximation of $\mathcal{X}$. In other words, one seeks an elementary (or decomposable) tensor $\hat{\mathcal{X}} = \mathbf{v}^{(1)} \otimes \cdots \otimes \mathbf{v}^{(N)}$ minimizing the cost function in (1). Note that no distinction exists between tensor rank and mrank in this case. The SeROAP algorithm [3] computes an approximate solution by proceeding as follows:

1. *Order reduction stage:*
   (i) Let $\mathcal{W}^{(1)} \triangleq \mathcal{X}$.
   (ii) For $n = 2, \ldots, N$, recursively calculate the tensor $\mathcal{W}^{(n)} \in \left(\bigotimes_{m=1}^{n-1} \mathbb{C}^1\right) \otimes \left(\bigotimes_{m=n}^{N} \mathbb{C}^{I_m}\right)$ whose vectorization $\mathbf{w}^{(n)} \triangleq \mathrm{vec}(\mathcal{W}^{(n)})$ is a minimizer of

$$\min_{\hat{\lambda}\in\mathbb{R},\, \hat{\mathbf{u}}^{(n)}\in\mathbb{C}^{I_{n-1}}\, \hat{\mathbf{w}}^{(n)}\in\mathbb{C}^{I_N\cdots I_n}} \left\| \mathbf{W}_{\langle n-1\rangle}^{(n-1)} - \hat{\lambda}\, \hat{\mathbf{u}}^{(n)} \left(\hat{\mathbf{w}}^{(n)}\right)^H \right\|_F^2 .$$

   This can be done by computing the dominant singular triplet of the matrix $\mathbf{W}_{\langle n-1\rangle}^{(n-1)} \in \mathbb{C}^{I_{n-1}\times I_N\ldots I_n}$.
2. *Projection stage:*
   (i) Let $\mathbf{z}^{(N-1)} \triangleq \mathbf{w}^{(N)^*} \boxtimes \mathbf{u}^{(N)} \in \mathbb{C}^{I_N I_{N-1}}$.
   (ii) For $n = N-2, \ldots, 1$, project the rows of $\mathbf{W}_{\langle n\rangle}^{(n)}$ onto $\mathrm{span}(\mathbf{z}^{(N+1)})$, i.e.,

$$\mathbf{Z}_{\langle n\rangle}^{(n)} = \mathbf{W}_{\langle n\rangle}^{(n)} \left[\frac{1}{\|\mathbf{z}^{(n+1)}\|_2^2} \left(\mathbf{z}^{(n+1)} \mathbf{z}^{(n+1)^H} \right) \right] \in \mathbb{C}^{I_n \times I_N \ldots I_{n+1}}, \tag{16}$$

   and then obtain $\mathbf{z}^{(n)}$ as $\mathbf{z}^{(n)} = \mathrm{vec}(\mathcal{Z}^{(n)})$.
3. Construct the estimate $\hat{\mathcal{X}}$ such that $\mathrm{vec}(\hat{\mathcal{X}}) = \mathbf{z}^{(1)} \in \mathbb{C}^{I_N \cdots I_1}$, or, equivalently, such that $\hat{\mathbf{X}}_{\langle 1\rangle} = \mathbf{Z}_{\langle 1\rangle}^{(1)}$.

It can be verified that the order reduction stage is identical to SeMP's dimension reduction stage when $R_n = 1$ for all n. Indeed, using the above notation, the rank-one approximation delivered by SeMP is proportional to $\mathbf{u}^{(2)} \otimes \cdots \otimes \mathbf{u}^{(N)} \otimes \mathbf{w}^{(N)}$. Intuitively, the "backward" projection stage performed by SeMP attempts to improve this initial recursive approximation. For third-order tensors, the following result holds.

THEOREM 2 (Theorem 1 of [3]). *Let $\mathcal{X} \in \bigotimes_{n=1}^{3} \mathbb{C}^{I_n}$ and denote by $\hat{\mathcal{X}}_{SeROAP}$ and $\hat{\mathcal{X}}_{SeMP}$ the rank-one approximations of $\mathcal{X}$ produced by SeROAP and SeMP, respectively, both processing the modes in the same (any) order. Then,*

$$\left\| \mathcal{X} - \hat{\mathcal{X}}_{SeROAP} \right\|_F^2 \leq \left\| \mathcal{X} - \hat{\mathcal{X}}_{SeMP} \right\|_F^2 . \tag{17}$$

By employing a k-step Lanczos-type algorithm of cost $\mathcal{O}(kIM)$ to compute the dominant singular triplet of an $I \times M$ matrix, the order reduction stage has cost $\sum_{n=1}^{N-1} \mathcal{O}\left(k \prod_{m=n}^{N} I_m\right)$. The overall complexity of SeROAP can thus be expressed as

$$C_{\mathrm{SeROAP}} = \sum_{n=1}^{N-1} \left[\mathcal{O}\left(k \prod_{m=n}^{N} I_m\right) + \mathcal{O}\left(\prod_{m=n}^{N} I_m\right) \right].$$

3. Sequential low-rank approximation and projection.

3.1. Formulation and algorithm. The same principle underlying SeROAP can also be employed for computing an LMA of arbitrary mrank $\mathbf{m} = (R_1, \ldots, R_N)$. In the projection stage of SeROAP, the rows of each unfolding $\mathbf{W}_{\langle n\rangle}^{(n)}$ are projected onto the subspace spanned by a Kronecker-structured vector representing a tensor product of one-dimensional subspaces. This suggests that a general procedure should project these rows onto the span of a Kronecker-structured basis representing a tensor product of low-dimensional subspaces, in consonance with $\mathbf{m}$.

Such a generalization is accomplished by the following procedure (SeLRAP):

Algorithm 1 Sequentially low-rank approximation and projection (SeLRAP).

Inputs: $\mathcal{X} \in \bigotimes_{n=1}^{N} \mathbb{C}^{I_n}$ and target mrank $\mathbf{m} = (R_1, \ldots, R_N)$
Output: An $\mathbf{m}$-mrank approximation of $\mathcal{X}$, $\hat{\mathcal{X}}$

$\mathbf{W}_{\langle 1 \rangle}^{(1)} \leftarrow \mathbf{X}_{\langle 1 \rangle}$
for $n = 2, \ldots, N+1$ **do**
  $\left(\mathbf{W}_{\langle n-1 \rangle}^{(n)}, \overline{\mathbf{U}}^{(n-1)}\right) \leftarrow \texttt{projdomsp}\left(\mathbf{W}_{\langle n-1 \rangle}^{(n-1)}, R_{n-1}\right)$
end for
$\mathbf{Z}_{\langle N \rangle}^{(N)} \leftarrow \overline{\mathbf{U}}^{(N)} \mathbf{W}_{\langle N \rangle}^{(N+1)}$
for $n = N-1, \ldots, 1$ **do**
  compute the "thin" QR decomposition: $\mathbf{Z}_{\langle n \rangle}^{(n+1)^H} = \mathbf{Q}^{(n+1)} \mathbf{R}^{(n+1)}$
  $\mathbf{Z}_{\langle n \rangle}^{(n)} \leftarrow \left(\mathbf{W}_{\langle n \rangle}^{(n)} \mathbf{Q}^{(n+1)}\right) \mathbf{Q}^{(n+1)^H}$
end for
return $\hat{\mathcal{X}} = \mathcal{Z}^{(1)}$

1. *Dimension reduction stage:* In this stage, one computes a sequence of tensors $\mathcal{W}^{(n)} \in \left(\bigotimes_{m=1}^{n-1} \mathbb{C}^{R_n}\right) \otimes \left(\bigotimes_{m=n}^{N} \mathbb{C}^{I_n}\right)$ in exactly the same way as in SeMP.
2. *Projection stage:* Then, one recursively obtains tensors $\mathcal{Z}^{(n)}$ of same dimensions as $\mathcal{W}^{(n)}$ by performing a sequence of orthogonal projections, similarly to SeROAP. Specifically:
   (i) Let $\mathcal{Z}^{(N)} = \mathcal{W}^{(N+1)} \bullet_N \overline{\mathbf{U}}^{(N)}$.
   (ii) For $n = N-1, \ldots, 1$, the mode-n unfolding of $\mathcal{Z}^{(n)}$ is computed as

$$\mathbf{Z}_{\langle n \rangle}^{(n)} = \mathbf{W}_{\langle n \rangle}^{(n)} \mathbf{Z}_{\langle n \rangle}^{(n+1)^H} \left(\mathbf{Z}_{\langle n \rangle}^{(n+1)} \mathbf{Z}_{\langle n \rangle}^{(n+1)^H}\right)^{-1} \mathbf{Z}_{\langle n \rangle}^{(n+1)}. \tag{18}$$

3. The desired rank-$(R_1, \ldots, R_N)$ approximation is then $\hat{\mathcal{X}} = \mathcal{Z}^{(1)}$.

It is not hard to check that the above procedure reduces to SeROAP when $R_n = 1$ for all n. As a matter of fact, in this particular case (16) and (18) are equivalent. In subsection 3.3, we will give examples of application of SeLRAP which will showcase the Kronecker structure of the matrices $\mathbf{Z}_{\langle n \rangle}^{(n+1)}$.

An explicit pseudocode for SeLRAP is given in Algorithm 1. For simplicity, the modes are processed in the natural order $(1, \ldots, N)$. If one wishes to follow a different order, it suffices to permute the modes of $\mathcal{X}$ before running the algorithm and to invert this permutation afterwards. In the dimension reduction stage, SeLRAP employs the `projdomsp` routine given in Algorithm 2. When $I_{n-1} < H_{n-1}$, instead of computing the SVD of $\mathbf{W}_{\langle n-1 \rangle}^{(n-1)}$ this routine computes the eigenvalue decomposition of $\mathbf{Y}^{(n-1)} \triangleq \mathbf{W}_{\langle n-1 \rangle}^{(n-1)} \mathbf{W}_{\langle n-1 \rangle}^{(n-1)^H}$, which provides $\overline{\mathbf{U}}^{(n-1)}$. Although the asymptotic complexities are the same due to the computation of $\mathbf{Y}^{(n-1)}$, this choice can save much time in practice, because of the reduced size of the decomposition problem. On the other hand, if $I_{n-1} \geq H_{n-1}$, then after computing the SVD of $\mathbf{W}_{\langle n-1 \rangle}^{(n-1)}$, one can obtain $\mathcal{W}^{(n)}$ by taking the first R_{n-1} right singular vectors multiplied by their corresponding singular values; this is cheaper than calculating $\overline{\mathbf{U}}^{(n-1)^H} \mathbf{W}_{\langle n-1 \rangle}^{(n-1)}$.

Algorithm 2 `projdomsp`$(\mathbf{W}, R)$: projects $\mathbf{W}$ onto its R-dimensional dominant column subspace.

Inputs: $\mathbf{W} \in \mathbb{C}^{I \times M}$, target dimension $R \leq I$

Outputs: $\hat{\mathbf{W}}$ and $\overline{\mathbf{U}}$, where $\overline{\mathbf{U}} = \arg\min_{\mathbf{U} \in \mathcal{S}(I,R)} \|\mathbf{W} - \mathbf{U}\mathbf{U}^H\mathbf{W}\|_F$ and $\hat{\mathbf{W}} = \overline{\mathbf{U}}^H\mathbf{W}$

if $I < M$ **then**
  $\mathbf{Y} \leftarrow \mathbf{W}\mathbf{W}^H$
  compute the EVD: $\mathbf{Y} = \begin{bmatrix}\overline{\mathbf{U}} & \tilde{\mathbf{U}}\end{bmatrix} \mathbf{\Lambda} \begin{bmatrix}\overline{\mathbf{U}} & \tilde{\mathbf{U}}\end{bmatrix}^H$,
  where $\overline{\mathbf{U}} \in \mathbb{C}^{I \times R}$ and $\mathbf{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_I)$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_I$
  $\hat{\mathbf{W}} \leftarrow \overline{\mathbf{U}}^H\mathbf{W}$
else
  compute the SVD: $\mathbf{W} = \begin{bmatrix}\overline{\mathbf{U}} & \tilde{\mathbf{U}}\end{bmatrix} \begin{bmatrix}\overline{\mathbf{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{\Sigma}}\end{bmatrix} \begin{bmatrix}\overline{\mathbf{V}} & \tilde{\mathbf{V}}\end{bmatrix}^H$,
  where $\overline{\mathbf{U}} \in \mathbb{C}^{I \times R}$, $\overline{\mathbf{\Sigma}} \in \mathbb{R}^{R \times R}$ and $\overline{\mathbf{V}} \in \mathbb{C}^{M \times R}$
  $\hat{\mathbf{W}} \leftarrow \overline{\mathbf{\Sigma}}\,\overline{\mathbf{V}}^H$
end if
return $\left(\hat{\mathbf{W}}, \overline{\mathbf{U}}\right)$

Remark 3. When $I_{n-1} < H_{n-1}$, it is actually more appropriate to first compute the decomposition $\mathbf{W}_{\langle n-1 \rangle}^{(n-1)^H} = \mathbf{Q}\mathbf{R}$, with $\mathbf{Q} \in \mathcal{S}(H_{n-1}, I_{n-1})$ and $\mathbf{R} \in \mathbb{C}^{I_{n-1} \times I_{n-1}}$ via a modified Gram-Schmidt orthogonalization. Then, the SVD of $\mathbf{R}$ is computed, providing the right singular vectors of $\mathbf{R}$, which are premultiplied by $\mathbf{Q}$ to yield those of $\mathbf{W}_{\langle n-1 \rangle}^{(n-1)}$, and also its singular values, which equal those of $\mathbf{W}_{\langle n-1 \rangle}^{(n-1)}$. Finally, $\mathcal{W}^{(n)}$ is calculated from the obtained right singular vectors and singular values of $\mathbf{W}_{\langle n-1 \rangle}^{(n-1)}$. This procedure requires less flops than `projdomsp` (though their order-wise complexities are the same) and is more accurate. However, Algorithm 2 runs faster in MATLAB, because the computation of $\mathbf{Y}^{(n-1)} = \mathbf{W}_{\langle n-1 \rangle}^{(n-1)}\mathbf{W}_{\langle n-1 \rangle}^{(n-1)^H}$ is highly optimized. Therefore, we have adopted it in our MATLAB implementation of SeLRAP.[4]

3.2. Computational complexity. In practice, Algorithm 1 performs (18) with the aid of an orthonormal basis for the row space of $\mathbf{Z}_{\langle n \rangle}^{(n+1)}$, obtained thanks to a QR decomposition. One could also use an SVD. In any case, the constructed projector must have the same rank as $\mathbf{Z}_{\langle n \rangle}^{(n+1)}$. If its rank is smaller than R_n, then this amounts to replacing the inverse matrix of (18) by the Moore-Penrose pseudo-inverse. Computing the orthonormal basis costs $\mathcal{O}(H_n R_n^2)$ flops (assuming $H_n \geq R_n$), since $\mathbf{Z}_{\langle n \rangle}^{(n+1)}$ has dimensions $R_n \times H_n$, while performing the projection costs $\mathcal{O}(H_n R_n I_n)$. SeLRAP thus has the overall complexity

$$C_{\text{SeLRAP}} = \sum_{n=1}^{N} \left[C_{\text{SVD}}(I_n, H_n) + \mathcal{O}(H_n R_n I_n)\right] + \mathcal{O}(H_N R_N I_N) + \sum_{n=1}^{N-1} C_{\text{proj}}(H_n, R_n, I_n),$$

[4] For a fair comparison, our SeMP implementation also uses Algorithm 2 for dimensionality reduction and our THOSVD implementation computes the EVD of $\mathbf{X}_{\langle n \rangle}\mathbf{X}_{\langle n \rangle}^H$ to obtain the dominant subspace of $\mathbf{X}_{\langle n \rangle}$ when $I_n < \bar{I}_n$.

TABLE 1
Operations involved in non-iterative LMA algorithms and their costs.

THOSVD		
(I) Comput. projectors	(II) Comput. truncated core	(III) LMA construction
$\sum_{n=1}^{N} C_{\mathrm{SVD}}(I_n, \bar{I}_n)$	$\sum_{n=1}^{N} \mathcal{O}(H_n R_n I_n)$	$\sum_{n=1}^{N} \mathcal{O}(J_n R_n I_n)$
SeMP		
(I) Comput. projectors	(II) Dimension reduction	(III) LMA construction
$\sum_{n=1}^{N} C_{\mathrm{SVD}}(I_n, H_n)$	$\sum_{n=1}^{N} \mathcal{O}(H_n R_n I_n)$	$\sum_{n=1}^{N} \mathcal{O}(J_n R_n I_n)$
SeLRAP		
(I) Comput. projectors	(II) Dimension reduction	(III) Backward projections
$\sum_{n=1}^{N} C_{\mathrm{SVD}}(I_n, H_n)$	$\sum_{n=1}^{N} \mathcal{O}(H_n R_n I_n)$	$\mathcal{O}(H_N R_N I_N) + \sum_{n=1}^{N-1} C_{\mathrm{proj}}(H_n, R_n, I_n)$

where $C_{\mathrm{proj}}(H, R, I) = \mathcal{O}(HR^2) + \mathcal{O}(HRI)$. The term $\mathcal{O}(H_N R_N I_N)$ outside the brackets corresponds to step 5 of Algorithm 1, while the second term of the first summand comprehends the cost of the dimension-reducing contractions. Yet, as discussed in subsection 3.1, when $I_n \geq H_n$ then the nth term is $\mathcal{O}(H_n R_n)$ rather than $\mathcal{O}(H_n R_n I_n)$, because one uses the right singular vectors of $\mathbf{W}^{(n-1)}_{\langle n-1 \rangle}$ to form $\mathcal{W}^{(n)}$. Note that `projdomsp` can be employed also in SeMP, and so the same remark applies to the second summation of (14). On the other hand, THOSVD can only partially benefit from the strategy followed in `projdomsp`, by computing the eigenvectors of $\mathbf{X}_{\langle n \rangle}\mathbf{X}_{\langle n \rangle}{}^H$ when $I_n < \bar{I}_n$.

The same heuristic described in subsection 2.2, of processing the modes in ascending order of dimensions, usually yields a significant economy of computing time when applying SeLRAP. This economy is all the more relevant when the ratios R_n/I_n are small and approximately equal.

Table 1 summarizes the operations involved in THOSVD, SeMP and SeLRAP. Though the operations (I) and (II) are sequentially performed in THOSVD while they are interleaved in SeLRAP and SeMP, there is a clear parallel among equally numbered operations of different algorithms.

3.3. Analysis.

3.3.1. Fulfillment of rank constraint. We now show that the approximation delivered by SeLRAP actually meets the desired mrank constraint.

LEMMA 4. *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* $\mathrm{mrank}(\mathcal{P}) = (R_1, \ldots, R_N)$, *then*

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

is the mode-n unfolding of a tensor $\mathcal{Y} \in \bigotimes_{m=1}^{N} \mathbb{C}^{K_m}$, *with* $\mathrm{mrank}(\mathcal{Y}) \leq (R_1, \ldots, R_N)$.

Proof. Since $\mathrm{mrank}(\mathcal{P}) = (R_1, \ldots, R_N)$, 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

$$\mathbf{P}_{\langle n \rangle} = \mathbf{G}_{\langle n \rangle} \underbrace{\left(\mathbf{U}^{(N)} \boxtimes \ldots \boxtimes \mathbf{U}^{(n+1)} \boxtimes \mathbf{U}^{(n-1)} \boxtimes \ldots \boxtimes \mathbf{U}^{(1)}\right)^T}_{\triangleq \mathbf{U}^T}.$$

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 $\operatorname{rank}(\mathbf{Y}_{\langle n \rangle}) \leq R_n$. Defining $\mathcal{C} \in \bigotimes_{m=1}^{N} \mathbb{C}^{R_m}$ such that $\mathbf{C}_{\langle n \rangle} = \mathbf{X}_{\langle n \rangle}\mathbf{U}^{*}\mathbf{G}_{\langle n \rangle}^{H}(\mathbf{G}_{\langle n \rangle}\mathbf{G}_{\langle n \rangle}^{H})^{-1}\mathbf{G}_{\langle n \rangle}$, it follows that $\mathcal{Y} = \mathcal{C} \bullet_{m \neq n} \mathbf{U}^{(m)}$, implying $\operatorname{rank}(\mathbf{Y}_{\langle m \rangle}) \leq R_m$ for all $m \in \mathbb{N}_N \backslash \{n\}$. ∎

3.3.2. Comparison with SeMP. In the following, we analytically compare the quadratic errors incurred by SeMP and SeLRAP for third-order tensors. Let $\mathbf{m} = (R_1, R_2, R_3)$ and denote by $\hat{\mathcal{X}}_{\text{SeMP}} = \mathcal{S} \bullet_{n=1}^{3} \overline{\mathbf{U}}^{(n)}$ the approximation delivered by SeMP, where $\overline{\mathbf{U}}^{(n)}$ is as defined in subsection 2.2. Because $\mathcal{S} = \mathcal{X} \bullet_{n=1}^{3} \overline{\mathbf{U}}^{(n)H}$, the resulting quadratic error can be written as

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

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

$$\varepsilon_{\text{SeMP}} = \|\mathcal{X}\|_F^2 - \left\| \overline{\mathbf{\Sigma}}^{(1)} \overline{\mathbf{V}}^{(1)H} \left(\overline{\mathbf{U}}^{(3)*} \boxtimes \overline{\mathbf{U}}^{(2)*} \right) \right\|_F^2 , \tag{19}$$

where the columns of $\overline{\mathbf{V}}^{(1)}$ are the first R_n singular vectors of $\mathbf{X}_{\langle 1 \rangle} = \mathbf{W}_{\langle 1 \rangle}^{(1)}$, while $\overline{\mathbf{\Sigma}}^{(1)}$ contains the corresponding singular values in its diagonal. This is a direct generalization of the expression derived in [3] for the case $\mathbf{m} = (1, 1, 1)$.

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

$$\mathbf{P} \triangleq \mathbf{Z}_{\langle 1 \rangle}^{(2)H} \left(\mathbf{Z}_{\langle 1 \rangle}^{(2)} \mathbf{Z}_{\langle 1 \rangle}^{(2)H} \right)^{-1} \mathbf{Z}_{\langle 1 \rangle}^{(2)}. \tag{20}$$

Using this definition along with (18) 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

$$\varepsilon_{\text{SeLRAP}} \triangleq \|\mathcal{X} - \hat{\mathcal{X}}_{\text{SeLRAP}}\|_F^2 = \|\mathcal{X}\|_F^2 - \left\| \mathbf{W}_{\langle 1 \rangle}^{(1)} \mathbf{P} \right\|_F^2 . \tag{21}$$

Writing $\mathbf{W}_{\langle n \rangle}^{(n)} = \overline{\mathbf{U}}^{(n)} \overline{\mathbf{\Sigma}}^{(n)} \overline{\mathbf{V}}^{(n)H} + \mathbf{E}^{(n)}$, the second norm in (21) can be rewritten as

$$\left\| \mathbf{W}_{\langle 1 \rangle}^{(1)} \mathbf{P} \right\|_F^2 = \operatorname{Tr} \left\{ \mathbf{P} \mathbf{W}_{\langle 1 \rangle}^{(1)H} \mathbf{W}_{\langle 1 \rangle}^{(1)} \mathbf{P} \right\} = \left\| \overline{\mathbf{\Sigma}}^{(1)} \overline{\mathbf{V}}^{(1)H} \mathbf{P} \right\|_F^2 + \left\| \mathbf{E}^{(1)} \mathbf{P} \right\|_F^2 .$$

Plugging the result into (21), we have

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

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

$$\left\| \overline{\mathbf{\Sigma}}^{(1)} \overline{\mathbf{V}}^{(1)^H} \mathbf{P} \right\|_F^2 \geq \left\| \overline{\mathbf{\Sigma}}^{(1)} \overline{\mathbf{V}}^{(1)^H} \left(\overline{\mathbf{U}}^{(3)^*} \boxtimes \overline{\mathbf{U}}^{(2)^*} \right) \right\|_F^2 . \tag{22}$$

In 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 $R_2 = R_3 = L$. (This can be easily seen from the mode-2 and mode-3 unfoldings of an rank-$(1, R_2, R_3)$ Tucker model.) For this case, the following result holds.

THEOREM 5. *Let $\mathcal{X} \in \bigotimes_{n=1}^{3} \mathbb{C}^{I_n}$ and denote by $\hat{\mathcal{X}}_{SeLRAP}$ and $\hat{\mathcal{X}}_{SeMP}$ the rank-$(1, L, L)$ approximations of $\mathcal{X}$ produced by SeLRAP and SeMP, respectively, by processing the modes in the natural order (1,2,3). Then,*

$$\left\| \mathcal{X} - \hat{\mathcal{X}}_{SeLRAP} \right\|_F^2 \leq \left\| \mathcal{X} - \hat{\mathcal{X}}_{SeMP} \right\|_F^2 . \tag{23}$$

Proof. First, SeLRAP computes the SVDs

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

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

where $\mathbf{W}_{\langle 2 \rangle}^{(2)}$ is such that $\mathbf{W}_{\langle 1 \rangle}^{(2)} = \overline{\sigma}^{(1)} \overline{\mathbf{v}}^{(1)^H}$. Observe that, for $R_1 = 1$,

$$\text{vec}\left(\mathbf{W}_{\langle 1 \rangle}^{(2)} \right) = \overline{\sigma}^{(1)} \overline{\mathbf{v}}^{(1)^*} = \text{vec}\left(\mathbf{W}_{\langle 2 \rangle}^{(2)} \right) = \text{vec}\left(\overline{\mathbf{U}}^{(2)} \overline{\mathbf{\Sigma}}^{(2)} \overline{\mathbf{V}}^{(2)^H} + \mathbf{E}^{(2)} \right) \tag{24}$$

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

$$\mathbf{Z}_{\langle 3 \rangle}^{(3)} = \mathbf{W}_{\langle 3 \rangle}^{(3)} = \overline{\mathbf{U}}^{(3)} \overline{\mathbf{\Sigma}}^{(3)} \in \mathbb{C}^{I_3 \times R}, \tag{25}$$

because $\text{rank}(\mathbf{W}_{\langle 3 \rangle}^{(3)}) \leq L$. Furthermore, $\mathbf{Z}_{\langle 2 \rangle}^{(3)} = \mathbf{Z}_{\langle 3 \rangle}^{(3)^T}$. Thus, plugging $\mathbf{Z}_{\langle 2 \rangle}^{(3)}$ into (18) for $n = 2$ we obtain

$$\mathbf{Z}_{\langle 2 \rangle}^{(2)} = \mathbf{W}_{\langle 2 \rangle}^{(2)} \overline{\mathbf{V}}^{(2)} \overline{\mathbf{V}}^{(2)^H} = \overline{\mathbf{U}}^{(2)} \overline{\mathbf{\Sigma}}^{(2)} \overline{\mathbf{V}}^{(2)^H} \in \mathbb{C}^{I_2 \times I_3}. \tag{26}$$

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

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

which implies $\mathbf{P} = \| \overline{\boldsymbol{\sigma}}^{(2)} \|_2^{-2} \, \mathbf{U}^* \, \overline{\boldsymbol{\sigma}}^{(2)} \, \overline{\boldsymbol{\sigma}}^{(2)^T} \, \mathbf{U}^T$, where $\overline{\boldsymbol{\sigma}}^{(2)} \triangleq \text{vec}\left(\overline{\mathbf{\Sigma}}^{(2)} \right)$ and $\mathbf{U} \triangleq \overline{\mathbf{U}}^{(3)} \boxtimes \overline{\mathbf{U}}^{(2)}$. Applying these definitions to (24), we have also

$$\overline{\sigma}^{(1)} \overline{\mathbf{v}}^{(1)^*} = \mathbf{U} \, \overline{\boldsymbol{\sigma}}^{(2)} + \text{vec}\left(\mathbf{E}^{(2)} \right). \tag{27}$$

In view of the derived expressions, computing the left-hand side of (22) for $\mathbf{m} = (1, L, L)$ yields

$$\left\|\overline{\sigma}^{(1)}\overline{\mathbf{v}}^{(1)^H}\mathbf{P}\right\|_2^2 = \left(\overline{\sigma}^{(1)}\right)^2 \overline{\mathbf{v}}^{(1)^H}\mathbf{P}\overline{\mathbf{v}}^{(1)} = \left(\overline{\sigma}^{(1)}\right)^2 \|\overline{\boldsymbol{\sigma}}^{(2)}\|_2^{-2} \left|\overline{\boldsymbol{\sigma}}^{(2)^T}\mathbf{U}^T\overline{\mathbf{v}}^{(1)}\right|^2. \tag{28}$$

Due to (27), $\mathbf{U}^T\overline{\mathbf{v}}^{(1)} = \left(\overline{\sigma}^{(1)}\right)^{-1}\left[\overline{\boldsymbol{\sigma}}^{(2)} + \mathbf{U}^T \operatorname{vec}\left(\mathbf{E}^{(2)}\right)^*\right]$. But, by definition of the SVD, the column space of $\mathbf{E}^{(2)}$ is orthogonal to $\overline{\mathbf{U}}^{(2)}$ while its row space is orthogonal to $\overline{\mathbf{V}}^{(2)^*} = \overline{\mathbf{U}}^{(3)}$. Thus, it turns out that $\mathbf{U}^T \operatorname{vec}\left(\mathbf{E}^{(2)}\right)^* = \mathbf{0}$, leading to $\mathbf{U}^T\overline{\mathbf{v}}^{(1)} = \left(\overline{\sigma}^{(1)}\right)^{-1}\overline{\boldsymbol{\sigma}}^{(2)}$. Substituting this expression into (28) yields $\left\|\overline{\sigma}^{(1)}\overline{\mathbf{v}}^{(1)^H}\mathbf{P}\right\|_2^2 = \|\overline{\boldsymbol{\sigma}}^{(2)}\|_2^2$. On the other hand, for $R_1 = 1$ the right-hand side of (22) is given by

$$\left(\overline{\sigma}^{(1)}\right)^2 \left\|\overline{\mathbf{v}}^{(1)^H}\mathbf{U}^*\right\|_2^2 = \|\overline{\boldsymbol{\sigma}}^{(2)}\|_2^2.$$

Therefore, (22) holds with equality, implying (23). ∎

Theorem 5 generalizes Theorem 2. Furthermore, together with Theorem 1, it implies the following.

COROLLARY 6. *Let $\mathcal{X} \in \bigotimes_{n=1}^3 \mathbb{C}^{I_n}$ and denote the rank-$(1, L, L)$ approximations of $\mathcal{X}$ produced by SeLRAP, SeMP and THOSVD by $\hat{\mathcal{X}}_{SeLRAP}$, $\hat{\mathcal{X}}_{SeMP}$ and $\hat{\mathcal{X}}_{THOSVD}$, respectively. Suppose that the modes are processed in the natural order (1,2,3) by both SeLRAP and SeMP. Then,*

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

The same results evidently apply to the cases $\mathbf{m} = (L, 1, L)$ and $\mathbf{m} = (L, L, 1)$, as long as the mode associated with the component $R_n = 1$ be the first one to be processed. Another consequence of Theorem 5 is that SeLRAP also satisfies the bound (13) in the rank-$(1, L, L)$ case.

COROLLARY 7. *Let $\mathcal{X} \in \bigotimes_{n=1}^3 \mathbb{C}^{I_n}$. For any solution $\mathcal{X}^\star$ of (1) with $\mathbf{m} = (1, L, L)$, the rank-$(1, L, L)$ approximation of $\mathcal{X}$ produced by SeLRAP satisfies*

$$\left\|\mathcal{X} - \hat{\mathcal{X}}_{SeLRAP}\right\|_F^2 \leq N\|\mathcal{X} - \mathcal{X}^\star\|_F^2.$$

3.3.3. The general case $\mathbf{m} = (R_1, R_2, R_3)$. For arbitrary $\mathbf{m}$, $\mathbf{W}_{\langle 2\rangle}^{(2)}$ is the mode-2 unfolding of an $R_1 \times I_2 \times I_3$ tensor whose best rank-(R_1, R_2, R_3) approximation is sought. Therefore, unlike the previous case, an explicit expression for this approximation is not available, and a proof of superiority of SeLRAP is harder to undertake. We sketch below one possible way of writing the resulting projector $\mathbf{P}$ in this case, in order to give an idea of the increased complexity. Consider the matrix

$$\mathbf{Z}_{\langle 3\rangle}^{(3)} = \overline{\mathbf{U}}^{(3)}\overline{\boldsymbol{\Sigma}}^{(3)}\overline{\mathbf{V}}^{(3)^H} \in \mathbb{C}^{I_3 \times R_2 R_1} \tag{29}$$

obtained at the end of the dimension reduction stage. Here, we introduce the notation $\mathbf{G}_{\langle 3\rangle}^{(3)} \triangleq \overline{\boldsymbol{\Sigma}}^{(3)}\overline{\mathbf{V}}^{(3)^H} \in \mathbb{C}^{R_3 \times R_2 R_1}$, which allows writing $\mathcal{Z}^{(3)} = \mathcal{G}^{(3)} \bullet_3 \overline{\mathbf{U}}^{(3)}$. Hence,

$$\mathbf{Z}_{\langle 2\rangle}^{(3)} = \mathbf{G}_{\langle 2\rangle}^{(3)}\left(\overline{\mathbf{U}}^{(3)} \boxtimes \mathbf{I}_{R_1}\right)^T \in \mathbb{C}^{R_2 \times I_3 R_1}. \tag{30}$$

Now,

$$\mathbf{Z}_{\langle 2\rangle}^{(3)^H}\left(\mathbf{Z}_{\langle 2\rangle}^{(3)}\mathbf{Z}_{\langle 2\rangle}^{(3)^H}\right)^{-1}\mathbf{Z}_{\langle 2\rangle}^{(3)} = \tag{31}$$

$$\left(\overline{\mathbf{U}}^{(3)}\boxtimes\mathbf{I}_{R_1}\right)^*\mathbf{G}_{\langle 2\rangle}^{(3)^H}\left(\mathbf{G}_{\langle 2\rangle}^{(3)}\mathbf{G}_{\langle 2\rangle}^{(3)^H}\right)^{-1}\mathbf{G}_{\langle 2\rangle}^{(3)}\left(\overline{\mathbf{U}}^{(3)}\boxtimes\mathbf{I}_{R_1}\right)^T\in\mathbb{C}^{I_3R_1\times I_3R_1}$$

Next, we define

$$\tilde{\mathbf{G}}_{\langle 2\rangle}^{(2)}\triangleq\mathbf{W}_{\langle 2\rangle}^{(2)}\left(\overline{\mathbf{U}}^{(3)}\boxtimes\mathbf{I}_{R_1}\right)^*\mathbf{G}_{\langle 2\rangle}^{(3)^H}\left(\mathbf{G}_{\langle 2\rangle}^{(3)}\mathbf{G}_{\langle 2\rangle}^{(3)^H}\right)^{-1}\mathbf{G}_{\langle 2\rangle}^{(3)}\in\mathbb{C}^{I_2\times R_3R_1}. \tag{32}$$

Since its rank is bounded by R_2, it can be decomposed as $\tilde{\mathbf{G}}_{\langle 2\rangle}^{(2)}=\tilde{\mathbf{U}}^{(2)}\mathbf{G}_{\langle 2\rangle}^{(2)}$, where $\tilde{\mathbf{U}}^{(2)}\in\mathbb{C}^{I_2\times R_2}$ has orthonormal columns (one can take, e.g., its QR decomposition). Moreover, $\mathbf{G}_{\langle 2\rangle}^{(2)}$ can be thought of as the mode-2 unfolding of a $R_1\times R_2\times R_3$ tensor $\mathcal{G}^{(2)}$. Thus,

$$\mathbf{Z}_{\langle 2\rangle}^{(2)}=\mathbf{W}_{\langle 2\rangle}^{(2)}\mathbf{Z}_{\langle 2\rangle}^{(3)^H}\left(\mathbf{Z}_{\langle 2\rangle}^{(3)}\mathbf{Z}_{\langle 2\rangle}^{(3)^H}\right)^{-1}\mathbf{Z}_{\langle 2\rangle}^{(3)}=\tilde{\mathbf{U}}^{(2)}\mathbf{G}_{\langle 2\rangle}^{(2)}\left(\overline{\mathbf{U}}^{(3)}\boxtimes\mathbf{I}_{R_1}\right)^T\in\mathbb{C}^{I_2\times I_3R_1},$$

leading to $\mathbf{Z}_{\langle 1\rangle}^{(2)}=\mathbf{G}_{\langle 1\rangle}^{(2)}\left(\overline{\mathbf{U}}^{(3)}\boxtimes\tilde{\mathbf{U}}^{(2)}\right)^T\in\mathbb{C}^{R_1\times I_3I_2}$. The projector $\mathbf{P}$ thus reads

$$\mathbf{P}=\left(\overline{\mathbf{U}}^{(3)}\boxtimes\tilde{\mathbf{U}}^{(2)}\right)^*\mathbf{G}_{\langle 1\rangle}^{(2)^H}\left(\mathbf{G}_{\langle 1\rangle}^{(2)}\mathbf{G}_{\langle 1\rangle}^{(2)^H}\right)^{-1}\mathbf{G}_{\langle 1\rangle}^{(2)}\left(\overline{\mathbf{U}}^{(3)}\boxtimes\tilde{\mathbf{U}}^{(2)}\right)^T\in\mathbb{C}^{I_3I_2\times I_3I_2}.$$

Clearly, whether $\left\|\overline{\mathbf{\Sigma}}^{(1)}\overline{\mathbf{V}}^{(1)^H}\mathbf{P}\right\|_F^2\geq\left\|\overline{\mathbf{\Sigma}}^{(1)}\overline{\mathbf{V}}^{(1)^H}\left(\overline{\mathbf{U}}^{(3)}\boxtimes\overline{\mathbf{U}}^{(2)}\right)^*\right\|_F^2$ holds or not is now a more complicated matter.

4. Decomposition in rank-$(1,L,L)$ terms via deflation. We propose in this section a deflationary algorithm whose purpose is to compute the low-mrank blocks constituting a BTD-$(1,L,L)$ of a tensor $\mathcal{X}$.

4.1. Algorithm. The proposed deflationary block-term decomposition (DBTD) algorithm is described by Algorithm 3. We have employed the symbol $\mathcal{P}_{\mathbf{m}}$ to denote an approximate projection onto $\mathcal{L}_{\mathbf{m}}\triangleq\{\mathcal{Y}:\operatorname{mrank}(\mathcal{Y})\leq\mathbf{m}\}$ which can be computed, e.g., by SeLRAP, SeMP or THOSVD. The resulting DBTD algorithm has a very simple form: at each iteration, one sequentially obtains a new estimate of each block, $\hat{\mathcal{X}}_{r,k}$, by computing an mrank-$\mathbf{m}^{(r)}$ approximation of the current residue tensor $\mathcal{E}_{r-1,k}$ plus the current block estimate $\hat{\mathcal{X}}_{r,k-1}$. The new residue tensor $\mathcal{E}_{r,k}$ then corresponds to the resulting approximation error. Because $\hat{\mathcal{X}}_{r,0}=\mathcal{O}$ for all r and the initial residue tensor equals $\mathcal{X}$, the first iteration amounts to extracting the R blocks from $\mathcal{X}$ in a greedy fashion. In general, this first iteration does not provide the sought blocks. Thus, DBTD refines these estimates from the second iteration onwards.

An appropriate stopping criterion for Algorithm 3 consists in checking whether $\|\mathcal{E}_{R,k}\|_F$ is sufficiently small or the ratio $\psi_k=\|\mathcal{E}_{R,k}\|_F\,\|\mathcal{E}_{R,k-1}\|_F^{-1}$ is close to 1. To avoid a premature stop, one can verify whether the average ψ_k among the K_s most recent iterations approximately equals 1, which yields the stopping criteria

$$\|\mathcal{E}_{R,k}\|_F\leq\epsilon_1\quad\text{or}\quad\left|1-\frac{1}{K_s}\sum_{l=0}^{K_s-1}\psi_{k-l}\right|\leq\epsilon_2, \tag{33}$$

Algorithm 3 Deflationary block-term decomposition (DBTD) algorithm.

Inputs: $\mathcal{X} \in \bigotimes_{n=1}^{N} \mathbb{C}^{I_n}$, number of blocks R and mranks $\mathbf{m}^{(r)} = (R_1^{(r)}, \ldots, R_N^{(r)})$
Output: Estimates of the blocks $\hat{\mathcal{X}}_r \in \mathcal{L}_{\mathbf{m}^{(r)}}$, $r = 1, \ldots, R$

$\hat{\mathcal{X}}_{r,0} \leftarrow \mathcal{O}$, for $r = 1, \ldots, R$
$\mathcal{E}_{0,1} \leftarrow \mathcal{X}$
$k \leftarrow 0$
while $k < K$ and (33) is not satisfied **do**
  $k \leftarrow k + 1$
  for $r = 1, \ldots, R$ **do**
    $\mathcal{E}_{r,k} \leftarrow \mathcal{E}_{r-1,k} + \hat{\mathcal{X}}_{r,k-1}$
    $\hat{\mathcal{X}}_{r,k} \leftarrow \mathcal{P}_{\mathbf{m}^{(r)}}(\mathcal{E}_{r,k})$
    $\mathcal{E}_{r,k} \leftarrow \mathcal{E}_{r,k} - \hat{\mathcal{X}}_{r,k}$
  end for
  $\mathcal{E}_{0,k+1} \leftarrow \mathcal{E}_{R,k}$
end while
$\hat{\mathcal{X}}_r \leftarrow \hat{\mathcal{X}}_{r,k}$, for $r = 1, \ldots, R$

for two sufficiently small constants ϵ_1, ϵ_2. (In practice, one usually imposes also a maximum number of iterations, K).

The cost of a DBTD iteration equals that of two additions of two $I_1 \times \cdots \times I_N$ tensors plus R applications of $\mathcal{P}_{\mathbf{m}^{(r)}}$ with the R specified block mranks:

$$C_{\text{DBTD}} = \mathcal{O}(\bar{I}) + \textstyle\sum_{r=1}^{R} C_{\mathcal{P}_{\mathbf{m}^{(r)}}}, \tag{34}$$

where $C_{\mathcal{P}_{\mathbf{m}^{(r)}}}$ denotes the cost of the LMA method. In particular, if $\mathbf{m}^{(r)} = \mathbf{m} = (1, L, L)$ and SeLRAP is used, we have

$$C_{\mathcal{P}_{\mathbf{m}}} = C_{\text{SVD}}(I_1, I_2 I_3) + \mathcal{O}(I_1 I_2 I_3) + C_{\text{SVD}}(I_2, I_3) + \mathcal{O}(L I_2 I_3) + C_{\text{proj}}(I_2 I_3, 1, I_1),$$

with $C_{\text{proj}}(I_2 I_3, 1, I_1) = \mathcal{O}(I_2 I_3) + \mathcal{O}(I_1 I_2 I_3)$. To derive this expression, we have taken into account the simplifications which apply to SeLRAP when $\mathbf{m} = (1, L, L)$, as described in subsection 3.3.2.

4.2. Discussion on convergence. The partial convergence analysis presented in [2] can be straightforwardly extended to our present case. In the following, we briefly discuss the main implication of such an extension and the assumption which underlies it. Namely, if we assume that the best rank-$\mathbf{m}^{(r)}$ approximation is achieved by $\mathcal{P}_{\mathbf{m}^{(r)}}$ when updating the rth block, then the following result holds.

PROPOSITION 8. *If $\mathcal{P}_{\mathbf{m}^{(r)}}$ delivers the best rank-$\mathbf{m}^{(r)}$ approximation of $\mathcal{E}_{r-1,k} + \mathcal{X}_{r,k-1}$ for all r and k, then $\|\mathcal{E}_{R,k}\|_F \leq \|\mathcal{E}_{R,k-1}\|$ for all k.*

Proof. Let $\mathcal{P}_{\mathcal{X}_{r,k-1}}$ denote the orthogonal projection onto the modal subspaces of $\mathcal{X}_{r,k-1}$, i.e., onto $\bigotimes_{n=1}^{N} \text{span}((\mathcal{X}_{r,k-1})_{\langle n \rangle})$. Since $\text{mrank}(\mathcal{P}_{\mathcal{X}_{r,k-1}}(\mathcal{Y})) \leq \mathbf{m}^{(r)}$ for any $\mathcal{Y}$, from the optimality of $\mathcal{P}_{\mathbf{m}^{(r)}}$ we have

$$\begin{aligned} &\| (\mathcal{E}_{r-1,k} + \mathcal{X}_{r,k-1}) - \mathcal{P}_{\mathbf{m}^{(r)}}(\mathcal{E}_{r-1,k} + \mathcal{X}_{r,k-1}) \|_F \\ &\qquad \leq \| (\mathcal{E}_{r-1,k} + \mathcal{X}_{r,k-1}) - \mathcal{P}_{\mathcal{X}_{r,k-1}}(\mathcal{E}_{r-1,k} + \mathcal{X}_{r,k-1}) \|_F. \end{aligned} \tag{35}$$

But, by definition of $\mathcal{P}_{\mathcal{X}_{r,k-1}}$, the right-hand side cannot be larger than $\|\mathcal{E}_{r-1,k}\|_F$. Since $\hat{\mathcal{X}}_{r,k} = \mathcal{P}_{\mathbf{m}^{(r)}}(\mathcal{E}_{r-1,k} + \mathcal{X}_{r,k-1})$ by construction, then the left-hand side of the

TABLE 2
Statistics of Δ for third-order tensors

Scenario			Δ(THOSVD)		Δ(SeMP)	
#	**i**	**r**	mean	std. dev.	mean	std. dev.
1	(40,80,200)	(1,2,2)	8.61e-04	1.58e-05	2.86e-05	4.55e-06
2	(40,80,200)	(2,4,8)	2.55e-03	2.98e-05	5.36e-05	6.17e-06
3	(40,80,200)	(10,20,50)	1.38e-02	1.22e-04	5.63e-04	1.55e-05
4	(200,80,40)	(2,2,1)	6.33e-04	2.23e-05	4.27e-04	2.44e-05
5	(200,80,40)	(8,4,2)	1.88e-03	3.98e-05	1.41e-03	3.80e-05
6	(200,80,40)	(50,20,10)	9.24e-03	1.26e-04	5.81e-03	6.16e-05
7	(40,200,200)	(1,5,5)	1.20e-03	1.15e-05	1.11e-05	1.78e-06
8	(40,200,200)	(2,10,10)	2.24e-03	2.01e-05	2.50e-04	6.78e-06
9	(40,200,200)	(10,50,50)	1.11e-02	7.40e-05	1.26e-03	1.46e-05
10	(200,200,200)	(5,5,5)	3.10e-04	3.87e-06	1.07e-04	2.48e-06
11	(200,200,200)	(10,10,10)	6.73e-04	6.18e-06	2.15e-04	3.50e-06
12	(200,200,200)	(50,50,50)	4.44e-03	2.95e-05	7.47e-04	6.14e-06

above inequality is precisely the norm of the residue $\boldsymbol{\mathcal{E}}_{r,k}$. Therefore, at the end of the kth iteration, one has $\|\boldsymbol{\mathcal{E}}_{R,k}\|_F \leq \|\boldsymbol{\mathcal{E}}_{R-1,k}\|_F \leq \cdots \leq \|\boldsymbol{\mathcal{E}}_{0,k}\|_F = \|\boldsymbol{\mathcal{E}}_{R,k-1}\|_F$. ∎

Under optimality of $\mathscr{P}_{\mathbf{m}^{(r)}}$, the above result implies $\|\boldsymbol{\mathcal{E}}_{R,k}\|_F \to C$ for some $C \geq 0$, since the sequence $\|\boldsymbol{\mathcal{E}}_{R,k}\|_F$ is non-negative and decreases monotonically with k. (Consequently, the second criterion of (33) is eventually satisfied.) However, in practice one can only resort to sub-optimal approximation schemes such as SeLRAP. Nonetheless, note that $\mathscr{P}_{\mathbf{m}^{(r)}}$ doesn't have to be optimal, but only at least as accurate as $\mathscr{P}_{\boldsymbol{\mathcal{X}}_{r,k-1}}$. Whilst it is currently unclear whether such a weaker condition can be proven, our numerical results suggest that it indeed holds in practice (at least with high probability), as a strictly monotonic decrease of $\|\boldsymbol{\mathcal{E}}_k\|_F$ is generally observed. We point out that there is some similarity between this discussion and that concerning the convergence of IHT schemes employing suboptimal LMAs for low-rank tensor recovery [21, 14].

5. 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 $\mathbf{i} = (I_1, \ldots, I_N)$ specifies the tensor dimensions in each scenario.

5.1. Non-iterative low-mrank approximation. First, we compare the performance of SeLRAP with respect to those of THOSVD and SeMP in the task of LMA. To do so, similarly to [3] we measure

$$\Delta(\text{Alg}) = 1 - \frac{\|\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}_{\text{SeLRAP}}\|_F}{\|\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}_{\text{Alg}}\|_F},$$

with $\text{Alg} \in \{\text{THOSVD}, \text{SeMP}\}$, for 10^4 realizations of third- and fourth-order complex tensors having entries whose real and imaginary parts are drawn from $(-1, 1)$. We have varied the dimensions and target mranks, yielding 12 different scenarios whose results are reported in Table 2. In all these scenarios, Δ was strictly positive for all realizations, meaning SeLRAP always found a better approximation. Note that Corollary 6 only guarantees that for scenarios 1 and 7. However, inspecting Table 2 we see that the average improvement in accuracy is generally small. This is similar to the conclusions reached by [25] concerning the comparison of SeMP with THOSVD.

Figure 1 displays, for six selected scenarios, the empirical cumulative distribution function (ECDF) of the time spent by each algorithm at each operation specified by Table 1. When the target mrank is quite small, SeMP and SeLRAP have much smaller overall computing time than THOSVD, as seen in Figure 1{(a),(c),(e)}. In Figure 1{(a),(c)}, this is mainly due to the fact that only two SVDs are required by SeMP and SeLRAP. In Figure 1(e), the gain comes from the dimension-reducing contractions which alleviate the cost of the SVDs. This observation also applies to the second SVD computed in Figure 1{(a),(c)}. Moreover, note that in thee cases the projection stage of SeLRAP is faster than the operations which construct the LMA in THOSVD and SeMP. The gap between THOSVD and the other algorithms is greatly reduced when mrank components are increased, cf. Figure 1{(b),(d),(f)}, since the ratio $H_n/\bar{I}_n$ is increased. Furthermore, the backward projections of SeLRAP now take longer than operations (III) of THOSVD and SeMP, causing its overall time to overcome that of SeMP.

Note that the dimensions increase with the mode number in all scenarios of Figure 1. As we process them in the natural order, this choice is consistent with the heuristic mentioned in subsection 3.2. For reference, the average overall times of SeLRAP in scenarios 1 and 3 of Table 2 are 0.0228 sec and 0.1497 sec, respectively, while in scenarios 4 and 6 (where $\mathcal{X}$ has the same dimensions but in reverse order) these times are 0.1255 sec and 0.1688 sec, respectively. The advantage of adopting the heuristic is thus more pronounced for smaller ratios R_n/I_n.

5.2. Decomposition in rank-$(1, L, L)$ terms. In this section, we compare DBTD with existing BTD-$(1, L, L)$ computation algorithms. Three variants of DBTD are considered, depending on the choice of $\mathscr{P}_{\mathbf{m}}$: DBTD-SeLRAP, DBTD-SeMP and DBTD-THOSVD. The other included algorithms are ALS-ELS [18] and the Gauss-Newton algorithm with dogleg trust region (GN-DL) of Tensorlab [26]. Among the algorithms provided by Tensorlab, the latter has been chosen for its superior performance in our simulations, consonantly with the findings of [23].

Our simulation scenarios encompass various levels of average correlation among the rank-$(1, L, L)$ blocks. To this end, given a target correlation coefficient $\rho \in [0, 1)$, the blocks are given by

$$\overline{\mathcal{X}}_r = \|\mathcal{X}_r\|_F^{-1} \mathcal{X}_r, \quad \mathcal{X}_r = \mathbf{a}_r \otimes \left(\mathbf{B}_r \mathbf{C}_r^T\right), \tag{36}$$

where $\mathbf{a} \in \mathbb{C}^{I_1}$, $\mathbf{B} \in \mathbb{C}^{I_2 \times L}$ and $\mathbf{C} \in \mathbb{C}^{I_3 \times L}$, are generated as $\begin{bmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_R\end{bmatrix} = \mathbf{Q}_a\mathbf{J}$, $\begin{bmatrix}\mathrm{vec}(\mathbf{B}_1) & \dots & \mathrm{vec}(\mathbf{B}_R)\end{bmatrix} = \mathbf{Q}_b\mathbf{J}$ and $\begin{bmatrix}\mathrm{vec}(\mathbf{C}_1) & \dots & \mathrm{vec}(\mathbf{C}_R)\end{bmatrix} = \mathbf{Q}_c\mathbf{J}$, with $\mathbf{Q}_a \in \mathcal{S}(I_1, R)$, $\mathbf{Q}_b \in \mathcal{S}(LI_2, R)$ and $\mathbf{Q}_C \in \mathcal{S}(LI_3, R)$ denoting random column-wise orthonormal matrices and

$$\mathbf{J}^T\mathbf{J} = \begin{bmatrix} 1 & \rho^{1/3} & \dots & \rho^{1/3} \\ \rho^{1/3} & 1 & \dots & \rho^{1/3} \\ \vdots & & \ddots & \vdots \\ \rho^{1/3} & \rho^{1/3} & \dots & 1 \end{bmatrix}. \tag{37}$$

Since $\langle \mathcal{X}_r, \mathcal{X}_s \rangle = \langle \mathbf{a}_r, \mathbf{a}_s \rangle \left\langle \mathbf{B}_r\mathbf{C}_r^T, \mathbf{B}_s\mathbf{C}_s^T \right\rangle = \langle \mathbf{a}_r, \mathbf{a}_s \rangle \,\mathrm{Tr}\left\{\mathbf{C}_r^T\mathbf{C}_s^*\,\mathbf{B}_s^H\mathbf{B}_r\right\}$, for $r = s$ we have $\|\mathcal{X}_r\|_F^2 = \mathrm{Tr}\left\{\mathbf{C}_s^T\mathbf{C}_s^*\,\mathbf{B}_s^H\mathbf{B}_s\right\}$ and, for $r \neq s$,

$$\left|\left\langle \overline{\mathcal{X}}_r, \overline{\mathcal{X}}_s \right\rangle\right| = \rho^{1/3} \left(\mathrm{Tr}\left\{\mathbf{C}_r^T\mathbf{C}_r^*\,\mathbf{B}_r^H\mathbf{B}_r\right\} \mathrm{Tr}\left\{\mathbf{C}_s^T\mathbf{C}_s^*\,\mathbf{B}_s^H\mathbf{B}_s\right\}\right)^{-\frac{1}{2}} \left|\mathrm{Tr}\left\{\mathbf{C}_r^T\mathbf{C}_s^*\,\mathbf{B}_s^H\mathbf{B}_r\right\}\right|.$$

If $\mathbf{Q}_a$, $\mathbf{Q}_b$ and $\mathbf{Q}_c$ are generated by orthogonalizing random matrices having standard circularly symmetric Gaussian (SCSG) entries, then $\left\langle \overline{\mathcal{X}}_r, \overline{\mathcal{X}}_s \right\rangle$ is normally distributed

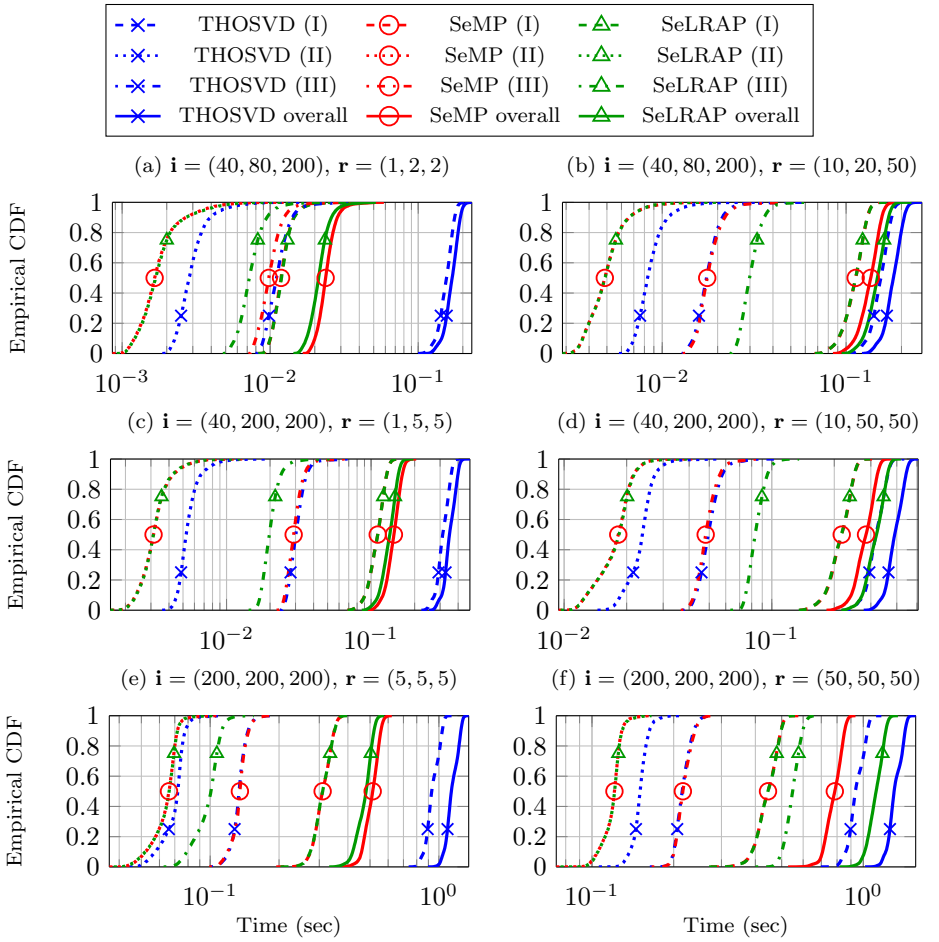

FIGURE 1. *Empirical CDFs of the times spent by each LMA algorithm at each stage.*

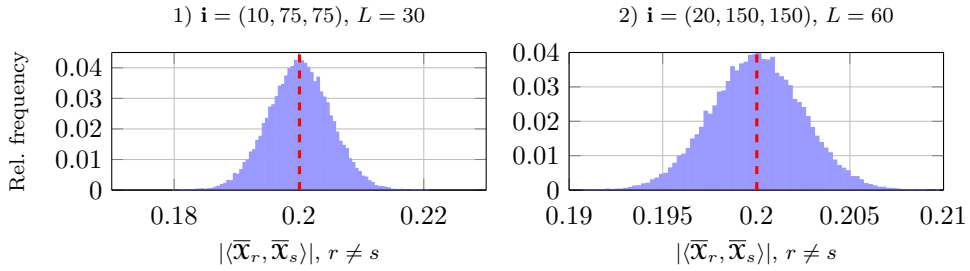

FIGURE 2. *Histogram of measured correlation among blocks generated with $\rho = 0.2$. The red dashed line indicates the sample mean.*

around $\mathbb{E}\{\langle \overline{\mathbf{x}}_r, \overline{\mathbf{x}}_s \rangle\} = \rho$ with a small standard deviation, which decays with the tensor dimensions. The histograms in Figure 2 illustrate this behavior for $\rho = 0.2$ in two cases: 1) $\mathbf{i} = (10, 75, 75)$, $L = 30$ and 2) $\mathbf{i} = (20, 150, 150)$, $L = 60$.

We consider tensors of dimensions $\mathbf{i} = (20, 150, 150)$ composed by three rank-$(1, 60, 60)$ blocks. For this choice of L, the uniqueness theorems of [4] do not hold. In

particular, the algebraic solution via generalized eigenvalue decomposition (GEVD) of [4, Theorem 4.1] does not apply. Nevertheless, with blocks generated as above, the decomposition is almost surely unique [5, Theorem 2.4]. The complete model reads

$$\boldsymbol{\mathcal{X}} = \textstyle\sum_{r=1}^{3} \lambda_r \, \overline{\boldsymbol{\mathcal{X}}}_r + \sigma \boldsymbol{\mathcal{N}}, \tag{38}$$

where $\boldsymbol{\mathcal{N}}$ has SCSG entries and σ is adjusted to impose a desired signal-to-noise ratio

$$\text{SNR} = \left\| \textstyle\sum_{r=1}^{3} \lambda_r \, \overline{\boldsymbol{\mathcal{X}}}_r \right\|_F^2 \sigma^{-2} \|\boldsymbol{\mathcal{N}}\|_F^{-2}. \tag{39}$$

In the first scenario, for each realization of $\boldsymbol{\mathcal{X}}$ we draw the weights λ_r independently from $\mathcal{N}(1, 0.2)$ and take $\rho \in \{0, 0.2, 0.4\}$. In ALS-ELS and GN-DL, THOSVD is initially applied to compress $\boldsymbol{\mathcal{X}}$ for reducing cost. Then, after a compressed solution is found, it is decompressed and refined. The stopping criteria of all algorithms were adjusted for accurately recovering the blocks while keeping computing time reasonably low.

The results for 200 realizations with SNR = 50 dB are shown in Figure 3. Specifically, we plot the ECDFs of the average normalized squared error (ANSE) over the blocks, defined as

$$\text{ANSE} = \tfrac{1}{3} \textstyle\sum_{r=1}^{3} \lambda_r^{-2} \, \|\lambda_r \, \overline{\boldsymbol{\mathcal{X}}}_r - \hat{\boldsymbol{\mathcal{X}}}_r\|_F, \tag{40}$$

and of the time spent by each algorithm. The superiority of DBTD for small ρ is clear. However, as ρ is increased, the amount of iterations required by DBTD quickly grows. For $\rho = 0.4$, the established maximum number of iterations $K = 400$ causes an early stop of all DBTD variants, which explains its minimum achieved ANSE of around -50 dB. Nonetheless, the mean ANSE of DBTD-SeLRAP is lower (-38.8 dB) than that of ALS-ELS (-35.3 dB). For even larger values of ρ, a higher K is required, and thus DBTD is not recommended due to the added computing cost. Among the DBTD variants, DBTD-SeLRAP attains the best compromise between cost and estimation accuracy, as its iterations are the least costly. Perhaps somewhat surprisingly, the estimation performance of ALS-ELS and GN-DL is poorer for $\rho = 0$ than for $\rho = 0.2$. This is due to the larger proportion of realizations for which these algorithms enter into some region of very slow convergence and are unable to achieve sufficient progress within reasonable time, despite the fact that their convergence is faster for the other realizations.

Figure 4 shows the results of a similar scenario, still with SNR = 50 dB, but this time the weights λ_r are drawn from $\mathcal{N}(1, 0.1)$. The better conditioning due to the less disparate block norms explains the better performances in comparison with the previous scenario, especially for the ALS-ELS and GN-DL algorithms. The DBTD algorithm, on the other hand, seems less sensitive in this regard. Overall, DBTD-SeLRAP still provides the best performance.

In Figure 5, we fix $\rho = 0.2$ and vary L and SNR separately, with $\lambda_r \sim \mathcal{N}(1, 0.1)$. With $L = 30$ (and SNR = 50 dB), it becomes possible to algebraically compute an approximate solution via a GEVD, because $RL = 90 < \min\{I_2, I_3\} = 150$ [4, Theorem 4.1]. We therefore initialize all algorithms with this solution. As seen in Figure 5, they are all able to satisfactorily refine it withing a few iterations, with ALS-ELS and GN-DL achieving the best overall performance due to their low computing times. For $L = 90$ and same SNR, the results of ALS-ELS and GN-DL are considerably degraded with respect to the $L = 60$ case. By contrast, DBTD is still able to accurately estimate the factors in the vast majority of realizations, though it typically needs

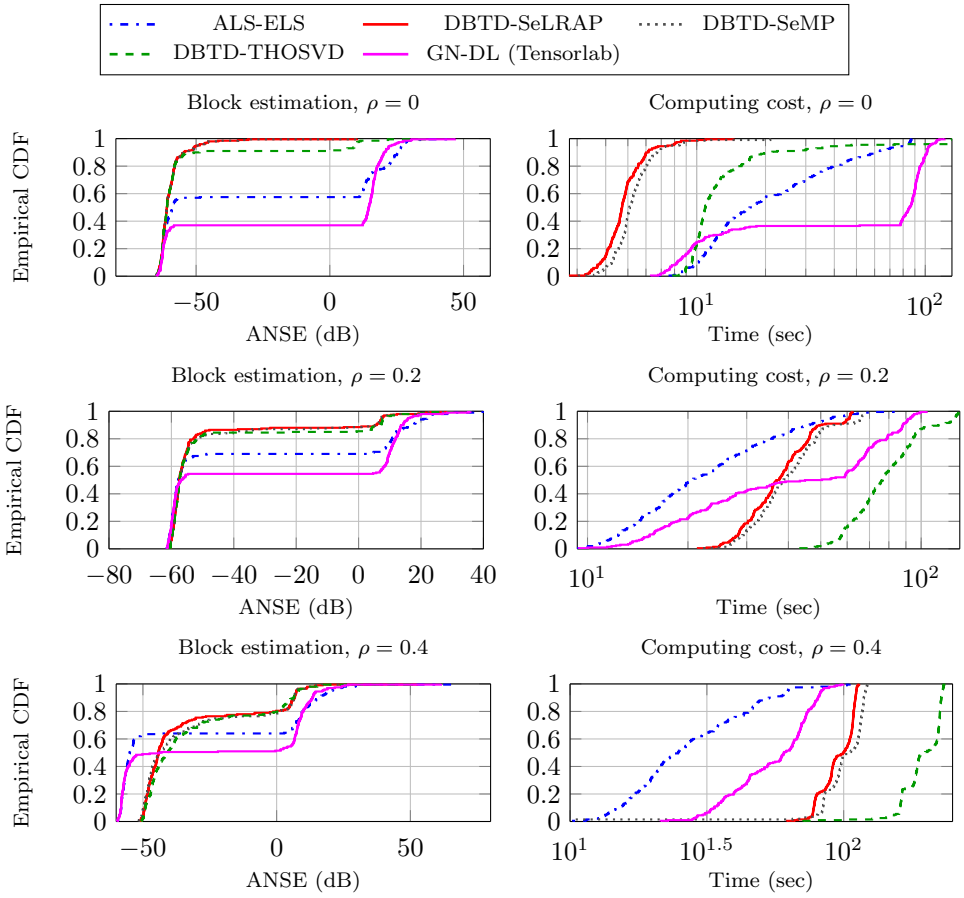

FIGURE 3. *Performance of BTD-*$(1, L, L)$ *algorithms for* $\mathbf{i} = (20, 150, 150)$, $R = 3$, $L = 60$, *SNR* $= 50$ *dB,* $\rho \in \{0, 0.2, 0.4\}$ *and* $\lambda_r \sim \mathcal{N}(1, 0.2)$.

more iterations (here, we have set $K = 500$, against $K = 200$ in the case $L = 60$). Finally, if we fix $L = 60$ and take SNR $\in \{20, 80\}$ dB, the conclusions are similar to the case where SNR $= 50$ dB; the most perceptible difference is perhaps the increased computing time of ALS-ELS when SNR $= 20$ dB.

6. Conclusion. We have proposed a novel non-iterative low-multilinear-rank approximation algorithm, SeLRAP, which generalizes the recently proposed rank-one approximation algorithm SeROAP. As we have demonstrated, this algorithm always performs at least as well as the truncated HOSVD and SeMP (also known as sequentially truncated HOSVD) for rank-$(1, L, L)$ approximation. In our numerical experiments with third-order random tensors, SeLRAP's backward projection stage was actually able to improve upon SeMP's solution for all employed target mranks, though generally by a small margin. Moreover, for small mranks it requires less computing effort.

As a second contribution, we have proposed an iterative deflationary algorithm named DBTD for decomposing a tensor in mrank-reduced block terms. This algorithm is in effect a generalization of the deflationary solution proposed for computing the canonical polyadic decomposition, DCPD. Despite the generality of DBTD, we have

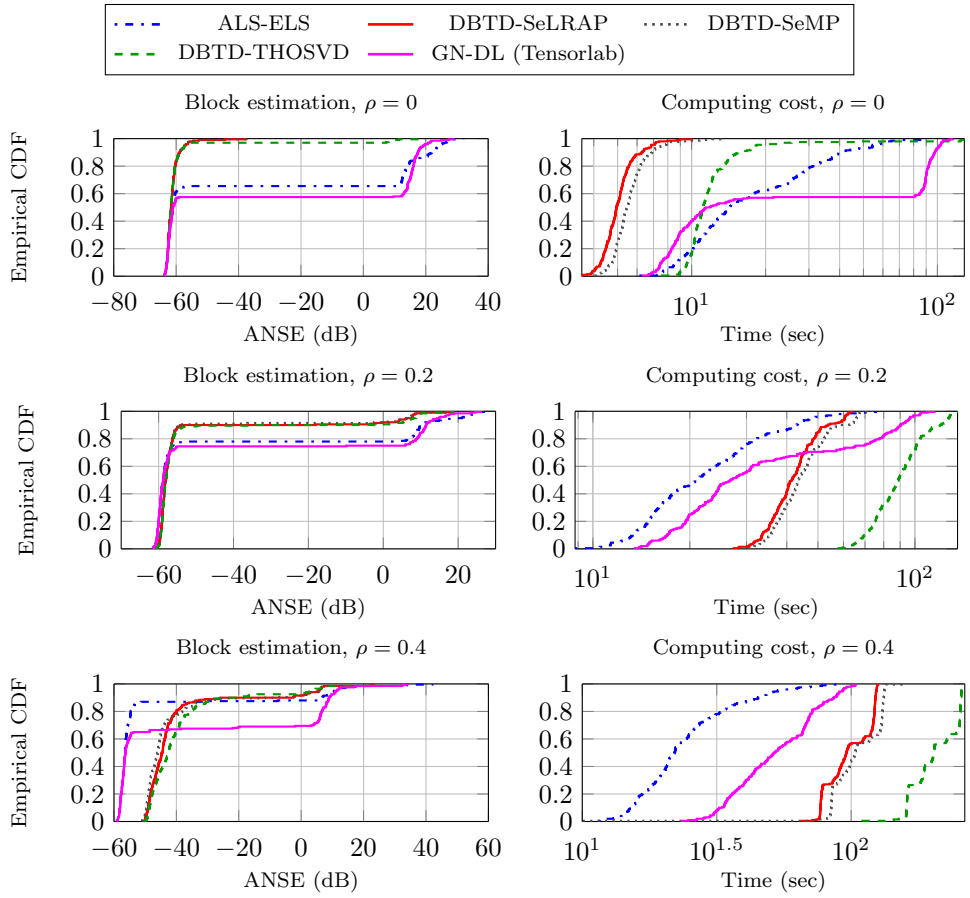

FIGURE 4. *Performance of BTD-*$(1, L, L)$ *algorithms for* $\mathbf{i} = (20, 150, 150)$, $R = 3$, $L = 60$, $SNR = 50$ *dB*, $\rho \in \{0.2, 0.4, 0.6\}$ *and* $\lambda_r \sim \mathcal{N}(1, 0.1)$.

kept our focus on the computation of rank-$(1, L, L)$ blocks. Our simulations show that, outside the regime where an approximate algebraic solution can be computed, DBTD is much more effective than existing algorithms whenever the correlation among blocks is low. Interestingly, under these conditions it was much less sensitive than other alternatives with respect to the discrepancy among the norms of the different blocks.

At the theoretical level, future research should attempt to establish convergence results for DBTD, perhaps imposing conditions for ensuring that (35) is met by SeLRAP. The more general BTD-$(1, L_r, L_r)$, where blocks have different mranks, is also left for future consideration. In this case, additional measures could possibly be taken to avoid local minima corresponding to wrong mrank matchings.

REFERENCES

[1] J. CULLUM AND R. WILLOUGHBY, *Lanczos Algorithms*, vol. I, Birkhauser, 1985.
[2] A. P. DA SILVA, P. COMON, AND A. L. F. DE ALMEIDA, *Rank-1 tensor approximation methods and application to deflation*, arXiv preprint arXiv:1508.05273, (2015).
[3] A. P. DA SILVA, P. COMON, AND A. L. F. DE ALMEIDA, *A finite algorithm to compute rank-1 tensor approximations*, IEEE Signal Processing Letters, 23 (2016), pp. 959–963.

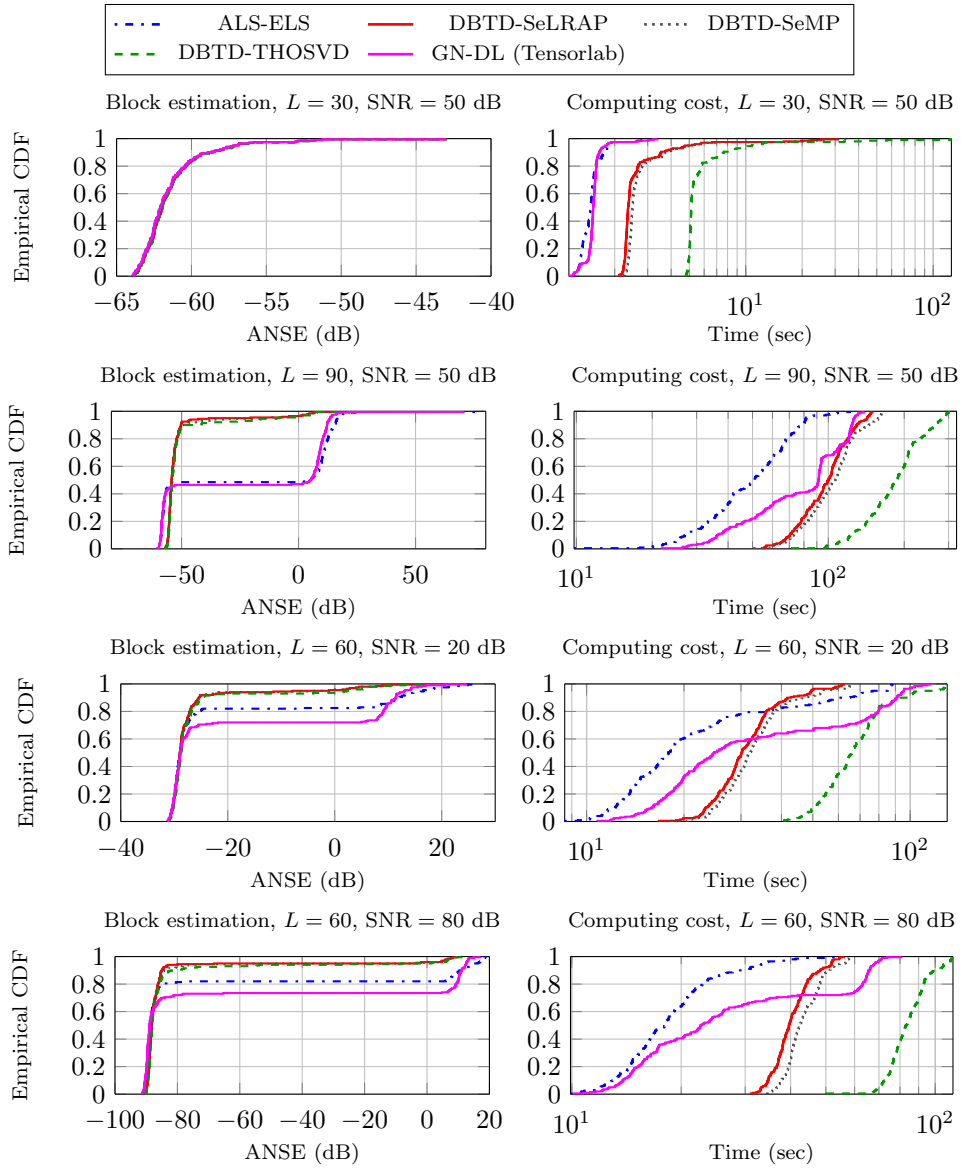

FIGURE 5. *Performance of BTD-*$(1, L, L)$ *algorithms for* $\mathbf{i} = (20, 150, 150)$, $R = 3$, $\rho = 0.2$, $\lambda_r \sim \mathcal{N}(1, 0.1)$, $L \in \{30, 90\}$ *and SNR* $\in \{20, 80\}$ *dB.*

[4] L. DE LATHAUWER, *Decompositions of a higher-order tensor in block terms–Part II: Definitions and uniqueness*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1033–1066.

[5] ———, *Blind separation of exponential polynomials and the decomposition of a tensor in rank-*$(L_r, L_r, 1)$ *terms*, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 1451–1474.

[6] L. DE LATHAUWER AND A. DE BAYNAST, *Blind deconvolution of ds-cdma signals by means of decomposition in rank-*$(1, L, L)$ *terms*, IEEE Transactions on Signal Processing, 56 (2008), pp. 1562–1571.

[7] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *A multilinear singular value decomposition*, SIAM Journal on Matrix Analysis and Applications, 21 (2000), pp. 1253–1278.

[8] L. De Lathauwer, B. De Moor, and J. Vandewalle, *On the best rank-1 and rank-(r_1,r_2,...,r_n) approximation of higher-order tensors*, SIAM J. Matrix Anal. Applicat., 21 (2000), pp. 1324–1342.

[9] L. De Lathauwer and D. Nion, *Decompositions of a higher-order tensor in block terms–Part III: Alternating least squares algorithms*, SIAM journal on Matrix Analysis and Applications, 30 (2008), pp. 1067–1083.

[10] V. De Silva and L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1084–1127.

[11] L. Eldén and B. Savas, *A Newton–Grassmann method for computing the best multilinear rank-(r_1,r_2,r_3) approximation of a tensor*, SIAM J. Matrix Anal. Applicat., 31 (2009), pp. 248–271.

[12] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Matrix Computations, Johns Hopkins University Press, 2012.

[13] J. H. de M. Goulart and P. Comon, *A novel non-iterative algorithm for low-multilinear-rank tensor approximation*, in European Signal Processing Conference (EUSIPCO), Kos, Greece, Sept. 2017. (submitted).

[14] J. H. de M. Goulart and G. Favier, *Low-rank tensor recovery using sequentially optimal modal projections in iterative hard thresholding (SeMPIHT)*, SIAM Journal on Scientific Computing, (2017). (to appear).

[15] 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*, 56 (2008), pp. 3198–3213.

[16] 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., 34 (2013), pp. 651–672.

[17] 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.

[18] D. Nion and L. De Lathauwer, *An enhanced line search scheme for complex-valued tensor decompositions. application in DS-CDMA*, Signal Processing, 88 (2008), pp. 749–755.

[19] D. Nion and N. D. Sidiropoulos, *Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar*, IEEE Transactions on Signal Processing, 58 (2010), pp. 5693–5705.

[20] H. Rauhut, R. Schneider, and Z. Stojanac, *Low rank tensor recovery via iterative hard thresholding*, in Proceedings of the 10th International Conference on Sampling Theory and Applications, 2013.

[21] H. Rauhut, R. Schneider, and Ž. Stojanac, *Low rank tensor recovery via iterative hard thresholding*, Linear Algebra and its Applications, 523 (2017), pp. 220–262.

[22] L. N. Ribeiro, A. L. F. De Almeida, and V. Zarzoso, *Enhanced block term decomposition for atrial activity extraction in atrial fibrillation ecg*, in IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), IEEE, 2016, pp. 1–5.

[23] L. Sorber, M. Van Barel, and L. De Lathauwer, *Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-($L_r, L_r, 1$) terms, and a new generalization*, SIAM Journal on Optimization, 23 (2013), pp. 695–720.

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

[25] N. Vannieuwenhoven, R. Vandebril, and K. Meerbergen, *A new truncation strategy for the higher-order singular value decomposition*, SIAM Journal on Scientific Computing, 34 (2012), pp. A1027–A1052.

[26] N. Vervliet, O. Debals, L. Sorber, M. V. Barel, and L. De Lathauwer, *Tensorlab 3.0*, Mar. 2016. Available online. URL: http://www.tensorlab.net.

[27] A. Wołczowski and R. Zdunek, *Electromyography and mechanomyography signal recognition: Experimental analysis using multi-way array decomposition methods*, Biocybern. Biomed. Eng., 37 (2017), pp. 103–113.

[28] T. Zhang and G. H. Golub, *Rank-one approximation to high order tensors*, SIAM Journal on Matrix Analysis and Applications, 23 (2001), pp. 534–550.