



**HAL**  
open science

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

José Henrique de Morais Goulart, Pierre Comon

## ► To cite this version:

José Henrique de Morais Goulart, Pierre Comon. Non-iterative low-multilinear-rank tensor approximation with application to decomposition in rank-(1,L,L) terms. 2017. hal-01516167

**HAL Id: hal-01516167**

**<https://hal.science/hal-01516167>**

Preprint submitted on 28 Apr 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.

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

4                    JOSÉ HENRIQUE DE MORAIS GOULART<sup>†</sup> AND PIERRE COMON<sup>†</sup>

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

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

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

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

29                    (1)                    
$$\min_{\hat{\mathcal{X}} \in \bigotimes_{n=1}^N \mathcal{U}_n} \|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 \quad \text{subj. to} \quad \begin{cases} \mathcal{U}_n \subseteq \mathbb{C}^{I_n} \\ \dim(\mathcal{U}_n) = R_n \end{cases} \quad \text{for } n = 1, \dots, N,$$

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

36                    A direct connection exists between mrank and the so-called Tucker decomposition:  
37                    every finite-dimensional complex rank- $(R_1, \dots, R_N)$  tensor<sup>1</sup>  $\hat{\mathcal{X}} \in \bigotimes_{n=1}^N \mathbb{C}^{I_n}$  can be  
38                    expressed in the form

39                    (2)                    
$$\hat{\mathcal{X}} = \mathcal{G} \underset{n=1}{\bullet} \mathbf{U}^{(n)} \triangleq \mathcal{G} \bullet_1 \mathbf{U}^{(1)} \bullet_2 \dots \bullet_N \mathbf{U}^{(N)},$$

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

<sup>†</sup>Univ. Grenoble Alpes, CNRS, Gipsa-Lab, F-38000 Grenoble ([jose-henrique.de-morais-goulart@gipsa-lab.fr](mailto:jose-henrique.de-morais-goulart@gipsa-lab.fr), [pierre.comon@gipsa-lab.fr](mailto: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.”

<sup>1</sup>Though  $(R_1, \dots, R_N)$  is the *mrank* of  $\hat{\mathcal{X}}$  rather than its rank, we employ the usual terminology “rank- $(R_1, \dots, R_N)$ ” without confusion since we exclusively work with the mrank throughout.

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

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

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

$$73 \quad (3) \quad \|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 \leq N \|\mathcal{X} - \mathcal{X}^*\|_F^2, \quad \text{where } \mathcal{X}^* \text{ is a minimizer of (1).}$$

74 The alternative proposed in [25], which we refer to as sequentially optimal modal  
 75 projections (SeMP), is less computationally intensive than the THOSVD, especially  
 76 for small dimensions  $R_n$ . It also computes  $N$  SVDs, but they are interleaved with  
 77 contractions which gradually reduce the tensor dimensions, each one being optimal  
 78 given the preceding ones. Moreover, the resulting approximation error also obeys the  
 79 bound in (3). In particular, for rank- $(1, L, L)$  approximations SeMP was shown to  
 80 perform at least as well as THOSVD. Simulation results presented in [25] with ran-  
 81 dom tensors suggest that this superiority actually holds in most cases. Concerning the  
 82 special case of rank-one approximations, [3] has come up with a two-stage algorithm  
 83 called sequential rank-one approximation and projection (SeROAP). It first reduces  
 84 dimensionality similarly to SeMP, and then performs a sequence of “backward” pro-  
 85 jections to refine the approximation. For third-order tensors, it has been proven in  
 86 [3] that SeROAP performs at least as well as SeMP and, consequently, as THOSVD  
 87 too.

88 **1.1. A generalization of SeROAP.** Our first contribution is a generalization  
 89 of SeROAP to arbitrary target mranks. We call this algorithm sequential low-rank  
 90 approximation and projection (SeLRAP). In the case of third-order tensors, we show  
 91 that SeLRAP performs at least as well as SeMP (and, consequently, as THOSVD) for  
 92 rank- $(1, L, L)$  approximations. Despite the lack of proof for other cases, such a supe-  
 93 riority was also observed in the overwhelming majority of our numerical simulations.  
 94 This situation is similar to that described in [25] concerning the relationship between  
 95 SeMP and THOSVD.

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

101 **1.2. Decomposition in rank- $(1, L, L)$  terms via deflation.** The block term  
 102 decomposition (BTD) problem consists in decomposing a tensor  $\mathcal{Y} \in \bigotimes_{n=1}^N \mathbb{C}^{I_n}$  as [4]

103 (4) 
$$\mathcal{Y} = \sum_{r=1}^R \mathcal{Y}_r, \quad \text{such that } \text{mrk}(\mathcal{Y}_r) \leq \mathbf{m}^{(r)} = (R_1^{(r)}, \dots, R_N^{(r)}).$$

104 In particular, when  $N = 3$  and  $\mathbf{m}^{(r)} = (1, L, L)$  for all  $r$ , it is known as the decompo-  
 105 sition of  $\mathcal{Y}$  in rank- $(1, L, L)$  terms,<sup>2</sup> and can be written as

106 (5) 
$$\mathcal{Y} = \sum_{r=1}^R \mathbf{a}_r \otimes (\mathbf{B}_r \mathbf{C}_r^T),$$

107 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  
 108 denote by BTD- $(1, L, L)$ , has received a great deal of attention in the literature due  
 109 to the various applications it finds—examples include, e.g., blind deconvolution [6],  
 110 multidimensional harmonic retrieval [19], blind source separation [22] and electron  
 111 energy loss spectroscopy [24].

112 On the theoretical side, conditions for the uniqueness of the blocks of a BTD-  
 113  $(1, L, L)$  have been derived (up to a permutation of their indices) in [4, 5]. Such  
 114 conditions are of central importance in applications because these blocks are typically  
 115 computed as a means of estimating some quantities of interest. For the numerical  
 116 computation of (5), an ALS algorithm has been put forth by [9]. In [18], an enhanced  
 117 line search (ELS) scheme with exact (complex) step computation is incorporated into  
 118 this algorithm, greatly improving its convergence speed. More recently, [23] has pro-  
 119 posed conjugate gradient, quasi-Newton, Gauss-Newton and Levenberg-Marquardt  
 120 algorithms for the BTD- $(1, L, L)$  problem.

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

---

<sup>2</sup>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.

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

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

141 **1.3. Basic definitions and notation.** Before proceeding, we introduce some  
 142 basic definitions and notation. Scalars, vectors, matrices and tensors are denoted by  
 143 lowercase, bold lowercase, bold uppercase and calligraphic uppercase letters, respec-  
 144 tively (e.g.,  $x$ ,  $\mathbf{x}$ ,  $\mathbf{X}$ ,  $\mathcal{X}$ ). The symbols  $\otimes$  and  $\boxtimes$  stand for the tensor and Kronecker  
 145 products, respectively. The symbol  $\mathcal{O}$  denotes the null tensor. Vector inequalities  
 146  $\mathbf{x} \leq \mathbf{y}$  are meant entry-wise. Submatrices and subtensors are denoted by MATLAB-  
 147 like notation, as in  $[\mathbf{X}]_{:,1:R}$ , which holds the first  $R$  columns of  $\mathbf{X}$ , and  $[\mathbf{X}]_{i,:}$ , which  
 148 holds its  $i$ th row. The notation  $\mathbf{X}_{\langle n \rangle} = (\mathcal{X})_{\langle n \rangle}$  stands for the mode- $n$  (flat) matrix  
 149 unfolding of  $\mathcal{X}$ , whose columns are subtensors  $[\mathcal{X}]_{i_1, \dots, i_{n-1}, :, i_{n+1}, \dots, i_N}$  sorted in reverse  
 150 lexicographical order with respect to the fixed indices. Given  $\mathcal{X} \in \bigotimes_{n=1}^N \mathbb{C}^{I_n}$  and  
 151  $\mathbf{P} \in \mathbb{C}^{M \times I_n}$ , the mode- $n$  contraction (or product) is defined such that  $(\mathcal{X} \bullet_n \mathbf{P})_{\langle n \rangle} =$   
 152  $\mathbf{P} \mathbf{X}_{\langle n \rangle}$ . For brevity, we employ the shorthands  $\mathbb{N}_N \triangleq \{1, \dots, N\}$ ,  $\bar{I} \triangleq \prod_n I_n$  and  
 153  $\bar{I}_n \triangleq \bar{I}/I_n$ . Finally,  $\mathbf{I}_M$  stands for the  $M \times M$  identity matrix.

154 **1.4. Paper organization.** The rest of this work is organized in the following  
 155 manner. Section 2 provides a brief review of existing non-iterative LMA algorithms,  
 156 their properties and computational complexity. Then, our proposed approach is de-  
 157 scribed and analyzed in section 3. Following that, section 4 introduces the DBTD  
 158 algorithm and investigates its properties. Numerical results of computer simulations  
 159 are presented in section 5, encompassing comparisons of SeLRAP and DBTD with  
 160 competing alternatives for LMA and DBTD-(1, L, L) computation, respectively. Con-  
 161 cluding remarks and perspectives are then drawn in section 6.

## 162 2. State of the art.

163 **2.1. Truncated higher-order singular value decomposition.** Let us de-  
 164 note 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  
 165 orthonormal matrices and define

$$166 \quad (6) \quad \mathcal{P}(I, R) = \{\mathbf{P} \in \mathbb{C}^{I \times I} : \mathbf{P} = \mathbf{U} \mathbf{U}^H, \mathbf{U} \in \mathcal{S}(I, R)\}.$$

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

$$169 \quad (7) \quad \min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I_n, R_n)} \left\| \mathcal{X} - \mathcal{X} \bullet_{n=1}^N \hat{\mathbf{P}}^{(n)} \right\|_F^2.$$

170 Introducing a telescoping sum inside the norm, one obtains

$$171 \quad (8) \quad \min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I, R)} \left\| \mathcal{X} - \sum_{n=1}^N \mathcal{X} \bullet_{m=1}^n \hat{\mathbf{P}}^{(m)} + \sum_{n=1}^{N-1} \mathcal{X} \bullet_{m=1}^n \hat{\mathbf{P}}^{(m)} \right\|_F^2.$$

172 Regrouping the terms, we have [25]

$$173 \quad (9) \quad \min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I, R)} \left\| \sum_{m=1}^N \mathbf{X} \bullet_{m=1}^{n-1} \hat{\mathbf{P}}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2 = \min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I, R)} \sum_{m=1}^N \left\| \mathbf{X} \bullet_{m=1}^{n-1} \hat{\mathbf{P}}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2,$$

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

$$177 \quad (10) \quad \sum_{n=1}^N \left\| \mathbf{X} \bullet_{m=1}^{n-1} \hat{\mathbf{P}}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2 \leq \sum_{n=1}^N \left\| \mathbf{X} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2 = \sum_{n=1}^N \left\| \hat{\mathbf{P}}_{\perp}^{(n)} \mathbf{X}_{\langle n \rangle} \right\|_F^2.$$

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  $R_n$ th column. These matrices are the factors of the HOSVD of  $\mathbf{X}$  [7]. Then, a truncated core is computed as  $\mathbf{S} = \mathbf{X} \bullet_{n=1}^N ([\mathbf{U}^{(n)}]_{:,1:R_n})^H$ , from which  $\hat{\mathbf{X}}$  is obtained via  $\hat{\mathbf{X}} = \mathbf{S} \bullet_{n=1}^N [\mathbf{U}^{(n)}]_{:,1:R_n}$ . Now, because these projectors  $\mathbf{P}^{(n)}$  are optimal when considered separately (but not jointly), any solution  $\mathbf{X}^*$  of (1) satisfies

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

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

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

$$182 \quad (11) \quad C_{\text{THOSVD}} = \sum_{n=1}^N C_{\text{SVD}}(I_n, \bar{I}_n) + \sum_{n=1}^N \mathcal{O}(H_n R_n I_n) + \sum_{n=1}^N \mathcal{O}(J_n R_n I_n),$$

183 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  
 184 third summations correspond to the calculation of  $\mathbf{S}$  and  $\hat{\mathbf{X}}$ , respectively.<sup>3</sup>

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

190 **2.2. Sequentially optimal modal projections.** Another way of computing  
 191 an approximate solution of (9) is by sequentially minimizing the cost function with  
 192 respect to the projectors. This leads to the SeMP solution [25], defined as:

$$193 \quad (12) \quad \text{Given } \mathbf{P}^{(1)}, \dots, \mathbf{P}^{(n-1)}, \text{ compute } \mathbf{P}^{(n)} = \arg \min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I_n, R_n)} \left\| \mathbf{X} \bullet_{m=1}^{n-1} \mathbf{P}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2.$$

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

<sup>3</sup>We assume that the contractions needed to calculate  $\mathbf{S}$  and  $\hat{\mathbf{X}}$  are performed in the order  $n = 1, \dots, N$ . This simplifies the comparison with the other algorithms.

- 197 1. Let  $\mathcal{W}^{(1)} = \mathcal{X}$ .  
198 2. For  $n = 1, \dots, N$ :  
199 (i) compute the SVD of  $\mathcal{W}_{\langle n \rangle}^{(n)}$  to obtain  $\bar{\mathbf{U}}^{(n)} \in \mathbb{C}^{I_n \times R_n}$ , which holds its first  $R_n$   
200 left singular vectors;  
201 (ii) compute  $\mathcal{W}^{(n+1)} = \mathcal{W}^{(n)} \bullet_n \bar{\mathbf{U}}^{(n)H} \in \left(\otimes_{m=1}^n \mathbb{C}^{R_m}\right) \otimes \left(\otimes_{m=n+1}^N \mathbb{C}^{I_m}\right)$ .  
202 3. Finally, construct the solution  $\hat{\mathcal{X}} = \mathcal{W}^{(N+1)} \bullet_{n=1}^N \bar{\mathbf{U}}^{(m)}$ .  
203 It is easy to show that the resulting approximation error is subject to the same  
204 upper bound as the THOSVD. Indeed, we have [25]

$$\begin{aligned}
205 \quad \|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 &= \sum_{n=1}^N \min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I, R)} \left\| \mathcal{X} \bullet_{m=1}^{n-1} \mathbf{P}^{(m)} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2 \\
206 \quad (13) \quad &\leq \sum_{n=1}^N \min_{\hat{\mathbf{P}}^{(n)} \in \mathcal{P}(I, R)} \left\| \mathcal{X} \bullet_n \hat{\mathbf{P}}_{\perp}^{(n)} \right\|_F^2 \leq N \|\mathcal{X} - \mathcal{X}^*\|_F^2, \\
207
\end{aligned}$$

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

$$211 \quad (14) \quad C_{\text{SeMP}} = \sum_{n=1}^N [C_{\text{SVD}}(I_n, H_n) + \mathcal{O}(H_n R_n I_n)] + \sum_{n=1}^N \mathcal{O}(J_n R_n I_n)$$

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

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

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

$$226 \quad (15) \quad \left\| \mathcal{X} - \hat{\mathcal{X}}_{\text{SeMP}} \right\|_F^2 \leq \left\| \mathcal{X} - \hat{\mathcal{X}}_{\text{THOSVD}} \right\|_F^2.$$

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

232 **2.3. Sequential rank-one approximation and projection.** When  $R_1 =$   
233  $\dots = R_N = 1$ , problem (1) reduces to the best rank-one approximation of  $\mathcal{X}$ . In  
234 other words, one seeks an elementary (or decomposable) tensor  $\hat{\mathcal{X}} = \mathbf{v}^{(1)} \otimes \dots \otimes \mathbf{v}^{(N)}$   
235 minimizing the cost function in (1). Note that no distinction exists between tensor  
236 rank and mrank in this case. The SeROAP algorithm [3] computes an approximate  
237 solution by proceeding as follows:

238 1. *Order reduction stage:*

239 (i) Let  $\mathbf{W}^{(1)} \triangleq \mathbf{X}$ .

240 (ii) For  $n = 2, \dots, N$ , recursively calculate the tensor  $\mathbf{W}^{(n)} \in \left(\bigotimes_{m=1}^{n-1} \mathbb{C}^{I_m}\right) \otimes$

241  $\left(\bigotimes_{m=n}^N \mathbb{C}^{I_m}\right)$  whose vectorization  $\mathbf{w}^{(n)} \triangleq \text{vec}(\mathbf{W}^{(n)})$  is a minimizer of

$$242 \quad \min_{\hat{\lambda} \in \mathbb{R}, \hat{\mathbf{u}}^{(n)} \in \mathbb{C}^{I_{n-1}}, \hat{\mathbf{w}}^{(n)} \in \mathbb{C}^{I_N \dots I_n}} \left\| \mathbf{W}_{\langle n-1 \rangle}^{(n-1)} - \hat{\lambda} \hat{\mathbf{u}}^{(n)} \left( \hat{\mathbf{w}}^{(n)} \right)^H \right\|_F^2.$$

243 This can be done by computing the dominant singular triplet of the matrix

$$244 \quad \mathbf{W}_{\langle n-1 \rangle}^{(n-1)} \in \mathbb{C}^{I_{n-1} \times I_N \dots I_n}.$$

245 2. *Projection stage:*

246 (i) Let  $\mathbf{z}^{(N-1)} \triangleq \mathbf{w}^{(N)*} \boxtimes \mathbf{u}^{(N)} \in \mathbb{C}^{I_N I_{N-1}}$ .

247 (ii) For  $n = N-2, \dots, 1$ , project the rows of  $\mathbf{W}_{\langle n \rangle}^{(n)}$  onto  $\text{span}(\mathbf{z}^{(N+1)})$ , i.e.,

$$248 \quad (16) \quad \mathbf{Z}_{\langle n \rangle}^{(n)} = \mathbf{W}_{\langle n \rangle}^{(n)} \left[ \frac{1}{\|\mathbf{z}^{(n+1)}\|_2^2} \left( \mathbf{z}^{(n+1)} \mathbf{z}^{(n+1)H} \right) \right] \in \mathbb{C}^{I_n \times I_N \dots I_{n+1}},$$

249 and then obtain  $\mathbf{z}^{(n)}$  as  $\mathbf{z}^{(n)} = \text{vec}(\mathbf{Z}^{(n)})$ .

250 3. Construct the estimate  $\hat{\mathbf{X}}$  such that  $\text{vec}(\hat{\mathbf{X}}) = \mathbf{z}^{(1)} \in \mathbb{C}^{I_N \dots I_1}$ , or, equivalently, such

251 that  $\hat{\mathbf{X}}_{\langle 1 \rangle} = \mathbf{Z}_{\langle 1 \rangle}^{(1)}$ .

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

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

$$260 \quad (17) \quad \left\| \mathbf{X} - \hat{\mathbf{X}}_{\text{SeROAP}} \right\|_F^2 \leq \left\| \mathbf{X} - \hat{\mathbf{X}}_{\text{SeMP}} \right\|_F^2.$$

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

$$264 \quad C_{\text{SeROAP}} = \sum_{n=1}^{N-1} \left[ \mathcal{O}\left(k \prod_{m=n}^N I_m\right) + \mathcal{O}\left(\prod_{m=n}^N I_m\right) \right].$$

### 265 3. Sequential low-rank approximation and projection.

266 **3.1. Formulation and algorithm.** The same principle underlying SeROAP  
267 can also be employed for computing an LMA of arbitrary mrank  $\mathbf{m} = (R_1, \dots, R_N)$ .  
268 In the projection stage of SeROAP, the rows of each unfolding  $\mathbf{W}_{\langle n \rangle}^{(n)}$  are projected  
269 onto the subspace spanned by a Kronecker-structured vector representing a tensor  
270 product of one-dimensional subspaces. This suggests that a general procedure should  
271 project these rows onto the span of a Kronecker-structured basis representing a tensor  
272 product of low-dimensional subspaces, in consonance with  $\mathbf{m}$ .

273 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  $\mathbf{m} = (R_1, \dots, R_N)$

**Output:** An  $\mathbf{m}$ -mrank approximation of  $\mathcal{X}$ ,  $\hat{\mathcal{X}}$

```

1:  $\mathbf{W}_{\langle 1 \rangle}^{(1)} \leftarrow \mathbf{X}_{\langle 1 \rangle}$ 
2: for  $n = 2, \dots, N + 1$  do
3:    $\left( \mathbf{W}_{\langle n-1 \rangle}^{(n)}, \bar{\mathbf{U}}^{(n-1)} \right) \leftarrow \text{projdomsp} \left( \mathbf{W}_{\langle n-1 \rangle}^{(n-1)}, R_{n-1} \right)$ 
4: end for
5:  $\mathbf{Z}_{\langle N \rangle}^{(N)} \leftarrow \bar{\mathbf{U}}^{(N)} \mathbf{W}_{\langle N \rangle}^{(N+1)}$ 
6: for  $n = N - 1, \dots, 1$  do
7:   compute the “thin” QR decomposition:  $\mathbf{Z}_{\langle n \rangle}^{(n+1)H} = \mathbf{Q}^{(n+1)} \mathbf{R}^{(n+1)}$ 
8:    $\mathbf{Z}_{\langle n \rangle}^{(n)} \leftarrow \left( \mathbf{W}_{\langle n \rangle}^{(n)} \mathbf{Q}^{(n+1)} \right) \mathbf{Q}^{(n+1)H}$ 
9: end for
10: return  $\hat{\mathcal{X}} = \mathcal{Z}^{(1)}$ 

```

---

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

$$281 \quad (18) \quad \mathbf{Z}_{\langle n \rangle}^{(n)} = \mathbf{W}_{\langle n \rangle}^{(n)} \mathbf{Z}_{\langle n \rangle}^{(n+1)H} \left( \mathbf{Z}_{\langle n \rangle}^{(n+1)} \mathbf{Z}_{\langle n \rangle}^{(n+1)H} \right)^{-1} \mathbf{Z}_{\langle n \rangle}^{(n+1)}.$$

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

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

287 An explicit pseudocode for SeLRAP is given in [Algorithm 1](#). For simplicity,  
 288 the modes are processed in the natural order  $(1, \dots, N)$ . If one wishes to follow a  
 289 different order, it suffices to permute the modes of  $\mathcal{X}$  before running the algorithm  
 290 and to invert this permutation afterwards. In the dimension reduction stage, SeLRAP  
 291 employs the `projdomsp` routine given in [Algorithm 2](#). When  $I_{n-1} < H_{n-1}$ , instead of  
 292 computing the SVD of  $\mathbf{W}_{\langle n-1 \rangle}^{(n-1)}$  this routine computes the eigenvalue decomposition  
 293 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  $\bar{\mathbf{U}}^{(n-1)}$ . Although the asymptotic  
 294 complexities are the same due to the computation of  $\mathbf{Y}^{(n-1)}$ , this choice can save  
 295 much time in practice, because of the reduced size of the decomposition problem.  
 296 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  
 297 can obtain  $\mathcal{W}^{(n)}$  by taking the first  $R_{n-1}$  right singular vectors multiplied by their  
 298 corresponding singular values; this is cheaper than calculating  $\bar{\mathbf{U}}^{(n-1)H} \mathbf{W}_{\langle n-1 \rangle}^{(n-1)}$ .

---

**Algorithm 2**  $\text{projdomsp}(\mathbf{W}, R)$ : 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  $\bar{\mathbf{U}}$ , where  $\bar{\mathbf{U}} = \arg \min_{\mathbf{U} \in \mathcal{S}(I, R)} \|\mathbf{W} - \mathbf{U}\mathbf{U}^H \mathbf{W}\|_F$  and  $\hat{\mathbf{W}} = \bar{\mathbf{U}}^H \mathbf{W}$

- 1: **if**  $I < M$  **then**
  - 2:      $\mathbf{Y} \leftarrow \mathbf{W}\mathbf{W}^H$
  - 3:     compute the EVD:  $\mathbf{Y} = [\bar{\mathbf{U}} \quad \tilde{\mathbf{U}}] \mathbf{\Lambda} [\bar{\mathbf{U}} \quad \tilde{\mathbf{U}}]^H$ ,  
       where  $\bar{\mathbf{U}} \in \mathbb{C}^{I \times R}$  and  $\mathbf{\Lambda} = \text{Diag}(\lambda_1, \dots, \lambda_I)$ , with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_I$
  - 4:      $\hat{\mathbf{W}} \leftarrow \bar{\mathbf{U}}^H \mathbf{W}$
  - 5: **else**
  - 6:     compute the SVD:  $\mathbf{W} = [\bar{\mathbf{U}} \quad \tilde{\mathbf{U}}] \begin{bmatrix} \bar{\mathbf{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{\Sigma}} \end{bmatrix} [\bar{\mathbf{V}} \quad \tilde{\mathbf{V}}]^H$ ,  
       where  $\bar{\mathbf{U}} \in \mathbb{C}^{I \times R}$ ,  $\bar{\mathbf{\Sigma}} \in \mathbb{R}^{R \times R}$  and  $\bar{\mathbf{V}} \in \mathbb{C}^{M \times R}$
  - 7:      $\hat{\mathbf{W}} \leftarrow \bar{\mathbf{\Sigma}} \bar{\mathbf{V}}^H$
  - 8: **end if**
  - 9: **return**  $(\hat{\mathbf{W}}, \bar{\mathbf{U}})$
- 

299     *Remark 3.* When  $I_{n-1} < H_{n-1}$ , it is actually more appropriate to first compute  
 300 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  $\mathbf{R} \in \mathbb{C}^{I_{n-1} \times I_{n-1}}$   
 301 via a modified Gram-Schmidt orthogonalization. Then, the SVD of  $\mathbf{R}$  is computed,  
 302 providing the right singular vectors of  $\mathbf{R}$ , which are premultiplied by  $\mathbf{Q}$  to yield those  
 303 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,  $\mathbf{W}_{\langle n-1 \rangle}^{(n-1)}$   
 304 is calculated from the obtained right singular vectors and singular values of  $\mathbf{W}_{\langle n-1 \rangle}^{(n-1)}$ .  
 305 This procedure requires less flops than  $\text{projdomsp}$  (though their order-wise complex-  
 306 ities are the same) and is more accurate. However, [Algorithm 2](#) runs faster in MAT-  
 307 LAB, 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.  
 308 Therefore, we have adopted it in our MATLAB implementation of SeLRAP.<sup>4</sup>

309     **3.2. Computational complexity.** In practice, [Algorithm 1](#) performs (18) with  
 310 the aid of an orthonormal basis for the row space of  $\mathbf{Z}_{\langle n \rangle}^{(n+1)}$ , obtained thanks to a QR  
 311 decomposition. One could also use an SVD. In any case, the constructed projector  
 312 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  
 313 replacing the inverse matrix of (18) by the Moore-Penrose pseudo-inverse. Computing  
 314 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  
 315 dimensions  $R_n \times H_n$ , while performing the projection costs  $\mathcal{O}(H_n R_n I_n)$ . SeLRAP  
 316 thus has the overall complexity

$$317 \quad C_{\text{SeLRAP}} = \sum_{n=1}^N [C_{\text{SVD}}(I_n, H_n) + \mathcal{O}(H_n R_n I_n)] + \mathcal{O}(H_N R_N I_N) + \sum_{n=1}^{N-1} C_{\text{proj}}(H_n, R_n, I_n),$$

---

<sup>4</sup>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
$\sum_{n=1}^N C_{\text{SVD}}(I_n, \bar{I}_n)$	$\sum_{n=1}^N \mathcal{O}(H_n R_n I_n)$	$\sum_{n=1}^N \mathcal{O}(J_n R_n I_n)$
SeMP		
(I) Comput. projectors	(II) Dimension reduction	(III) LMA construction
$\sum_{n=1}^N C_{\text{SVD}}(I_n, H_n)$	$\sum_{n=1}^N \mathcal{O}(H_n R_n I_n)$	$\sum_{n=1}^N \mathcal{O}(J_n R_n I_n)$
SeLRAP		
(I) Comput. projectors	(II) Dimension reduction	(III) Backward projections
$\sum_{n=1}^N C_{\text{SVD}}(I_n, H_n)$	$\sum_{n=1}^N \mathcal{O}(H_n R_n I_n)$	$\mathcal{O}(H_N R_N I_N) + \sum_{n=1}^{N-1} C_{\text{proj}}(H_n, R_n, I_n)$

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

327 The same heuristic described in [subsection 2.2](#), of processing the modes in as-  
328 cending order of dimensions, usually yields a significant economy of computing time  
329 when applying SeLRAP. This economy is all the more relevant when the ratios  $R_n/I_n$   
330 are small and approximately equal.

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

### 335 3.3. Analysis.

336 **3.3.1. Fulfillment of rank constraint.** We now show that the approximation  
337 delivered by SeLRAP actually meets the desired mrank constraint.

338 LEMMA 4. Let  $\mathcal{X} \in \bigotimes_{m=1}^N \mathbb{C}^{K_m}$  and define the tensor  $\mathcal{P} \in \left( \bigotimes_{m=1}^{n-1} \mathbb{C}^{K_m} \right) \otimes \mathbb{C}^{R_n} \otimes$   
339  $\left( \bigotimes_{m=n+1}^N \mathbb{C}^{K_m} \right)$ , for some  $n \in \mathbb{N}_N$ . If  $\text{mrank}(\mathcal{P}) = (R_1, \dots, R_N)$ , then

$$340 \mathbf{Y}_{\langle n \rangle} = \mathbf{X}_{\langle n \rangle} \mathbf{P}_{\langle n \rangle}^H (\mathbf{P}_{\langle n \rangle} \mathbf{P}_{\langle n \rangle}^H)^{-1} \mathbf{P}_{\langle n \rangle}$$

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

342 *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  $\mathbf{U}^{(m)} \in$   
343  $\mathbb{C}^{K_m \times R_m}$  for  $m \in \mathbb{N}_N \setminus \{n\}$  such that  $\mathbf{U}^{(m)}$  has orthonormal columns and

$$344 \mathbf{P}_{\langle n \rangle} = \mathbf{G}_{\langle n \rangle} \underbrace{\left( \mathbf{U}^{(N)} \boxtimes \dots \boxtimes \mathbf{U}^{(n+1)} \boxtimes \mathbf{U}^{(n-1)} \boxtimes \dots \boxtimes \mathbf{U}^{(1)} \right)^T}_{\cong \mathbf{U}^T}.$$

345 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  $\text{rank}(\mathbf{Y}_{\langle n \rangle}) \leq R_n$ .  
 346 Defining  $\mathbf{C} \in \bigotimes_{m=1}^N \mathbb{C}^{R_m}$  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  
 347 that  $\mathbf{y} = \mathbf{C} \bullet_{m \neq n} \mathbf{U}^{(m)}$ , implying  $\text{rank}(\mathbf{Y}_{\langle m \rangle}) \leq R_m$  for all  $m \in \mathbb{N}_N \setminus \{n\}$ .  $\square$

348 **3.3.2. Comparison with SeMP.** In the following, we analytically compare  
 349 the quadratic errors incurred by SeMP and SeLRAP for third-order tensors. Let  
 350  $\mathbf{m} = (R_1, R_2, R_3)$  and denote by  $\hat{\mathbf{X}}_{\text{SeMP}} = \mathbf{S} \bullet_{n=1}^3 \bar{\mathbf{U}}^{(n)}$  the approximation delivered  
 351 by SeMP, where  $\bar{\mathbf{U}}^{(n)}$  is as defined in subsection 2.2. Because  $\mathbf{S} = \mathbf{X} \bullet_{n=1}^3 \bar{\mathbf{U}}^{(n)H}$ , the  
 352 resulting quadratic error can be written as

$$353 \quad \varepsilon_{\text{SeMP}} \triangleq \|\mathbf{X} - \hat{\mathbf{X}}_{\text{SeMP}}\|_F^2 = \|\mathbf{X}\|_F^2 - \left\| \mathbf{X} \bullet_{n=1}^3 \bar{\mathbf{U}}^{(n)H} \right\|_F^2$$

$$354 \quad = \|\mathbf{X}\|_F^2 - \left\| \bar{\mathbf{U}}^{(1)H} \mathbf{X}_{\langle 1 \rangle} \left( \bar{\mathbf{U}}^{(3)*} \boxtimes \bar{\mathbf{U}}^{(2)*} \right) \right\|_F^2.$$

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

$$357 \quad (19) \quad \varepsilon_{\text{SeMP}} = \|\mathbf{X}\|_F^2 - \left\| \bar{\boldsymbol{\Sigma}}^{(1)} \bar{\mathbf{V}}^{(1)H} \left( \bar{\mathbf{U}}^{(3)*} \boxtimes \bar{\mathbf{U}}^{(2)*} \right) \right\|_F^2,$$

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

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

$$363 \quad (20) \quad \mathbf{P} \triangleq \mathbf{Z}_{\langle 1 \rangle}^{(2)H} \left( \mathbf{Z}_{\langle 1 \rangle}^{(2)} \mathbf{Z}_{\langle 1 \rangle}^{(2)H} \right)^{-1} \mathbf{Z}_{\langle 1 \rangle}^{(2)}.$$

364 Using this definition along with (18) and the identities  $(\hat{\mathbf{X}}_{\text{SeLRAP}})_{\langle 1 \rangle} = \mathbf{Z}_{\langle 1 \rangle}^{(1)}$  and  
 365  $\mathbf{X}_{\langle 1 \rangle} = \mathbf{W}_{\langle 1 \rangle}^{(1)}$ , we derive

$$366 \quad (21) \quad \varepsilon_{\text{SeLRAP}} \triangleq \|\mathbf{X} - \hat{\mathbf{X}}_{\text{SeLRAP}}\|_F^2 = \|\mathbf{X}\|_F^2 - \left\| \mathbf{W}_{\langle 1 \rangle}^{(1)} \mathbf{P} \right\|_F^2.$$

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

$$369 \quad \left\| \mathbf{W}_{\langle 1 \rangle}^{(1)} \mathbf{P} \right\|_F^2 = \text{Tr} \left\{ \mathbf{P} \mathbf{W}_{\langle 1 \rangle}^{(1)H} \mathbf{W}_{\langle 1 \rangle}^{(1)} \mathbf{P} \right\} = \left\| \bar{\boldsymbol{\Sigma}}^{(1)} \bar{\mathbf{V}}^{(1)H} \mathbf{P} \right\|_F^2 + \left\| \mathbf{E}^{(1)} \mathbf{P} \right\|_F^2.$$

371 Plugging the result into (21), we have

$$372 \quad \varepsilon_{\text{SeLRAP}} = \|\mathbf{X}\|_F^2 - \left\| \bar{\boldsymbol{\Sigma}}^{(1)} \bar{\mathbf{V}}^{(1)H} \mathbf{P} \right\|_F^2 - \left\| \mathbf{E}^{(1)} \mathbf{P} \right\|_F^2$$

$$373 \quad \leq \|\mathbf{X}\|_F^2 - \left\| \bar{\boldsymbol{\Sigma}}^{(1)} \bar{\mathbf{V}}^{(1)H} \mathbf{P} \right\|_F^2.$$

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

$$376 \quad (22) \quad \left\| \overline{\Sigma}^{(1)} \overline{\mathbf{V}}^{(1)H} \mathbf{P} \right\|_F^2 \geq \left\| \overline{\Sigma}^{(1)} \overline{\mathbf{V}}^{(1)H} \left( \overline{\mathbf{U}}^{(3)*} \boxtimes \overline{\mathbf{U}}^{(2)*} \right) \right\|_F^2.$$

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

381 **THEOREM 5.** *Let  $\mathbf{X} \in \bigotimes_{n=1}^3 \mathbb{C}^{I_n}$  and denote by  $\hat{\mathbf{X}}_{\text{SeLRAP}}$  and  $\hat{\mathbf{X}}_{\text{SeMP}}$  the rank-*  
 382  *$(1, L, L)$  approximations of  $\mathbf{X}$  produced by SeLRAP and SeMP, respectively, by pro-*  
 383 *cessing the modes in the natural order  $(1, 2, 3)$ . Then,*

$$384 \quad (23) \quad \left\| \mathbf{X} - \hat{\mathbf{X}}_{\text{SeLRAP}} \right\|_F^2 \leq \left\| \mathbf{X} - \hat{\mathbf{X}}_{\text{SeMP}} \right\|_F^2.$$

385 *Proof.* First, SeLRAP computes the SVDs

$$386 \quad \mathbf{W}_{(1)}^{(1)} = \overline{\sigma}^{(1)} \overline{\mathbf{u}}^{(1)} \overline{\mathbf{v}}^{(1)H} + \mathbf{E}^{(1)} \in \mathbb{C}^{I_1 \times I_3 I_2},$$

$$387 \quad \mathbf{W}_{(2)}^{(2)} = \overline{\mathbf{U}}^{(2)} \overline{\Sigma}^{(2)} \overline{\mathbf{V}}^{(2)H} + \mathbf{E}^{(2)} \in \mathbb{C}^{I_2 \times I_3},$$

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

$$390 \quad (24) \quad \text{vec} \left( \mathbf{W}_{(1)}^{(2)} \right) = \overline{\sigma}^{(1)} \overline{\mathbf{v}}^{(1)*} = \text{vec} \left( \mathbf{W}_{(2)}^{(2)} \right) = \text{vec} \left( \overline{\mathbf{U}}^{(2)} \overline{\Sigma}^{(2)} \overline{\mathbf{V}}^{(2)H} + \mathbf{E}^{(2)} \right)$$

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

$$394 \quad (25) \quad \mathbf{Z}_{(3)}^{(3)} = \mathbf{W}_{(3)}^{(3)} = \overline{\mathbf{U}}^{(3)} \overline{\Sigma}^{(3)} \in \mathbb{C}^{I_3 \times R},$$

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

$$397 \quad (26) \quad \mathbf{Z}_{(2)}^{(2)} = \mathbf{W}_{(2)}^{(2)} \overline{\mathbf{V}}^{(2)} \overline{\mathbf{V}}^{(2)H} = \overline{\mathbf{U}}^{(2)} \overline{\Sigma}^{(2)} \overline{\mathbf{V}}^{(2)H} \in \mathbb{C}^{I_2 \times I_3}.$$

399 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

$$400 \quad \mathbf{Z}_{(1)}^{(2)} = \text{vec} \left( \overline{\Sigma}^{(2)} \right)^T \left( \overline{\mathbf{U}}^{(3)} \boxtimes \overline{\mathbf{U}}^{(2)} \right)^T \in \mathbb{C}^{1 \times I_3 I_2},$$

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

$$403 \quad (27) \quad \overline{\sigma}^{(1)} \overline{\mathbf{v}}^{(1)*} = \mathbf{U} \overline{\sigma}^{(2)} + \text{vec} \left( \mathbf{E}^{(2)} \right).$$

404 In view of the derived expressions, computing the left-hand side of (22) for  $\mathbf{m} =$   
 405  $(1, L, L)$  yields

$$406 \quad (28) \quad \left\| \bar{\sigma}^{(1)} \bar{\mathbf{v}}^{(1)H} \mathbf{P} \right\|_2^2 = \left( \bar{\sigma}^{(1)} \right)^2 \bar{\mathbf{v}}^{(1)H} \mathbf{P} \bar{\mathbf{v}}^{(1)} = \left( \bar{\sigma}^{(1)} \right)^2 \|\bar{\sigma}^{(2)}\|_2^{-2} \left| \bar{\sigma}^{(2)T} \mathbf{U}^T \bar{\mathbf{v}}^{(1)} \right|^2.$$

Due to (27),  $\mathbf{U}^T \bar{\mathbf{v}}^{(1)} = \left( \bar{\sigma}^{(1)} \right)^{-1} \left[ \bar{\sigma}^{(2)} + \mathbf{U}^T \text{vec}(\mathbf{E}^{(2)})^* \right]$ . 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  $\mathbf{U}^T \text{vec}(\mathbf{E}^{(2)})^* = \mathbf{0}$ , leading to  $\mathbf{U}^T \bar{\mathbf{v}}^{(1)} = \left( \bar{\sigma}^{(1)} \right)^{-1} \bar{\sigma}^{(2)}$ . Substituting this expression into (28) 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 (22) is given by

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

407 Therefore, (22) holds with equality, implying (23).  $\square$

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

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

$$414 \quad \left\| \mathbf{X} - \hat{\mathbf{X}}_{\text{SeLRAP}} \right\|_F^2 \leq \left\| \mathbf{X} - \hat{\mathbf{X}}_{\text{SeMP}} \right\|_F^2 \leq \left\| \mathbf{X} - \hat{\mathbf{X}}_{\text{THOSVD}} \right\|_F^2.$$

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

419 **COROLLARY 7.** *Let  $\mathbf{X} \in \bigotimes_{n=1}^3 \mathbb{C}^{I_n}$ . For any solution  $\mathbf{X}^*$  of (1) with  $\mathbf{m} =$*   
 420  *$(1, L, L)$ , the rank- $(1, L, L)$  approximation of  $\mathbf{X}$  produced by SeLRAP satisfies*

$$421 \quad \left\| \mathbf{X} - \hat{\mathbf{X}}_{\text{SeLRAP}} \right\|_F^2 \leq N \|\mathbf{X} - \mathbf{X}^*\|_F^2.$$

422 **3.3.3. The general case  $\mathbf{m} = (R_1, R_2, R_3)$ .** For arbitrary  $\mathbf{m}$ ,  $\mathbf{W}_{(2)}^{(2)}$  is the mode-  
 423 2 unfolding of an  $R_1 \times I_2 \times I_3$  tensor whose best rank- $(R_1, R_2, R_3)$  approximation is  
 424 sought. Therefore, unlike the previous case, an explicit expression for this approxi-  
 425 mation is not available, and a proof of superiority of SeLRAP is harder to undertake.  
 426 We sketch below one possible way of writing the resulting projector  $\mathbf{P}$  in this case, in  
 427 order to give an idea of the increased complexity. Consider the matrix

$$428 \quad (29) \quad \mathbf{Z}_{(3)}^{(3)} = \bar{\mathbf{U}}^{(3)} \bar{\Sigma}^{(3)} \bar{\mathbf{V}}^{(3)H} \in \mathbb{C}^{I_3 \times R_2 R_1}$$

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

$$431 \quad (30) \quad \mathbf{Z}_{(2)}^{(3)} = \mathbf{G}_{(2)}^{(3)} \left( \bar{\mathbf{U}}^{(3)} \boxtimes \mathbf{I}_{R_1} \right)^T \in \mathbb{C}^{R_2 \times I_3 R_1}.$$

432 Now,  
433

$$434 \quad (31) \quad \mathbf{Z}_{(2)}^{(3)H} \left( \mathbf{Z}_{(2)}^{(3)} \mathbf{Z}_{(2)}^{(3)H} \right)^{-1} \mathbf{Z}_{(2)}^{(3)} =$$

$$435 \quad \left( \bar{\mathbf{U}}^{(3)} \boxtimes \mathbf{I}_{R_1} \right)^* \mathbf{G}_{(2)}^{(3)H} \left( \mathbf{G}_{(2)}^{(3)} \mathbf{G}_{(2)}^{(3)H} \right)^{-1} \mathbf{G}_{(2)}^{(3)} \left( \bar{\mathbf{U}}^{(3)} \boxtimes \mathbf{I}_{R_1} \right)^T \in \mathbb{C}^{I_3 R_1 \times I_3 R_1}$$

436  
437 Next, we define

$$438 \quad (32) \quad \tilde{\mathbf{G}}_{(2)}^{(2)} \triangleq \mathbf{W}_{(2)}^{(2)} \left( \bar{\mathbf{U}}^{(3)} \boxtimes \mathbf{I}_{R_1} \right)^* \mathbf{G}_{(2)}^{(3)H} \left( \mathbf{G}_{(2)}^{(3)} \mathbf{G}_{(2)}^{(3)H} \right)^{-1} \mathbf{G}_{(2)}^{(3)} \in \mathbb{C}^{I_2 \times R_3 R_1}.$$

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

$$443 \quad \mathbf{Z}_{(2)}^{(2)} = \mathbf{W}_{(2)}^{(2)} \mathbf{Z}_{(2)}^{(3)H} \left( \mathbf{Z}_{(2)}^{(3)} \mathbf{Z}_{(2)}^{(3)H} \right)^{-1} \mathbf{Z}_{(2)}^{(3)} = \tilde{\mathbf{U}}^{(2)} \mathbf{G}_{(2)}^{(2)} \left( \bar{\mathbf{U}}^{(3)} \boxtimes \mathbf{I}_{R_1} \right)^T \in \mathbb{C}^{I_2 \times I_3 R_1},$$

444 leading to  $\mathbf{Z}_{(1)}^{(2)} = \mathbf{G}_{(1)}^{(2)} \left( \bar{\mathbf{U}}^{(3)} \boxtimes \tilde{\mathbf{U}}^{(2)} \right)^T \in \mathbb{C}^{R_1 \times I_3 I_2}$ . The projector  $\mathbf{P}$  thus reads

$$445 \quad \mathbf{P} = \left( \bar{\mathbf{U}}^{(3)} \boxtimes \tilde{\mathbf{U}}^{(2)} \right)^* \mathbf{G}_{(1)}^{(2)H} \left( \mathbf{G}_{(1)}^{(2)} \mathbf{G}_{(1)}^{(2)H} \right)^{-1} \mathbf{G}_{(1)}^{(2)} \left( \bar{\mathbf{U}}^{(3)} \boxtimes \tilde{\mathbf{U}}^{(2)} \right)^T \in \mathbb{C}^{I_3 I_2 \times I_3 I_2}.$$

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

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

451 **4.1. Algorithm.** The proposed deflationary block-term decomposition (DBTD)  
452 algorithm is described by [Algorithm 3](#). We have employed the symbol  $\mathcal{P}_{\mathbf{m}}$  to denote  
453 an approximate projection onto  $\mathcal{L}_{\mathbf{m}} \triangleq \{\mathcal{Y} : \text{mrank}(\mathcal{Y}) \leq \mathbf{m}\}$  which can be computed,  
454 e.g., by SeLRAP, SeMP or THOSVD. The resulting DBTD algorithm has a very  
455 simple form: at each iteration, one sequentially obtains a new estimate of each block,  
456  $\hat{\mathbf{X}}_{r,k}$ , by computing an mrank- $\mathbf{m}^{(r)}$  approximation of the current residue tensor  $\mathcal{E}_{r-1,k}$   
457 plus the current block estimate  $\hat{\mathbf{X}}_{r,k-1}$ . The new residue tensor  $\mathcal{E}_{r,k}$  then corresponds  
458 to the resulting approximation error. Because  $\hat{\mathbf{X}}_{r,0} = \mathbf{O}$  for all  $r$  and the initial residue  
459 tensor equals  $\mathbf{X}$ , the first iteration amounts to extracting the  $R$  blocks from  $\mathbf{X}$  in a  
460 greedy fashion. In general, this first iteration does not provide the sought blocks.  
461 Thus, DBTD refines these estimates from the second iteration onwards.

462 An appropriate stopping criterion for [Algorithm 3](#) consists in checking whether  
463  $\|\mathcal{E}_{R,k}\|_F$  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  
464 avoid a premature stop, one can verify whether the average  $\psi_k$  among the  $K_s$  most  
465 recent iterations approximately equals 1, which yields the stopping criteria

$$466 \quad (33) \quad \|\mathcal{E}_{R,k}\|_F \leq \epsilon_1 \quad \text{or} \quad \left| 1 - \frac{1}{K_s} \sum_{l=0}^{K_s-1} \psi_{k-l} \right| \leq \epsilon_2,$$

---

**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 m-ranks  $\mathbf{m}^{(r)} = (R_1^{(r)}, \dots, R_N^{(r)})$

**Output:** Estimates of the blocks  $\hat{\mathcal{X}}_r \in \mathcal{L}_{\mathbf{m}^{(r)}}$ ,  $r = 1, \dots, R$

```

1:  $\hat{\mathcal{X}}_{r,0} \leftarrow \mathbf{O}$ , for  $r = 1, \dots, R$ 
2:  $\mathcal{E}_{0,1} \leftarrow \mathcal{X}$ 
3:  $k \leftarrow 0$ 
4: while  $k < K$  and (33) is not satisfied do
5:    $k \leftarrow k + 1$ 
6:   for  $r = 1, \dots, R$  do
7:      $\mathcal{E}_{r,k} \leftarrow \mathcal{E}_{r-1,k} + \hat{\mathcal{X}}_{r,k-1}$ 
8:      $\hat{\mathcal{X}}_{r,k} \leftarrow \mathcal{P}_{\mathbf{m}^{(r)}}(\mathcal{E}_{r,k})$ 
9:      $\mathcal{E}_{r,k} \leftarrow \mathcal{E}_{r,k} - \hat{\mathcal{X}}_{r,k}$ 
10:  end for
11:   $\mathcal{E}_{0,k+1} \leftarrow \mathcal{E}_{R,k}$ 
12: end while
13:  $\hat{\mathcal{X}}_r \leftarrow \hat{\mathcal{X}}_{r,k}$ , for  $r = 1, \dots, R$ 

```

---

467 for two sufficiently small constants  $\epsilon_1, \epsilon_2$ . (In practice, one usually imposes also a  
468 maximum number of iterations,  $K$ ).

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

$$471 \quad (34) \quad C_{\text{DBTD}} = \mathcal{O}(\bar{I}) + \sum_{r=1}^R C_{\mathcal{P}_{\mathbf{m}^{(r)}}},$$

472 where  $C_{\mathcal{P}_{\mathbf{m}^{(r)}}}$  denotes the cost of the LMA method. In particular, if  $\mathbf{m}^{(r)} = \mathbf{m} =$   
473  $(1, L, L)$  and SeLRAP is used, we have

$$474 \quad C_{\mathcal{P}_{\mathbf{m}}} = C_{\text{SVD}}(I_1, I_2 I_3) + \mathcal{O}(I_1 I_2 I_3) + C_{\text{SVD}}(I_2, I_3) + \mathcal{O}(L I_2 I_3) + C_{\text{proj}}(I_2 I_3, 1, I_1),$$

475 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  
476 taken into account the simplifications which apply to SeLRAP when  $\mathbf{m} = (1, L, L)$ ,  
477 as described in [subsection 3.3.2](#).

478 **4.2. Discussion on convergence.** The partial convergence analysis presented  
479 in [2] can be straightforwardly extended to our present case. In the following, we  
480 briefly discuss the main implication of such an extension and the assumption which  
481 underlies it. Namely, if we assume that the best rank- $\mathbf{m}^{(r)}$  approximation is achieved  
482 by  $\mathcal{P}_{\mathbf{m}^{(r)}}$  when updating the  $r$ th block, then the following result holds.

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

485 *Proof.* Let  $\mathcal{P}_{\mathcal{X}_{r,k-1}}$  denote the orthogonal projection onto the modal subspaces of  
486  $\mathcal{X}_{r,k-1}$ , 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  
487  $\mathcal{Y}$ , from the optimality of  $\mathcal{P}_{\mathbf{m}^{(r)}}$  we have  
488

$$489 \quad (35) \quad \begin{aligned} & \|(\mathcal{E}_{r-1,k} + \mathcal{X}_{r,k-1}) - \mathcal{P}_{\mathbf{m}^{(r)}}(\mathcal{E}_{r-1,k} + \mathcal{X}_{r,k-1})\|_F \\ 490 & \leq \|(\mathcal{E}_{r-1,k} + \mathcal{X}_{r,k-1}) - \mathcal{P}_{\mathcal{X}_{r,k-1}}(\mathcal{E}_{r-1,k} + \mathcal{X}_{r,k-1})\|_F. \end{aligned}$$

492 But, by definition of  $\mathcal{P}_{\mathcal{X}_{r,k-1}}$ , the right-hand side cannot be larger than  $\|\mathcal{E}_{r-1,k}\|_F$ .  
493 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  $\Delta$  for third-order tensors*

Scenario			$\Delta$ (THOSVD)		$\Delta$ (SeMP)	
#	$\mathbf{i}$	$\mathbf{r}$	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

494 above inequality is precisely the norm of the residue  $\mathbf{E}_{r,k}$ . Therefore, at the end of  
495 the  $k$ th iteration, one has  $\|\mathbf{E}_{R,k}\|_F \leq \|\mathbf{E}_{R-1,k}\|_F \leq \dots \leq \|\mathbf{E}_{0,k}\|_F = \|\mathbf{E}_{R,k-1}\|_F$ .  $\square$

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

507 **5. Numerical results.** In the following experiments, the modes are always processed  
508 in the natural order in SeLRAP and SeMP, for simplicity. The reported computing  
509 times were measured in MATLAB R2013a running on a Intel Xeon ES-2630v2  
510 2.60 GHz with 32 GB RAM 1866 MHz. For conciseness, the notation  $\mathbf{i} = (I_1, \dots, I_N)$   
511 specifies the tensor dimensions in each scenario.

512 **5.1. Non-iterative low-mrank approximation.** First, we compare the performance  
513 of SeLRAP with respect to those of THOSVD and SeMP in the task of  
514 LMA. To do so, similarly to [3] we measure

$$515 \quad \Delta(\text{Alg}) = 1 - \frac{\|\mathbf{X} - \hat{\mathbf{X}}_{\text{SeLRAP}}\|_F}{\|\mathbf{X} - \hat{\mathbf{X}}_{\text{Alg}}\|_F},$$

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

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

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

545 **5.2. Decomposition in rank-(1, L, L) terms.** In this section, we compare  
546 DBTD with existing BTD-(1, L, L) computation algorithms. Three variants of DBTD  
547 are considered, depending on the choice of  $\mathcal{P}_m$ : DBTD-SeLRAP, DBTD-SeMP and  
548 DBTD-THOSVD. The other included algorithms are ALS-ELS [18] and the Gauss-  
549 Newton algorithm with dogleg trust region (GN-DL) of Tensorlab [26]. Among the  
550 algorithms provided by Tensorlab, the latter has been chosen for its superior perfor-  
551 mance in our simulations, consonantly with the findings of [23].

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

$$555 \quad (36) \quad \bar{\mathbf{x}}_r = \|\mathbf{x}_r\|_F^{-1} \mathbf{x}_r, \quad \mathbf{x}_r = \mathbf{a}_r \otimes (\mathbf{B}_r \mathbf{C}_r^T),$$

556 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  $[\mathbf{a}_1 \ \dots \ \mathbf{a}_R] =$   
557  $\mathbf{Q}_a \mathbf{J}$ ,  $[\text{vec}(\mathbf{B}_1) \ \dots \ \text{vec}(\mathbf{B}_R)] = \mathbf{Q}_b \mathbf{J}$  and  $[\text{vec}(\mathbf{C}_1) \ \dots \ \text{vec}(\mathbf{C}_R)] = \mathbf{Q}_c \mathbf{J}$ , with  
558  $\mathbf{Q}_a \in \mathcal{S}(I_1, R)$ ,  $\mathbf{Q}_b \in \mathcal{S}(LI_2, R)$  and  $\mathbf{Q}_c \in \mathcal{S}(LI_3, R)$  denoting random column-wise  
559 orthonormal matrices and

$$560 \quad (37) \quad \mathbf{J}^T \mathbf{J} = \begin{bmatrix} 1 & \rho^{1/3} & \dots & \rho^{1/3} \\ \rho^{1/3} & 1 & \dots & \rho^{1/3} \\ \vdots & & \ddots & \vdots \\ \rho^{1/3} & \rho^{1/3} & \dots & 1 \end{bmatrix}.$$

561 Since  $\langle \mathbf{x}_r, \mathbf{x}_s \rangle = \langle \mathbf{a}_r, \mathbf{a}_s \rangle \langle \mathbf{B}_r \mathbf{C}_r^T, \mathbf{B}_s \mathbf{C}_s^T \rangle = \langle \mathbf{a}_r, \mathbf{a}_s \rangle \text{Tr} \{ \mathbf{C}_r^T \mathbf{C}_s^* \mathbf{B}_s^H \mathbf{B}_r \}$ , for  $r = s$  we  
562 have  $\|\mathbf{x}_r\|_F^2 = \text{Tr} \{ \mathbf{C}_r^T \mathbf{C}_r^* \mathbf{B}_r^H \mathbf{B}_r \}$  and, for  $r \neq s$ ,

$$563 \quad |\langle \bar{\mathbf{x}}_r, \bar{\mathbf{x}}_s \rangle| = \rho^{1/3} (\text{Tr} \{ \mathbf{C}_r^T \mathbf{C}_r^* \mathbf{B}_r^H \mathbf{B}_r \} \text{Tr} \{ \mathbf{C}_s^T \mathbf{C}_s^* \mathbf{B}_s^H \mathbf{B}_s \})^{-\frac{1}{2}} |\text{Tr} \{ \mathbf{C}_r^T \mathbf{C}_s^* \mathbf{B}_s^H \mathbf{B}_r \}|.$$

564 If  $\mathbf{Q}_a$ ,  $\mathbf{Q}_b$  and  $\mathbf{Q}_c$  are generated by orthogonalizing random matrices having standard  
565 circularly symmetric Gaussian (SCSG) entries, then  $\langle \bar{\mathbf{x}}_r, \bar{\mathbf{x}}_s \rangle$  is normally distributed

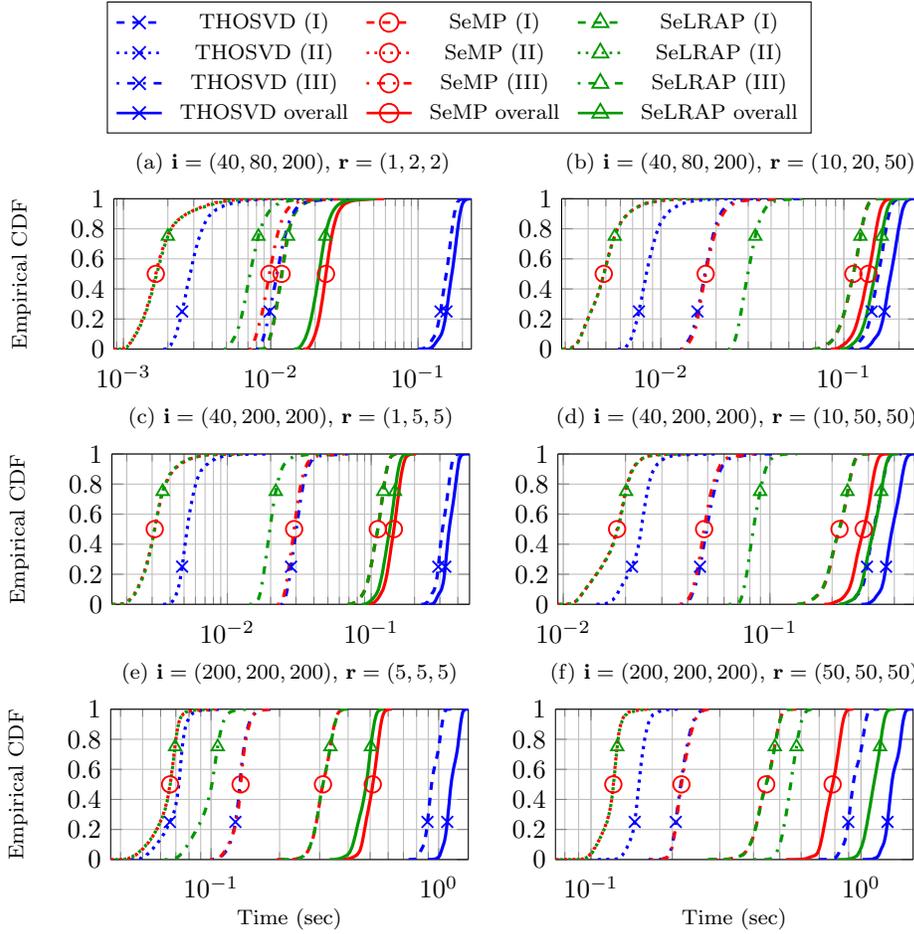


FIGURE 1. Empirical CDFs of the times spent by each LMA algorithm at each stage.

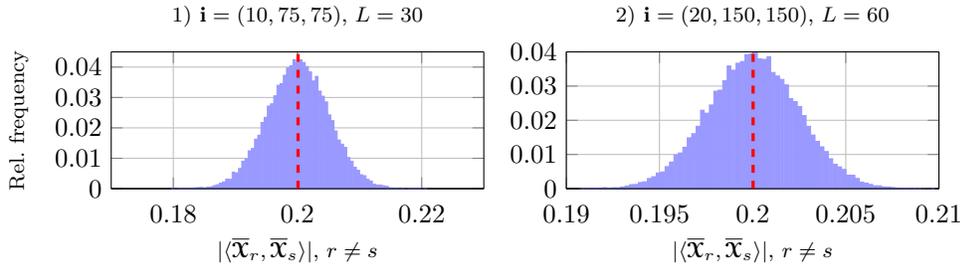


FIGURE 2. Histogram of measured correlation among blocks generated with  $\rho = 0.2$ . The red dashed line indicates the sample mean.

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

569 We consider tensors of dimensions  $\mathbf{i} = (20, 150, 150)$  composed by three rank-  
 570  $(1, 60, 60)$  blocks. For this choice of  $L$ , the uniqueness theorems of [4] do not hold. In

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

$$574 \quad (38) \quad \mathbf{X} = \sum_{r=1}^3 \lambda_r \bar{\mathbf{X}}_r + \sigma \mathbf{N},$$

575 where  $\mathbf{N}$  has SCSG entries and  $\sigma$  is adjusted to impose a desired signal-to-noise ratio

$$576 \quad (39) \quad \text{SNR} = \left\| \sum_{r=1}^3 \lambda_r \bar{\mathbf{X}}_r \right\|_F^2 \sigma^{-2} \|\mathbf{N}\|_F^{-2}.$$

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

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

$$586 \quad (40) \quad \text{ANSE} = \frac{1}{3} \sum_{r=1}^3 \lambda_r^{-2} \|\lambda_r \bar{\mathbf{X}}_r - \hat{\mathbf{X}}_r\|_F,$$

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

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

607 In Figure 5, we fix  $\rho = 0.2$  and vary  $L$  and SNR separately, with  $\lambda_r \sim \mathcal{N}(1, 0.1)$ .  
 608 With  $L = 30$  (and SNR = 50 dB), it becomes possible to algebraically compute an  
 609 approximate solution via a GEVD, because  $RL = 90 < \min\{I_2, I_3\} = 150$  [4, Theorem  
 610 4.1]. We therefore initialize all algorithms with this solution. As seen in Figure 5,  
 611 they are all able to satisfactorily refine it withing a few iterations, with ALS-ELS and  
 612 GN-DL achieving the best overall performance due to their low computing times. For  
 613  $L = 90$  and same SNR, the results of ALS-ELS and GN-DL are considerably degraded  
 614 with respect to the  $L = 60$  case. By contrast, DBTD is still able to accurately  
 615 estimate the factors in the vast majority of realizations, though it typically needs

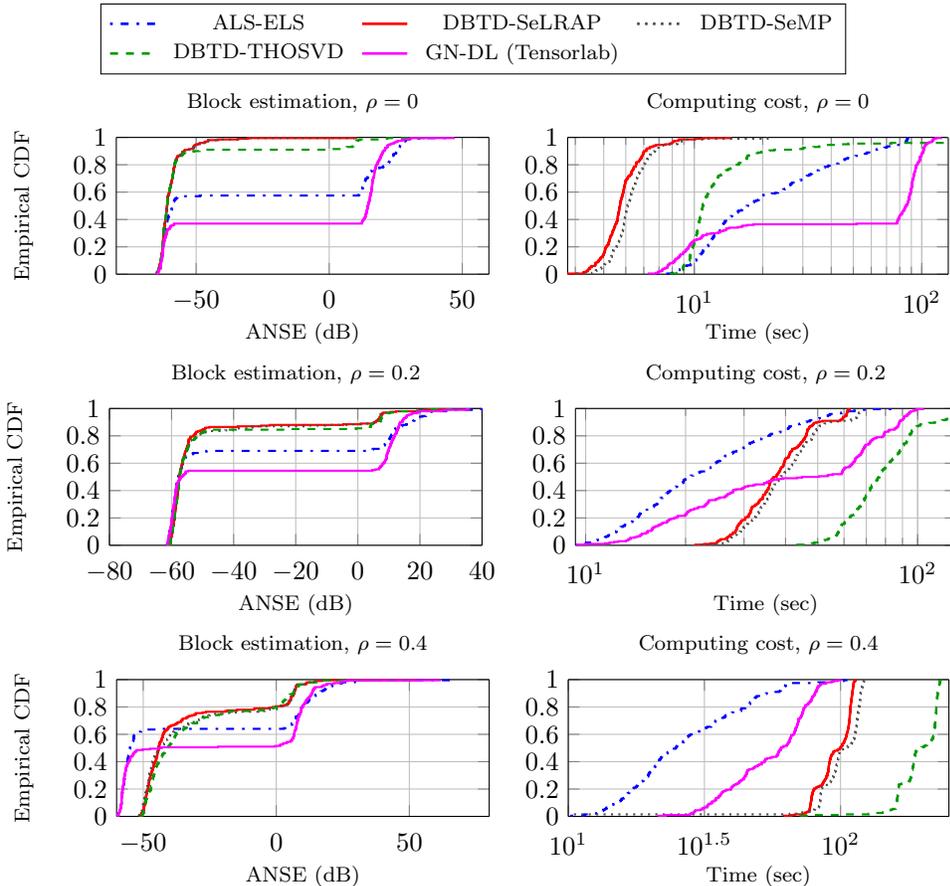


FIGURE 3. Performance of  $BTD-(1, L, L)$  algorithms for  $\mathbf{i} = (20, 150, 150)$ ,  $R = 3$ ,  $L = 60$ ,  $SNR = 50$  dB,  $\rho \in \{0, 0.2, 0.4\}$  and  $\lambda_r \sim \mathcal{N}(1, 0.2)$ .

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

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

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

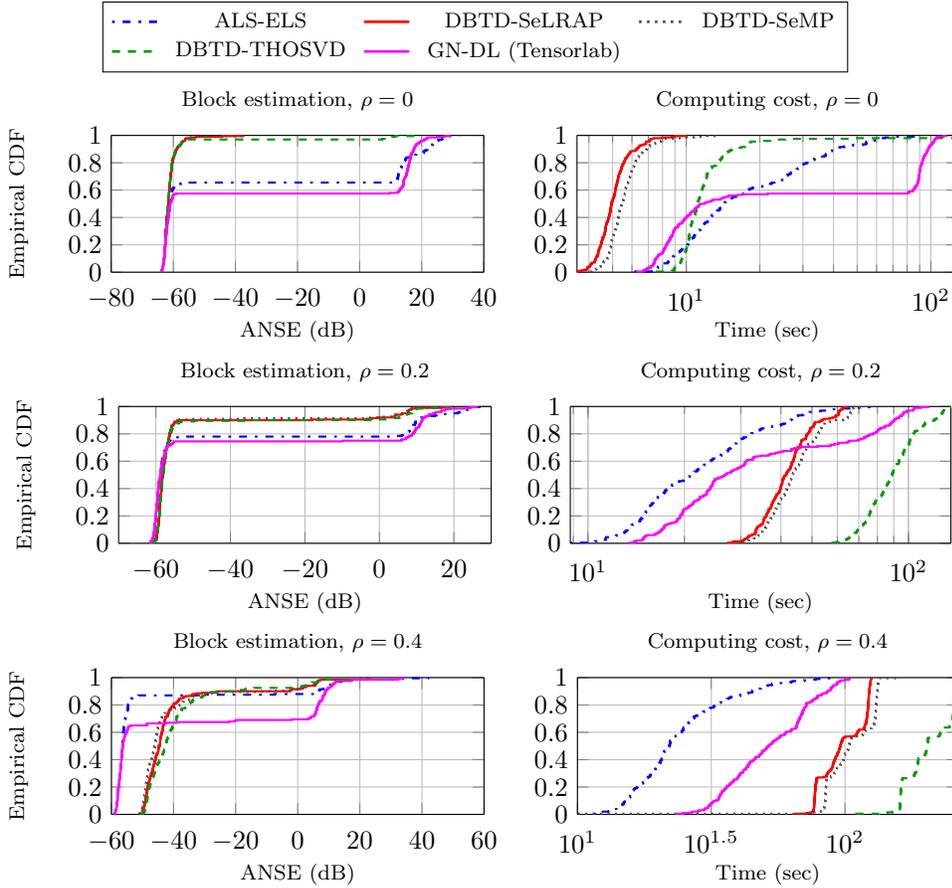


FIGURE 4. Performance of  $BTD-(1, L, L)$  algorithms for  $\mathbf{i} = (20, 150, 150)$ ,  $R = 3$ ,  $L = 60$ ,  $SNR = 50$  dB,  $\rho \in \{0.2, 0.4, 0.6\}$  and  $\lambda_r \sim \mathcal{N}(1, 0.1)$ .

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

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

643

#### REFERENCES

- 644 [1] J. CULLUM AND R. WILLOUGHBY, *Lanczos Algorithms*, vol. I, Birkhauser, 1985.  
 645 [2] A. P. DA SILVA, P. COMON, AND A. L. F. DE ALMEIDA, *Rank-1 tensor approximation methods*  
 646 *and application to deflation*, arXiv preprint arXiv:1508.05273, (2015).  
 647 [3] A. P. DA SILVA, P. COMON, AND A. L. F. DE ALMEIDA, *A finite algorithm to compute rank-1*  
 648 *tensor approximations*, IEEE Signal Processing Letters, 23 (2016), pp. 959–963.

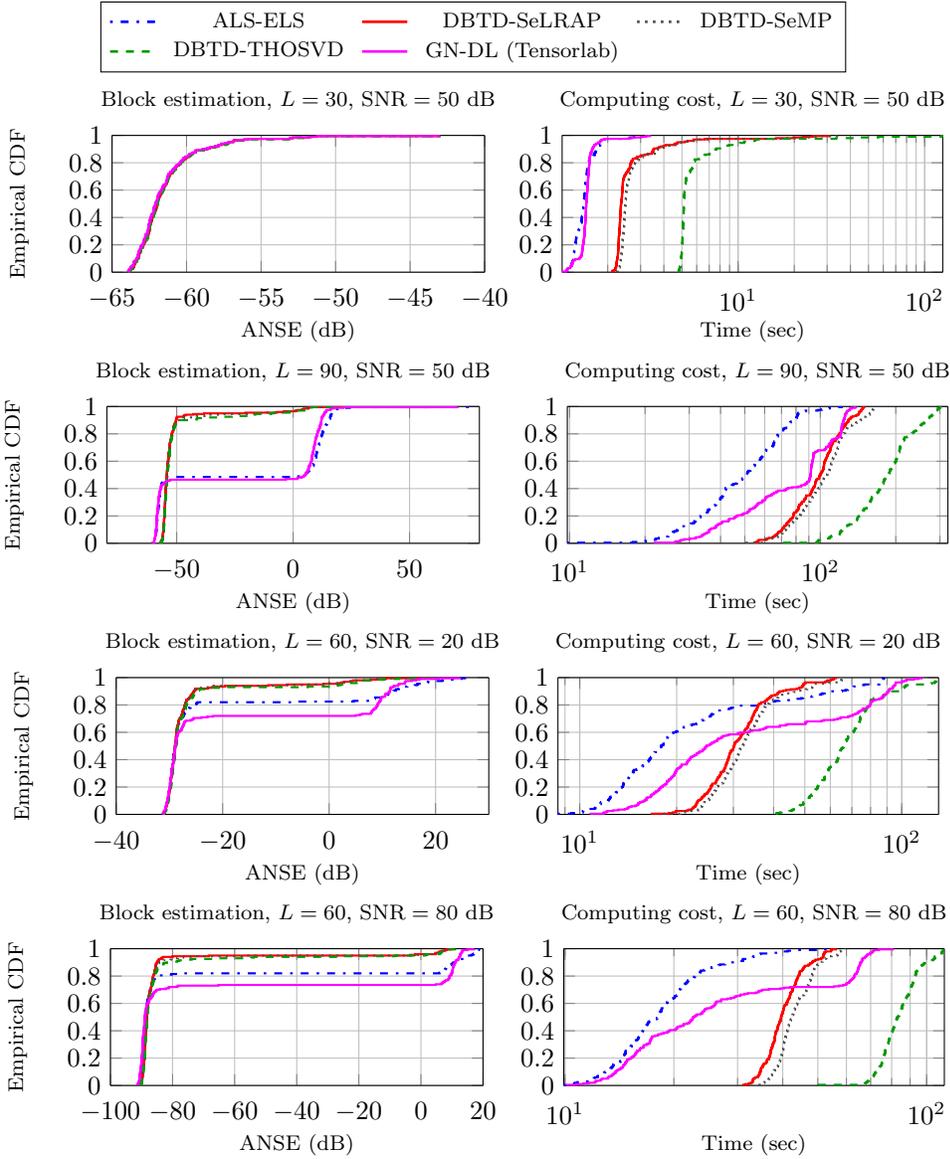


FIGURE 5. Performance of  $BTD-(1, L, L)$  algorithms for  $\mathbf{i} = (20, 150, 150)$ ,  $R = 3$ ,  $\rho = 0.2$ ,  $\lambda_r \sim \mathcal{N}(1, 0.1)$ ,  $L \in \{30, 90\}$  and  $SNR \in \{20, 80\}$  dB.

- 649 [4] L. DE LATHAUWER, *Decompositions of a higher-order tensor in block terms—Part II: Definitions*  
650 *and uniqueness*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1033–  
651 1066.
- 652 [5] ———, *Blind separation of exponential polynomials and the decomposition of a tensor in rank-*  
653  *$(L_r, L_r, 1)$  terms*, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 1451–  
654 1474.
- 655 [6] L. DE LATHAUWER AND A. DE BAYNAST, *Blind deconvolution of ds-cdma signals by means of*  
656 *decomposition in rank- $(1, L, L)$  terms*, IEEE Transactions on Signal Processing, 56 (2008),  
657 pp. 1562–1571.
- 658 [7] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *A multilinear singular value decom-*  
659 *position*, SIAM Journal on Matrix Analysis and Applications, 21 (2000), pp. 1253–1278.

- 660 [8] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *On the best rank-1 and rank-*  
661  *$(r_1, r_2, \dots, r_n)$  approximation of higher-order tensors*, SIAM J. Matrix Anal. Applicat., 21  
662 (2000), pp. 1324–1342.
- 663 [9] L. DE LATHAUWER AND D. NION, *Decompositions of a higher-order tensor in block terms—Part*  
664 *III: Alternating least squares algorithms*, SIAM journal on Matrix Analysis and Applica-  
665 tions, 30 (2008), pp. 1067–1083.
- 666 [10] V. DE SILVA AND L.-H. LIM, *Tensor rank and the ill-posedness of the best low-rank approxi-*  
667 *mation problem*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1084–  
668 1127.
- 669 [11] L. ELDÉN AND B. SAVAS, *A Newton–Grassmann method for computing the best multilinear*  
670 *rank- $(r_1, r_2, r_3)$  approximation of a tensor*, SIAM J. Matrix Anal. Applicat., 31 (2009),  
671 pp. 248–271.
- 672 [12] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Matrix Computations, Johns Hop-  
673 kins University Press, 2012.
- 674 [13] J. H. DE M. GOULART AND P. COMON, *A novel non-iterative algorithm for low-multilinear-*  
675 *rank tensor approximation*, in European Signal Processing Conference (EUSIPCO), Kos,  
676 Greece, Sept. 2017. (submitted).
- 677 [14] J. H. DE M. GOULART AND G. FAVIER, *Low-rank tensor recovery using sequentially optimal*  
678 *modal projections in iterative hard thresholding (SeMPIHT)*, SIAM Journal on Scientific  
679 Computing, (2017). (to appear).
- 680 [15] M. HAARDT, F. ROEMER, AND G. DEL GALDO, *Higher-order SVD-based subspace estimation to*  
681 *improve the parameter estimation accuracy in multidimensional harmonic retrieval prob-*  
682 *lems*, 56 (2008), pp. 3198–3213.
- 683 [16] M. ISHTEVA, P.-A. ABSIL, AND P. VAN DOOREN, *Jacobi algorithm for the best low multilinear*  
684 *rank approximation of symmetric tensors*, SIAM J. Matrix Anal. Applicat., 34 (2013),  
685 pp. 651–672.
- 686 [17] M. ISHTEVA, P.-A. ABSIL, S. VAN HUFFEL, AND L. DE LATHAUWER, *On the best low multilinear*  
687 *rank approximation of higher-order tensors*, in Recent Adv. Opt. Applicat. Eng., Springer,  
688 2010, pp. 145–164.
- 689 [18] D. NION AND L. DE LATHAUWER, *An enhanced line search scheme for complex-valued tensor*  
690 *decompositions. application in DS-CDMA*, Signal Processing, 88 (2008), pp. 749–755.
- 691 [19] D. NION AND N. D. SIDIROPOULOS, *Tensor algebra and multidimensional harmonic retrieval*  
692 *in signal processing for MIMO radar*, IEEE Transactions on Signal Processing, 58 (2010),  
693 pp. 5693–5705.
- 694 [20] H. RAUHUT, R. SCHNEIDER, AND Z. STOJANAC, *Low rank tensor recovery via iterative hard*  
695 *thresholding*, in Proceedings of the 10th International Conference on Sampling Theory and  
696 Applications, 2013.
- 697 [21] H. RAUHUT, R. SCHNEIDER, AND Ž. STOJANAC, *Low rank tensor recovery via iterative hard*  
698 *thresholding*, Linear Algebra and its Applications, 523 (2017), pp. 220–262.
- 699 [22] L. N. RIBEIRO, A. L. F. DE ALMEIDA, AND V. ZARZOSO, *Enhanced block term decomposition for*  
700 *atrial activity extraction in atrial fibrillation ecg*, in IEEE Sensor Array and Multichannel  
701 Signal Processing Workshop (SAM), IEEE, 2016, pp. 1–5.
- 702 [23] L. SORBER, M. VAN BAREL, AND L. DE LATHAUWER, *Optimization-based algorithms for ten-*  
703 *sor decompositions: Canonical polyadic decomposition, decomposition in rank- $(L_r, L_r, 1)$*   
704 *terms, and a new generalization*, SIAM Journal on Optimization, 23 (2013), pp. 695–720.
- 705 [24] J. SPIEGELBERG, J. RUSZ, AND K. PELCKMANS, *Tensor decompositions for the analysis of*  
706 *atomic resolution electron energy loss spectra*, Ultramicroscopy, (2017).
- 707 [25] N. VANNIEUWENHOVEN, R. VANDEBRIL, AND K. MEERBERGEN, *A new truncation strategy for*  
708 *the higher-order singular value decomposition*, SIAM Journal on Scientific Computing, 34  
709 (2012), pp. A1027–A1052.
- 710 [26] N. VERVLIET, O. DEBALS, L. SORBER, M. V. BAREL, AND L. DE LATHAUWER, *Tensorlab 3.0*,  
711 Mar. 2016. Available online. URL: <http://www.tensorlab.net>.
- 712 [27] A. WOŁCZOWSKI AND R. ZDUNEK, *Electromyography and mechanomyography signal recog-*  
713 *nition: Experimental analysis using multi-way array decomposition methods*, Biocy-  
714 bern. Biomed. Eng., 37 (2017), pp. 103–113.
- 715 [28] T. ZHANG AND G. H. GOLUB, *Rank-one approximation to high order tensors*, SIAM Journal  
716 on Matrix Analysis and Applications, 23 (2001), pp. 534–550.