Abstract—Because of the attractiveness of the canonical polyadic (CP) tensor decomposition in various applications, several algorithms have been designed to compute it, but efficient ones are still lacking. Iterative deflation algorithms based on successive rank-1 approximations can be used to perform this task, since the latter are rather easy to compute. We first present an algebraic rank-1 approximation method that performs better than the standard higher-order singular value decomposition (HOSVD) for three-way tensors. Second, we propose a new iterative rank-1 approximation algorithm that improves any other rank-1 approximation method. Third, we describe a probabilistic framework allowing to study the convergence of deflation CP decomposition (DCPD) algorithms based on successive rank-1 approximations. A set of computer experiments then validates theoretical results and demonstrates the efficiency of DCPD algorithms compared to other ones.

Index Terms—rank-1 approximation, Canonical Polyadic, tensor decomposition, iterative deflation, blind source separation

I. INTRODUCTION

In the last years, tensors have been playing an important role in many applications such as blind source separation [1], [2], telecommunications [3], chemometrics [4], neuroscience [5], sensor array processing [6] and data mining [7]. The attractiveness behind tensors lies in the uniqueness of their canonical polyadic (CP) decomposition under mild conditions [8], which is a powerful property not shared by standard matrix-based tools. There are several methods to compute the CP tensor decomposition. We will point out here some of the most used methods among many others. For the exact CP decomposition, [9] proposes a direct computation to decompose $2 \times n \times n$ tensors. In [10], a generalization of Sylvester’s algorithm is described for decomposing symmetric tensors. In [11], one can use simultaneous matrix diagonalization by congruence, provided that the rank of the tensor is smaller than its greatest dimension. An approach based on eigenvectors of tensors is proposed in [12].

In practice, tensors are corrupted by noise so that one needs to compute an approximate decomposition of given rank. Computing the exact CP decomposition is difficult [13], but finding a lower-rank approximation is even harder. In fact, this is an ill-posed problem in general [14]. Nevertheless, some useful algorithms have been conceived to solve locally the low-rank approximation problem. This kind of algorithms can be found in [15], [16], [17], [18], [19]. One of the most widely used is the alternating least squares (ALS) algorithm [3], which is an iterative method that consists in conditionally updating in an alternate way, the matrix factors composing the CP decomposition. Other gradient and Newton-based methods estimate the factor matrices all-at-once. However, all these methods have disappointing convergence properties [15], [20]. Another kind of algorithms is based on rank-1 deflation. It is known that the conventional deflation works for matrices but not for tensors [21]. In [22], the authors propose deflation methods that work only if the rank of the tensor is not larger than its dimensions. In [23], ALS is used to update the columns of matrix factors in a deflation procedure, but for non-negative tensors. However, these deflation methods strongly depend on initialization and no convergence study has been conducted.

In the same vein of iterative deflation, the authors proposed in [24] a deflation-based CP decomposition (DCPD), based on successive rank-1 approximations computed by low-complexity methods. Other rank-1 approximation methods can be used in DCPD, for instance, ALS. However, the latter exhibits an unbounded complexity and no satisfactory convergence study is available for rank-1 approximation apart [26], which shows results on the global convergence for generic tensors in the sense that, for any initialization, ALS converges to the same point in general. A quasi-Newton method defined on Grassmannians are developed in [27]. However, they exhibit an unbounded complexity to compute rank-1 approximations, since these methods need to be iterated. In [28], the best rank-1 approximation problem can be computed by means of an algebraic geometry moment method, but it is only applicable to very small tensor dimensions since the number of variables grows exponentially when building convex relaxations. Moreover, even for small dimensions, its convergence is very slow. In [29], the authors propose an improvement method for [28] based on border basis relaxation, but again the method is limited to small tensor dimensions. Semidefinite relaxations are proposed in [30] to compute rank-1 approximations, however the convergence becomes very slow when dimensions are large.

In this paper, we report mainly three contributions. First, we propose an algebraic rank-1 approximation method, namely the sequential rank-1 approximation and projection (SeROAP), which can perform better than the standard truncated high-order singular value decomposition (THOSVD) [25]. Indeed, we prove that the rank-1 approximation performed by SeROAP is always better than the one obtained from THOSVD for three-way tensors. Moreover, for large dimensions and small orders, we show that the computational complexity of SeROAP is dramatically lower than that of THOSVD.

Second, we propose an alternating eigenvalue rank-1 iterative algorithm for three-way tensors, namely CE (coupled-
eigenvalue), that improves other rank-1 approximation algorithms. We prove that if the solution obtained from some rank-1 approximation algorithm (e.g., SeROAP, THOSVD) is the input of the CE algorithm, the performed rank-1 approximation remains the same in the worst case. We also prove that the convergence to a stationary point is always guaranteed. Actually, results have shown that when the initialization of the CE algorithm is close enough to the global solution, it recovers the best rank-1 approximation. Furthermore, when one dimension is much larger than the other two dimensions, the computational complexity of the CE algorithm can be lower than that of the standard ALS algorithm.

Third, we perform a theoretical study on deflation in order to analyze the convergence of the DCPD algorithm. In a first stage, we show that the norm of residuals is monotonically reduced within the iterative deflation process. We also prove that the DCPD algorithm recovers the exact CP decomposition of a given tensor when residuals do not fall within a cone with an arbitrary small volume. In a second stage, we prove that the iterative deflation method can reduce the norm of the initial residual by a factor smaller than $(\sin(\beta))^{L-1}$ ($\beta$ being the angle of a suitable cone where the residuals can fall in) after $L$ iterations with high probability, when tensors are distributed according to an absolutely continuous probability measure, and the probability function of residuals is continuous on some suitable angular interval. We also present a conjecture stating the existence of probability measures ensuring the convergence of the DCPD algorithm to an exact decomposition with high probability.

The paper is organized as follows. In Section III some standard iterative methods and the DCPD algorithm as well as SeROAP and THOSVD methods are described. The computational complexity per iteration for each algorithm is provided. The next two sections form the core of the paper. In Section IV we first prove that SeROAP performs better than THOSVD as far as rank-1 approximations of three-way tensors is concerned. Then in a second stage, the CE algorithm is presented as well as the proof that it can refine any other rank-1 approximation method and the proof of its convergence. In Section V, among other theoretical results, we study conditions ensuring the convergence of the DCPD algorithm. Finally, in Section VI, computer results show satisfactory performances of the proposed DCPD and rank-1 approximation methods, compared to other related algorithms, even under noisy scenarios.

II. NOTATION

The notation employed in this paper is the following. Scalar numbers are denoted by lowercase letters and vectors by boldface lowercase ones. For matrices and tensors, boldface capital and calligraphic letters are used, respectively. Plain capitals are used to denote array dimensions. Depending on the context, greek letters can denote either scalars, vectors, matrices or tensors. The symbols $\square, \circ, \otimes$ and $\otimes$ denote the Hadamard, Khatri-Rao, Kronecker and tensorial products, respectively, and $\dagger$ denotes matrix pseudo-inversion. The Euclidean scalar product between tensors is denoted by $\langle \mathbf{T}, \mathbf{U} \rangle = \sum_{i_1, \ldots, i_N} T_{i_1, \ldots, i_N} U_{i_1, \ldots, i_N}^*$. The angle between two tensors $\mathbf{U}$ and $\mathbf{V}$ will refer to $\arccos\{|\mathbf{U}, \mathbf{V}|\}/\|\mathbf{U}\|_F \|\mathbf{V}\|_F \in [0, \pi/2]$. $\| \cdot \|$ then denotes the Frobenius norm induced by the previous scalar product. We shall also use the $\ell^2$ operator norm $\| \cdot \|_2$ for matrices, which corresponds to the largest singular value. The mode-$n$ unfolding of a tensor $\mathbf{T}$ is denoted as $\mathbf{T}(n)$, as proposed in [10]. $\mathbf{T}(i, j, k)$, $\mathbf{T}(i, j, k)$ and $\mathbf{T}(i, j, :)$ denote vector slices of tensor $\mathbf{T}$.

The operator vec is the vectorization operator that stacks the columns of a matrix into a long column vector, and Unvec is its reverse operator. $C^k$ is the set of functions having continuous $k$th derivatives. Finally, $\mathbb{K}$ is either the real or the complex field.

III. DESCRIPTION OF ALGORITHMS AND COMPLEXITY ANALYSIS

This section presents the description of some CP decomposition algorithms. In order to support further results, the complexity per iteration is calculated here for each algorithm using Landau’s notation, denoted by $O$, and counting only multiplications, as recommended in [31]. From Section III-A to Section III-B we describe two classical algorithms known in literature: ALS and conjugate gradient [15]. In Section III-C the DCPD algorithm is presented.

For the CP decomposition algorithms described in the following, the input parameter $R$ denotes the rank of the output tensor. Assuming $R_0$ is the rank of the input tensor $\mathbf{T}$, if $R_0 \leq R$, then the algorithms perform an exact decomposition. On the other hand, if $R_0 > R$, a lower rank-$R$ approximation is computed.

A. Alternating least squares (ALS)

The most commonly used algorithm for solving the CP decomposition is ALS [4]. The goal is to update alternately each factor matrix in each iteration by solving a least squares problem conditioned on previous updates of the other factor matrices. There is no guarantee of convergence to the global solution, nor even to a critical point. The implementation is quite simple and it is detailed in Alg.1

Algorithm 1: ALS algorithm

The complexity per iteration (repeat loop) of ALS may be calculated as follows. The computation of matrix $\mathbf{V}$ needs...
$(I_1 + \ldots + I_{n-1} + I_{n+1} \ldots I_N)R^2 + (N-2)R^2$ operations (multiplications). The pseudo-inverse of $V$ is calculated by resorting to an SVD. For a $m \times n$ rank-$r$ matrix, with $m \geq n \geq r$, the explicit calculation of diagonal, left singular, and right singular matrices require $2mn^2 - 2n^3/3, 5mr^2 - r^3/3$, and $5mr^2 - r^3/3$ multiplications, respectively [31]. Thus, assuming for simplicity that $V$ is a non-singular matrix, the number of operations to calculate its pseudo inverse is $35R^3/3 + R^2$. For updating $A^{(n)}$, we need

$$R \prod_{i=1, i \neq n}^N I_i + R \prod_{j=1}^N I_j + I_n R^2$$

multiplications. These calculations must be performed for each $n \in \{1, \ldots, N\}$. Thus, the number of operations per iteration for ALS is dominated by the term composed of the product of all dimensions. Hence,

$$\#op = O\left\{NR \prod_{j=1}^N I_j\right\}. \quad (1)$$

B. Conjugate gradient (CG)

The conjugate gradient algorithm (CG) is a faster algorithm than the well known gradient descent [15]. Here, we use the optimum step size and the Polak-Ribière implementation [22] for updating the parameter $\beta$ in the algorithm presented in Alg 2. The number of operations for computing each vector $g_n$ is

$$R^2 \sum_{j=1}^N I_j + (N-2)R^2 + I_n R^2 + R \prod_{i=1, i \neq n}^N I_i + R \prod_{j=1}^N I_j$$

The computation of the step size $\mu^*$ is dominated by the number of multiplications needed to determine all the coefficients but one of a $2N$-degree polynomial generated from the enhanced line search (ELS) method, which is given by [17]:

$$(2^N R + O\{N^2\}) \prod_{j=1}^N I_j$$

The parameter $\beta$ requires only $1 + 2R \sum_{j=1}^N I_j$ multiplications. Hence, the CG algorithm with ELS has a total complexity given by

$$\#op = O\left\{((2^N + 1)R + N^2) \prod_{j=1}^N I_j\right\}. \quad (2)$$

C. Deflation-based CP decomposition (DCPD)

The computation of a rank-1 approximation is the key of the DCPD algorithm. We present here two methods for computing a rank-1 approximation referred to as THOSVD and SeROAP [24].

---

**Algorithm 2: CG algorithm**

1) Truncation of higher order singular value decomposition - THOSVD: The algorithm is described in [33]. For computing the first right singular vector, we do not need compute the complete SVD. According to [33], we can compute the best rank-1 approximation of a $m \times n$ matrix in $k$ steps by using the Lanczos algorithm, with a complexity $O\{2kmn\}$. Hence, the accumulated complexity computed for all $u_n$ is equal to $O\{2Nk \prod_{j=1}^N I_j\}$. The computation of $\mathcal{U}$ requires $\prod_{j=1}^N I_j$ flops. The contraction to obtain $\lambda$ also needs $\prod_{j=1}^N I_j$ operations. To sum up, the total number of operations of THOSVD is of order:

$$O\left\{(2Nk + 2) \prod_{j=1}^N I_j\right\}$$

---

**Algorithm 3: THOSVD algorithm**

2) Sequential rank-1 approximation and projection - SeROAP: Without loss of generality, consider $I_1 \geq I_2 \geq \ldots \geq I_N$. The SeROAP algorithm [24] goes along the lines depicted in Alg 4. In the first for loop, we compute $N-2$ right singular vectors of matrices whose size is successively reduced. For this step, the complexity is $O\{2k \sum_{i=1}^{N-2} (\prod_{j=i+1}^N I_j)\}$. The
computation of vectors $u, v$ and $w$ has complexity $O((2k + 1)I_{N-1}I_N)$. Next, the second for loop performs $N - 2$ successive projections of the rows of matrices $V_n$ onto the vectors $w$. We need here $2 \sum_{i=1}^{N} (\prod_{j=i}^{N} I_j)$ operations.

Algorithm 4: SeROAP algorithm

For large dimensions and small $N$, the complexity of SeROAP is dominated by:

$$O\left((2k + 2) \prod_{j=1}^{N} I_j\right)$$

which can be significantly smaller than that of THOSVD. An example of typical execution is given in the Appendix.

3) Description and complexity of DCPD: The DCPD is an iterative deflation algorithm [24] that computes the CP decomposition for real or complex tensors. As summarized in Alg 5, it proceeds as follows. In the first for loop, we compute the rank-1 tensors $\mathcal{X}[1, 1], \ldots, \mathcal{X}[R, 1]$ by successive rank-1 approximations and subtractions. Since the rank of the tensor does not decrease with subtractions in general [21], a residual $\mathcal{E}[R, 1]$ is then produced. In the iterative process (repeat loop), a new rank-1 component is generated from the sum of the previous residual and $\mathcal{X}[1, 1]$, and a new residual $\mathcal{E}[1, 2]$ is produced with the subtraction $\mathcal{Y}[1,2] - \mathcal{X}[1,1]$. The tensor $\mathcal{Y}[1, 2]$ is updated within the if-else condition. By applying the same procedure to the other components, we update all $R$ rank-1 tensors, so that another residual $\mathcal{E}[R, 2]$ is generated in the end of the second for loop. The second loop continues to execute until some stopping criterion is satisfied, and all rank-1 components of $\mathcal{T}$ can be recovered.

The complexity per iteration is dominated by the rank-1 approximation function $\phi$, which is computed $R$ times. Therefore, the complexity of DCPD is

$$\#op = O\left(\left((2Nk + 2) R \prod_{j=1}^{N} I_j\right)\right), \quad \text{(3)}$$

1Matlab codes are available at http://www.gipsa-lab.grenoble-inp.fr/~pierrecomon/TensorPackage/tensorPackage.html.

Algorithm 5: DCPD algorithm

for the T-HSOVD algorithm, and

$$\#op = O\left(\left((2k + 2) R \prod_{j=1}^{N} I_j\right)\right), \quad \text{(4)}$$

for the SeROAP algorithm.

IV. RANK-1 APPROXIMATION

This section is divided in two parts. The first one presents a more detailed study of the THOSVD and SeROAP rank-1 approximation methods. For three-way tensors, we show that SeROAP is a better choice than THOSVD because the former presents a better rank-1 approximation, which ensures a more probable monotonic decrease of the residual $||\mathcal{E}[R, 1]||$ within the DCPD algorithm [24]. A new rank-1 approximation algorithm is described in a second part, and is proved to perform better than any other rank-1 approximation method.

A. THOSVD vs SeROAP

Since we do not have at our disposal an efficient method to compute quickly the best rank-1 approximation of a tensor, we should compute a suboptimal rank-1 approximation in a tractable way. The THOSVD and SeROAP algorithms can
perform this task, as presented in Section III. The question that arises is: which algorithm performs the best? The proposition 4.1 below shows that SeROAP performs better than THOSVD for three-way tensors. For simplicity, the notation of the unfolding matrices does not present in this section.

**Proposition 4.1:** Let $T \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ be a 3-order tensor. Let also $\phi^{TH}(T)$ and $\phi^{Se}(T)$ be the rank-1 approximations delivered by THOSVD and SeROAP algorithms, respectively. Then the inequality $\|T - \phi^{Se}(T)\| \leq \|T - \phi^{TH}(T)\|$ holds.

**Proof:** Let $T, T^{\phi}$ and $T^{Se}$ be some mode unfolding of tensors $T, \phi^{TH}(T)$ and $\phi^{Se}(T)$, respectively. Assuming mode-1 unfolding for the THOSVD algorithm, we have

$$\|T - T^{\phi}\|^2 = \|T - \lambda u_1(u_3 \otimes u_2)^T\|^2,$$

where $\lambda, u_1, u_2$, and $u_3$ are obtained from Alg. 3. Since $\lambda$ is the contraction of $T$ on $U$, we plug it into the previous equation and we obtain, after simplifications,

$$\|T - T^{\phi}\|^2 = \|T\|^2 - \|\lambda\|^2,$$

with $\lambda = u_1^H(T(u_3 \otimes u_2))$. Let, $u_1^H T = \|T\|^2 v_1^H$ since $(u_1, v_1, ||T||_2)$ is the dominant singular triplet of matrix $T$. Hence $\|\lambda\|^2 = \|T\|^2 \|v_1^H (u_3 \otimes u_2)\|^2$.

On the other hand for SeROAP, we have

$$\|T - T^{Se}\|^2 = \|T - \lambda w v_1^H w^H + S\|^2$$

where $S$ is a semidefinite positive matrix. Hence, we have

$$w^H T H T w = \|T\|^2 \|v_1 w_1^H w + c\|^2,$$

with $c \geq 0$. To complete the proof of the proposition, we just need to show that $\|v_1 w_1\|^2 \geq \|u_3 \otimes u_2\|^2$, or equivalently that

$$\langle w, v_1 \rangle \geq \langle u_3 \otimes u_2, v_1 \rangle.$$

This is true, because $w$ is by construction (cf. Alg. 4) the vector closest to $v_1$ among all vectors of the form $a \otimes b$ where $a$ and $b$ have unit norm.

**B. Coupled-eigenvalue rank-1 approximation**

This section presents an alternating eigenvalue method for three-way tensors that can improve local solutions obtained from any other rank-1 approximation method (e.g. SeROAP and THOSVD algorithms). Actually, simulations have shown that the global solution is always attained if the initial approximation is close enough.

Let $v_{i_3}$ be the vectorization of slice $i_3$, $1 \leq i_3 \leq I_3$ (we have chosen the third mode) of tensor $T$. The rank-1 approximation problem can be stated as

$$[\alpha_{opt}, x_{opt}, y_{opt}] = \arg \min_{\alpha, x, y} \mathcal{T}(x, y, \alpha)$$

s.t. $\|x\|_2 = 1$, $\|y\|_2 = 1$, \hspace{1cm} (5)

with $\mathcal{T}(x, y, \alpha) = \sum_{i_3=1}^{I_3} \|t_{i_3} - \alpha_i \langle y^* \otimes x \rangle\|^2$ and $\alpha = [\alpha_1 \cdots \alpha_{I_3}]$.

Plugging the optimal value of $\alpha_i$ into Problem (5), namely $(y^* \otimes x)^H t_{i_3}$, we can rewrite it as the following equivalent maximization problem

$$[x_{opt}, y_{opt}] = \arg \max_{x, y} z^H M z$$

s.t. $z = y^* \otimes x$, $\|y\|_2 = 1$, $\|x\|_2 = 1$, \hspace{1cm} (6)

where $M = \sum_{i_3=1}^{I_3} t_{i_3} t_{i_3}^H$.

Now, we decompose $M$ as a sum of Kronecker products. This can be done by reshaping $M$ and applying the SVD decomposition [54]. Thus, $M$ can be given by

$$M = \sum_{r=1}^{R'} \mathcal{Q}^{(r)} \circ \mathcal{P}^{(r)},$$

with the Hermitian matrices $\mathcal{P}^{(r)} \in \mathbb{C}^{I_1 \times I_1}$ and $\mathcal{Q}^{(r)} \in \mathbb{C}^{I_2 \times I_2}$. $R'$ is the Kronecker rank of $M$ satisfying $R' \leq \min(I_1^2, I_2^2)$. Substituting $M$ into Problem (6), we have:

$$[x_{opt}, y_{opt}] = \arg \max_{x, y} \sum_{r=1}^{R'} \langle y^H \mathcal{Q}^{(r)} x \rangle \langle x^H \mathcal{P}^{(r)} x \rangle$$

s.t. $\|y\|_2 = 1$, $\|x\|_2 = 1$. \hspace{1cm} (7)

Let $\mathcal{L}$ be the Lagrangian function given by

$$\mathcal{L} = -\sum_{r=1}^{R'} \langle y^H \mathcal{Q}^{(r)} x \rangle \langle x^H \mathcal{P}^{(r)} x \rangle + \eta_1(\|y\|_2^2 - 1) + \eta_2(\|x\|_2^2 - 1),$$

where $\eta_1$ and $\eta_2$ are the Lagrange multipliers. By computing the critical points, we obtain a pair of coupled eigenvalue problems

$$\begin{bmatrix}
y^H A_{(1,1)} y & \cdots & y^H A_{(1,I_1)} y \\
n_{1,2} & \cdots & n_{1,2} \\
y^H A_{(I_1,1)} y & \cdots & y^H A_{(I_1,I_1)} y
\end{bmatrix} x = \lambda x$$ \hspace{1cm} (8)

and

$$\begin{bmatrix}
x^H B_{(1,1)} x & \cdots & x^H B_{(1,I_2)} x \\
n_{2,3} & \cdots & n_{2,3} \\
x^H B_{(I_2,1)} x & \cdots & x^H B_{(I_2,I_2)} x
\end{bmatrix} y = \lambda y, \hspace{1cm} (9)$$

where $\lambda = \eta_1 = \eta_2$, $A_{(m,n)} = \sum_{r=1}^{R'} P^{(r)}_{mn} \mathcal{Q}^{(r)}$, and $B_{(k,l)} = \sum_{r=1}^{R'} Q^{(r)}_{kl} \mathcal{P}^{(r)}$, with $1 \leq m, n \leq I_1$ and $1 \leq k, l \leq I_2$.

The coupled-eigenvalue algorithm is presented in Alg. 5.

We can initialize the algorithm by computing $x_0$ and $y_0$ from the rank-1 approximation solution obtained with SeROAP, THOSVD or any other rank-1 approximation method.

The complexity per iteration of the CE algorithm is dominated by the construction of the matrices in the LHS of (8) and (9), which is of order $\mathcal{O}\{\min(I_1^2 I_2^2, I_1^3 I_2, I_2^3 I_1)\}$. Suppose $I_3$ is the largest dimension. If $I_3 \ll I_1 I_2$, then we can take advantage of the CE algorithm in terms of complexity in comparison with the ALS algorithm. Indeed, the complexity per iteration of ALS for rank-1 approximation is of order
Algorithm 6: CE rank-1 approximation. Above, we chose to start with $x$, but we could equivalently have started with $y$.

$O(3I_1I_2I_3)$, which is higher than that of the CE algorithm in this case. Notice, however, that a properly comparison makes sense if the same initialization is employed in both algorithms.

The following proposition shows that the above algorithm improves (in worst case the solution remains the same) any rank-1 approximation algorithm.

Proposition 4.2: Let $\phi(T)$ be a rank-1 approximation of a three-way tensor $T$. If $\phi(T)$ is the input of the CE algorithm and $\phi^*(T)$ the output, then the inequality $\|T - \phi^*(T)\| \leq \|T - \phi(T)\|$ holds.

Proof: Plugging the expression of $A_{(m,n)}$ into equation (8), we obtain, after simplifications,

$$\lambda = \sum_{r=1}^{R'} (y^H Q(r)^* y) (x^H P(r) x),$$

which is the objective function of Problem (7). The same result is obtained when the matrix $B_{(k,l)}$ is plugged into equation (9). Now $\forall t \geq 1$, let $\lambda_{t+1}^{(y)}$ and $\lambda_{t+1}^{(x)}$ be the maximal eigenvalues whose eigenvectors are $x_t$ and $y_t$, respectively.

The eigenpair $(\lambda_{t+1}^{(y)}, y_{t+1})$ obtained by solving equation (9) with $x = x_t$, is solution of the maximization problem

$$\lambda_{t+1}^{(y)} = \max_{\|y\|_2 = 1} \sum_{r=1}^{R'} (y^H Q(r)^* y) (x^H_t P(r) x_t).$$

Also, the eigenpair $(\lambda_{t+1}^{(x)}, x_{t+1})$ obtained by setting $y = y_{t+1}$ in equation (8), is solution of the problem

$$\lambda_{t+1}^{(x)} = \max_{\|x_t\|_2 = 1} \sum_{r=1}^{R'} (y_{t+1}^H Q(r)^* y_{t+1}) (x^H_t P(r) x_t).$$

Since

$$\max_{\|x_t\|_2 = 1} \sum_{r=1}^{R'} (y_{t+1}^H Q(r)^* y_{t+1}) (x^H_t P(r) x_t) = \sum_{r=1}^{R'} (y_{t+1}^H Q(r)^* y_{t+1}) (x_{t+1}^H P(r) x_{t+1}),$$

it follows in particular that

$$\sum_{r=1}^{R'} (y_{t+1}^H Q(r)^* y_{t+1}) (x_{t+1}^H P(r) x_{t+1}) \leq \sum_{r=1}^{R'} (y_{t+1}^H Q(r)^* y_{t+1}) (x_t^H P(r) x_t),$$

which implies that $\lambda_{t+1}^{(y)} \leq \lambda_{t+1}^{(x)}$.

Similarly, plugging $x_{t+1}$ into equation (9), we can conclude that $\lambda_{t+1}^{(x)} \leq \lambda_{t+1}^{(x)}$ for the reason that

$$\sum_{r=1}^{R'} (y_{t+1}^H Q(r)^* y_{t+1}) (x_{t+1}^H P(r) x_{t+1}) \leq \sum_{r=1}^{R'} (y_{t+1}^H Q(r)^* y_{t+1}) (x_{t+1}^H P(r) x_{t+1}).$$

Hence, the sequence

$$\{\lambda_t\}_{t \in \mathbb{N}} = \{\lambda_1^{(y)}, \lambda_{t+1}^{(x)}, \lambda_{t+1}^{(y)}, \lambda_{t+1}^{(x)}, \ldots\}$$

is monotonically non-decreasing. The same conclusion would be achieved if we begin by plugging $x_t$ into equation (9).

Now, let $\phi(T) = x_0 \otimes y_0 \otimes \alpha_0$ be a rank-1 approximation obtained with any other method. Assume $x_0$ and $y_0$ are unit vectors, and define $\lambda_0 = \sum_{r=1}^{R'} (y_0^H Q(r)^* y_0) (x_0^H P(r) x_0)$. By setting $x = x_0$ in equation (9) in the first iteration (a similar operation would be possible for $y_0$ in equation (8)), we clearly have $\lambda_0 \leq \lambda_1^{(y)} \leq \lambda_{t_{\text{max}}}^{(y)}$, where $t_{\text{max}}$ is the iteration in which the stopping criterion is satisfied.

Since the optimization problems (5) and (7) are equivalent, $\alpha_{t_{\text{max}}}$, $1 \leq i \leq I_3$, can be obtained by performing the scalar product between vectors $t_{i_3}$ and $y_{t_{\text{max}}}^* \otimes x_{i_{\text{max}}}$ (which is equivalent to contracting tensor $T$ on $x_{t_{\text{max}}}$ and $y_{t_{\text{max}}}^*$). Hence, the tensor

$$\phi^*(T) = x_{t_{\text{max}}}^* \otimes y_{t_{\text{max}}}^* \otimes \alpha^*$$

is a better rank-1 approximation of $T$ than $\phi(T)$, implying $\|T - \phi^*(T)\| \leq \|T - \phi(T)\|$.

Proposition 4.3: For any input $\phi(T)$, the CE algorithm converges to a stationary point.

Proof: In the proof of Proposition 4.2 we have shown that $\{T_t\}_{t \in \mathbb{N}}$ is monotonically non-decreasing for any input $\phi(T)$. Let $p^*$ be the maximum of the objective (7). Since the best rank-1 approximation problem always has a solution, then $p^* < \infty$. But $\max_{x} \mathcal{Y}(x, y_{t+1}, \alpha) \leq \max_{x} \mathcal{Y}(x, y, \alpha)$, which implies that $\{T_t\}_{t \in \mathbb{N}}$ is bounded above by $p^*$. Since $\{T_t\}$ is a real non decreasing sequence bounded above, it converges to a limit $\hat{Y}$, $\hat{Y} \leq p^*$.

V. DEFLECTION

In [24], we proved that for a rank-$R$ tensor, the normalized residual $(\|E[R, l]\|)$ is a monotonically decreasing sequence when the best rank-1 approximation is assumed within.
DCPD. In this section, a thorough theoretical analysis and new results are presented. Based on a geometric approach, we sketch an analysis of the convergence of the DCDP algorithm, including a conjecture that it converges to an exact decomposition with high probability when tensors within \( \mathcal{T}(R) = \{ \mathcal{T} \in \mathcal{F} : \text{rank} \{ \mathcal{T} \} \leq R \} \) are distributed according to absolutely continuous probability measures.

First, let us take a closer look at the 2D geometric interpretation of the DCPD algorithm. Figure 1 depicts the \([r, l]\)-iteration for \( r > 1 \), so that \( \gamma[r, l] \) is the angle between the tensors \( \mathcal{E}[r-1, l] \) and \( \mathcal{X}[1, l-1] \). For \( r = 1 \), the residual \( \mathcal{E}[1, l-1] \) can be just replaced with \( \mathcal{E}[R, l-1] \) in the figure, and \( \gamma[1, l] \) is then defined from \( \mathcal{E}[R, l-1] \) and \( \mathcal{X}[1, l-1] \).

Before stating some theoretical results on the DCPD algorithm, we present a fundamental lemma related to the error in rank-1 approximations of tensors of the form \( \mathcal{X} + \mathcal{E} \), where \( \mathcal{X} \) is a rank-1 tensor and \( \mathcal{E} \) any other tensor, both with entries in some field \( \mathbb{K} \).

**Lemma 5.1:** Let \( \mathcal{X} \) be a rank-1 tensor and \( \mathcal{E} \) the best rank-1 approximation operator. For any tensor \( \mathcal{E} \),

\[
\| \mathcal{X} + \mathcal{E} - \phi(\mathcal{X} + \mathcal{E}) \| \leq \sin(\gamma) \| \mathcal{E} \|
\]

where \( \gamma \) denotes the angle between \( \mathcal{E} \) and \( \mathcal{X} \).

**Proof:** Let \( \mathcal{P}_X(\mathcal{X} + \mathcal{E}) \) be the orthogonal projection of \( \mathcal{X} + \mathcal{E} \) onto \( \text{span}(\mathcal{X}) \). Because \( \phi(\mathcal{X} + \mathcal{E}) \) is a best rank-1 approximation of \( \mathcal{X} + \mathcal{E} \), \( \mathcal{P}_X(\mathcal{X} + \mathcal{E}) \) cannot be a strictly better rank-1 approximation than \( \phi(\mathcal{X} + \mathcal{E}) \). Thus,

\[
\| \mathcal{X} + \mathcal{E} - \phi(\mathcal{X} + \mathcal{E}) \| \leq \| \mathcal{X} + \mathcal{E} - \mathcal{P}_X(\mathcal{X} + \mathcal{E}) \|
\]

On the other hand, \( \mathcal{X} + \mathcal{E} - \mathcal{P}_X(\mathcal{X} + \mathcal{E}) \perp \mathcal{X} \). Hence, we have \( \| \mathcal{X} + \mathcal{E} - \mathcal{P}_X(\mathcal{X} + \mathcal{E}) \| = \sin(\gamma) \| \mathcal{E} \| \) by using basic trigonometry. This concludes the proof.

The following results for the DCPD algorithm stems from the previous lemma.

**Corollary 5.2:** The inequality \( \| \mathcal{E}[r, l] \| \leq \sin(\gamma[r, l]) \| \mathcal{E}[r-1, l] \| \) holds for any \( 1 < r \leq R \).

**Proof:** By replacing \( \mathcal{X}, \mathcal{E} \) and \( \gamma \) in Lemma 5.1 with \( \mathcal{X}[r, l-1], \mathcal{E}[r-1, l] \) and \( \gamma[r, l] \) respectively, the result follows directly.

**Corollary 5.3:** For any \( l > 1 \) and \( c_l = \prod_{r=1}^{R} \sin(\gamma[r, l]) \), the inequality \( \| \mathcal{E}[R, l] \| \leq c_l \| \mathcal{E}[R, l-1] \| \) holds.

**Proof:** By applying \( R - 1 \) times the result of Corollary 5.2, we have

\[
\| \mathcal{E}[r, l] \| \leq (\sin(\gamma[r, l]) \cdots \sin(\gamma[2, l])) \| \mathcal{E}[1, l] \|
\]

From Lemma 5.1 we know that

\[
\| \mathcal{E}[1, l] \| \leq \sin(\gamma[1, l]) \| \mathcal{E}[R, l-1] \|
\]

Thus, it follows that \( \| \mathcal{E}[R, l] \| \leq c_l \| \mathcal{E}[R, l-1] \| \).

Notice that the same result brought by Proposition (4.4) in [24] can be deduced from Corollary 5.3 since \( 0 \leq c_l \leq 1 \) for every iteration \( l \), which implies the monotonic decrease of the sequence \( \{ \| \mathcal{E}[R, l] \| \}_{l \in \mathbb{N}_0} \).

**Lemma 5.4:** If \( \| \mathcal{E}[R, l] \| = \| \mathcal{E}[R, l-1] \| \), then \( c_l = 1 \).

**Proof:** Because \( c_l \leq 1 \) and \( \| \mathcal{E}[R, l] \| = \| \mathcal{E}[R, l-1] \| \), one concludes directly from Corollary 5.3 that \( c_l = 1 \).

**Proposition 5.5:** Let \( \mathcal{T} \) be a tensor such that \( \text{rank} \{ \mathcal{T} \} \leq R \). An exact decomposition is recovered by the DCPD algorithm if and only if there exists for every \( (r, l) \) a half cone of angle \( \beta \) (in white in Figure 1), \( 0 \leq \beta < \pi/2 \), such that

\[
\beta \geq \max_{l > 1} \min_{1 \leq r \leq R} \gamma[r, l].
\]

**Proof:** (\( \Leftarrow \)) For any iteration \( l > 1 \), take \( \gamma[0, l] = \min_{1 \leq r \leq R} \gamma[r, l] \). Notice that \( c_l \leq \sin(\gamma[0, l]) \) from Corollary 5.3. By hypothesis, \( \sin(\gamma[0, l]) \leq \sin(\beta) \), which implies that \( \| \mathcal{E}[R, l] \| \leq c_l \| \mathcal{E}[R, l-1] \| \) \( \leq \sin(\beta) \| \mathcal{E}[R, l-1] \| \).

Because \( \beta \) is an upper bound for \( \gamma[r, l] \), \( l > 1 \), we have \( \sin(\beta) \| \mathcal{E}[R, l] \| \geq \| \mathcal{E}[R, 2] \| \Rightarrow (\sin(\beta))^{l-1} \| \mathcal{E}[R, 1] \| \geq \| \mathcal{E}[R, l] \| \). Hence, when \( l \to \infty \), \( \| \mathcal{E}[R, l] \| \to 0 \).

(\( \Rightarrow \)) Let \( l_0 \) be some iteration such that \( l_0 > 1 \). Without loss of generality, assume \( \| \mathcal{E}[R, l] \| = 0 \) for \( l \geq l_0 \) (\( l_0 \) can be arbitrarily large). Then \( \{ \| \mathcal{E}[R, l] \| \}_{l \in \mathbb{N}_0} \) is a strictly monotonically decreasing sequence for \( 1 < l < l_0 \), otherwise the algorithm would converge to a nonzero constant for some iteration smaller than \( l_0 \). Hence, for every \( (r, l) \) we can choose \( \beta, 0 \leq \beta < \pi/2 \) such that \( \beta \geq \max_{l > 1} \min_{1 \leq r \leq R} \gamma[r, l] \), and the proof is complete.
As a conclusion, if for a given iteration \( l \), all tensors \( \mathcal{E}[r, l] \), \( 1 \leq r \leq R \), fall within the gray volume depicted in Fig. 1 (the complementary of the white cone), then the sequence \( \mathcal{E}[r, l] \) does not tend to zero. Even if this gray volume can be made arbitrarily small, it is not of zero measure. So the best we can do is to prove an almost sure convergence of the DCPD algorithm to the exact decomposition under some probabilistic conditions.

**Lemma 5.6:** If tensors \( \mathcal{T} \) are distributed within \( \mathcal{T}^{(R)} \) according to an absolutely continuous probability measure, then \( \| \mathcal{E}[r, l] \| \) are absolutely continuous random variables.

*Proof:* Let \( \mathcal{D} = [I_1 \cdots I_N] \) be a specific size of \( n \)-order tensors, and let \( \mathcal{T}^{(R)} = \{ \mathcal{T} \in \mathbb{R}^{I_1 \times \cdots \times I_N} : \mathcal{T} \subset \mathcal{T}^{(R)} \} \). Because \( \mathcal{T}^{(R)} \supset \mathcal{T}_D^{(R)} \), any tensor \( \mathcal{T} \) within \( \mathcal{T}_D^{(R)} \) is also distributed according to an absolutely continuous probability measure. Via the DCDP algorithm, each rank-1 component obtained in successive deflations is also in \( \mathcal{T}^{(R)} \). Hence, since the sum (subtraction) of continuous random variables does not affect the continuity, the residuals \( \mathcal{E}[r, l] \) are also absolutely continuous random variables. Since the norm is a \( C^0 \) function in finite dimension, \( \| \mathcal{E}[r, l] \| \) is also absolutely continuous.

For the next developments, let \( Z_1 = \| \mathcal{E}[R, l] \| \) and define the following probability for some iteration \( L > 1 \):

\[
F_L[\beta] = P \left( Z_L \leq \sin(\beta) Z_{L-1} \leq \cdots \leq (\sin(\beta))^L Z_1 \right).
\]

\( F_L[\beta] \) can be viewed as the probability that residuals fall within at least one of the \( R \) white cones in every iteration \( l \leq L \).

The following proposition ensures a reduction of \( Z_1 \) by a factor smaller than \( (\sin(\beta))^{L-1} \) after \( L \) iterations with high probability, if a condition on the continuity of \( F_L[\beta] \) is assumed.

**Proposition 5.7:** Let \( L \) be fixed. If \( \exists \beta_0 : \beta_0 \in [0, \pi/2) \) such that \( F_L[\beta] \) is continuous on \( [\beta_0, \pi/2) \), then \( \forall \varepsilon : \varepsilon \in (0, 1], \exists \beta \in [\beta_0, \pi/2) \) such that \( F_L[\beta] > 1 - \varepsilon \).

*Proof:* Since \( F_L[\pi/2] = 1 \) and \( F_L[\beta] \) is continuous on \( [\beta_0, \pi/2] \), the proof follows directly from the intermediate value theorem.

Although \( Z_l, 1 \leq l \leq L \), are absolutely continuous random variables and \( g_m(\beta) = \sin^m(\beta) \) are continuous functions for all \( m \geq 0 \), the continuity of \( F_L[\beta] \) in \( \beta \) is not guaranteed due to the dependence of the random variables \( Z_l, \ldots, Z_L \) (\( Z_l \) depends on \( Z_{l-1} \)). For example, for \( L = 2 \) and \( Z_1 = 2\pi/2 \) with probability 1, it is easy to check that \( F_2[\beta] \) is not continuous at \( \beta = \pi/2 \). Indeed, \( \lim_{\beta \rightarrow \pi/2} F_2[\beta] = 0 \) whereas \( F_2[\pi/2] = 1 \).

The following conjecture claims that there exists absolutely continuous distributions of tensors in \( \mathcal{T}^{(R)} \) such that the probability \( F_l[\beta(l)] \) tends to 1 for some function \( \beta(l) \) as \( l \rightarrow \infty \), and at the same time the norm of residuals tends to 0, which is suitable for the convergence of the DCPD algorithm to an exact CP decomposition.

**Conjecture 5.8:** There exists at least one absolutely continuous probability measure \( \mu \) for tensors \( \mathcal{T} \) within \( \mathcal{T}^{(R)} \) for which the following holds:

(i) \( \forall \varepsilon : \varepsilon \in (0, 1], \forall l > 1, \exists \beta : \beta \in [0, \pi/2) \), such that \( F_l[\beta] > 1 - \varepsilon \).

(ii) \( \forall l > 1, \exists \beta : \beta \in [0, \pi/2) \), such that \( (\sin(\beta))^{l-1} \) is a strictly monotonically decreasing sequence converging to 0.

Subsequent computer simulations support the existence of a uniform probability measure \( \mu \) for the entries of tensors within \( \mathcal{T}^{(R)} \), such that \( F_2[\beta] \approx 1 \) for large values of \( L \), and \( \| \mathcal{E}[R, L] \| \approx 0 \). This reinforces our conjecture.

### VI. Computer Results

#### A. Comparison between THOSVD and SeROAP

In this section, we compare the performance of rank-1 approximation methods SeROAP and THOSVD for different three-way tensor scenarios. For each case, 300 complex tensors whose real and imaginary parts are uniformly distributed in \([-1, 1]\) were generated. Figure 2 presents the difference between the Frobenius norms of the residuals computed as

\[
\Delta \phi = \| \mathcal{T} - \phi^{TH} (\mathcal{T}) \| - \| \mathcal{T} - \phi^{Se} (\mathcal{T}) \|.
\]

![Fig. 2. THOSVD and SeROAP comparison](image)

We note that \( \Delta \phi > 0 \) in all scenarios, as predicted by Proposition 4.1.

#### B. Performance of rank-1 approximations

The tables below compare different rank-1 approximation methods with respect to the best rank-1 approximation, which was obtained from the algebraic geometric moment method described in (28). Because the latter is infeasible to compute for high dimensions, we have focused on \( 2 \times 2 \times 2 \) and \( 3 \times 3 \times 3 \) real tensors.

The two iterative methods, namely ALS and CE, are initialized by the result obtained from SeROAP. A sample of 200 real tensors uniformly distributed in \([-1, 1]\) were generated for each of both scenarios. For comparison, we consider the MSE metric given by

\[
\text{MSE} = \frac{1}{200} \sum_{n=1}^{200} (\Delta \phi_{(n)}^{m})^2,
\]
where $\Delta m^{(n)} = \| T^{(n)} - \phi_m(T^{(n)}) \| - \| T^{(n)} - \phi^*(T^{(n)}) \|$. $\phi_m(T^{(n)})$ and $\phi^*(T^{(n)})$ are the rank-1 approximation for algorithm $m$ and the best rank-1 approximation of $T^{(n)}$, respectively.

| Algorithm   | MSE       | Mean iteration |
|-------------|-----------|----------------|
| THOSVD      | 0.02299   | N/A            |
| SeROAP      | 4.36155e-4| N/A            |
| CE          | 1.13056e-18| 6.115         |
| ALS         | 1.03406e-13| 6.135         |

The results show that SeROAP is a better rank-1 approximation than THOSVD as expected. For $2 \times 2 \times 2$ tensors, CE attains the best rank-1 approximation. In both scenarios, ALS and CE converge approximately in the same number of iterations.

C. Percentage of successful decompositions

Figure 3 presents the percentage of successful decompositions of rank-3 tensors for the algorithms ALS, CG with ELS, and DCPD. The ALS and the CG algorithms were randomly initialized. We have simulated DCPD with the algebraic methods THOSVD and SeROAP. Noise is not considered in this case so that the performance is evaluated for the computation of an exact decomposition of 300 tensors. We consider that a decomposition is succeeded if the residual $\| E \| \leq 10^{-5}$.

![Figure 3: Percentage of successful decomposition for rank-3 tensors.](image)

We note that the DCPD algorithm combined with SeROAP always presents a better performance than the standard ALS algorithm. Moreover, for higher dimensions, the percentage of successful decompositions is almost 100% for DCPD-SeROAP, which is a remarkable result, bearing in mind that the objective is multimodal.

D. Convergence rate

Figure 4 presents the performance of the algorithms in terms of the average of $\| E[R, l] \|$ per iteration for different values of the signal-to-noise (SNR) ratio for $5 \times 5 \times 5$-rank-3 tensors.

![Figure 4: Mean of $\| E[R, l] \|$ per iteration for different values of SNR.](image)

Again, ALS and CG algorithms were randomly initialized. Additive Gaussian noise is considered in our simulations.

We note in this figure that DCPD-SeROAP converges more quickly than the other algorithms. For an SNR of 40 dB, DCPD-SeROAP attains $\| E[R, l] \| \approx 0.01$ in approximately 100 iterations while, for the other algorithms, $\| E[R, l] \| > 0.02$ for the same number of iterations. Similar results are observed for other SNRs. The figure also shows that performances become similar when the SNR is decreased.

E. Residual vs rank

Now, we compare the algorithms for two SNRs by varying the rank of $8 \times 8 \times 8$ tensors. Again, we note the better performance of DCPD-SeROAP over the competing algorithms. Figure 5 also shows that the combination of DCPD and THOSVD yields the worst results. This is expected because the rank-1 approximation obtained by THOSVD is not good enough, so that DCPD does not come at a small residual.

![Figure 5: Mean of $\| E[l] \|$ under rank variation.](image)

VII. CONCLUSION

In this paper, we presented some CP tensor decompositions algorithms and provided an analysis of their compu-
tional complexities. Our contributions included: (i) a new algebraic rank-1 method, namely SeROAP, performing better than THOSVD for three-way tensors; (ii) an iterative rank-1 approximation algorithm, namely CE, that refines any rank-
1 approximation method, such as SeROAP and THOSVD, which converges in very few iterations; and (iii) an analysis of the convergence of the DCPD algorithm under a geometric point of view. Several computer experiments have confirmed the theoretical results.

APPENDIX A
COMPUTATION OF RANK-1 APPROXIMATION USING
SeROAP

We present an example of how SeROAP algorithm works for computing a rank-1 approximation of a given tensor.

Let $\mathcal{T}$ be a $2 \times 2 \times 2$ complex tensor whose mode-1 unfolding is given by

$$
T^{(1)} = \begin{bmatrix}
1 & -1 & 0 & 1 & 3 & i & 0 & 1 \\
0 & 1 & -i & 1 & 1 & 0 & 2 & -2i
\end{bmatrix}.
$$

In the first for loop of SeROAP algorithm, $v_1$ is the dominate right singular vector of $V = V_0 = T^{(1)}$ given by

$$
v_1 = \begin{bmatrix}
-0.1717 - 0.0914i \\
0.0245 + 0.1060i \\
-0.0146 - 0.1472i \\
-0.3189 - 0.0768i \\
-0.6624 - 0.2596i \\
-0.0914 + 0.1717i \\
-0.2944 + 0.0292i \\
-0.2010 - 0.3858i
\end{bmatrix}.
$$

By reshaping $v_1$ in a $2 \times 4$ matrix $V_1$ we have

$$
V_1 = \begin{bmatrix}
-0.1717 - 0.0914i & -0.0146 - 0.1472i & 0.0245 + 0.1060i & -0.3189 - 0.0768i \\
-0.6624 - 0.2596i & -0.2944 + 0.0292i & -0.0914 + 0.1717i & -0.2010 - 0.3858i
\end{bmatrix}.
$$

In next iteration ($n = 2$), we compute the dominate right singular vector of $V = V_1$. Hence,

$$
v_2 = \begin{bmatrix}
-0.1654 + 0.1657i \\
0.0611 + 0.2758i \\
-0.6190 + 0.6261i \\
-0.2033 + 0.2210i
\end{bmatrix},
$$

and $V = V_2 = U \text{vec}(v_2)$ is updated

$$
V = \begin{bmatrix}
-0.1654 + 0.1657i & -0.6190 + 0.6261i \\
0.0611 + 0.2758i & -0.2033 + 0.2210i
\end{bmatrix}.
$$

The next step is to compute the vector $w = v^* \bar{u}$, where $u$ and $v$ are the first left and right singular vectors, respectively. Thus,

$$
u = \begin{bmatrix}
0.6106 - 0.7024i \\
0.1758 - 0.3208i
\end{bmatrix},
$$

$$
v = \begin{bmatrix}
-0.3027 - 0.0545i \\
-0.9481 + 0.0809i
\end{bmatrix},
$$

$$
w = \begin{bmatrix}
-0.1466 + 0.2459i \\
-0.0357 + 0.1067i \\
-0.6357 + 0.6166i \\
-0.1926 + 0.2899i
\end{bmatrix}.
$$

Notice that $w$ can be viewed as a vectorization of a rank-
1 matrix. In the end of the first iteration of the second for loop, $w$ is updated so that it becomes a vectorization of a rank-
1 three-way tensor. This is achieved by performing the projection of each row of matrix $V_1$ onto $w$, by means of $V_1 w$, following by the multiplication with $w^H$. Hence

$$
X^{(2)} = \begin{bmatrix}
-0.1900 - 0.1178i & -0.0631 - 0.0611i \\
-0.0604 + 0.0051i & -0.0236 - 0.0031i \\
-0.6624 - 0.1989i & -0.2377 - 0.1317i \\
-0.1762 + 0.0638i & -0.0730 + 0.0098i
\end{bmatrix},
$$

and

$$
w = \begin{bmatrix}
-0.1900 - 0.1178i \\
-0.0604 + 0.0051i \\
-0.0631 - 0.0611i \\
-0.0236 - 0.0031i \\
-0.6624 - 0.1989i \\
-0.2377 - 0.1317i \\
-0.1762 + 0.0638i \\
-0.0730 + 0.0098i
\end{bmatrix}.
$$

In the next iteration, we perform $X_1 = (V_0)w^H$, which is an unfolding of a 4-order rank-1 tensor. Actually, in every iteration of the second for loop, $X_n$ is updated to an unfolding of a rank-1 tensor of order $N - n + 1$. Hence,

$$
X^{(1)} = \begin{bmatrix}
0.5373 - 0.0992i & 0.1329 + 0.0653i \\
0.2696 - 0.1010i & 0.0750 + 0.0216i \\
0.1981 - 0.0830i & 0.0565 + 0.0140i \\
0.0951 - 0.0613i & 0.0306 + 0.0020i \\
1.6851 + 0.1360i & 0.3446 + 0.3020i \\
0.8868 - 0.0849i & 0.2066 + 0.1249i \\
0.6585 - 0.0886i & 0.1576 + 0.0872i \\
0.3335 - 0.1067i & 0.0898 + 0.0307i
\end{bmatrix},
$$

Indeed, $X^{(1)}$ is the mode-1 unfolding of the rank-1 approximation of $\mathcal{T}$ computed by SeROAP algorithm.

REFERENCES

[1] P. Comon and C. Jutten, Eds., Handbook of Blind Source Separation, Independent Component Analysis and Applications, Academic Press, Oxford UK, Burlington USA, 2010, ISBN: 978-0-12-374726-6, hal-00460653.

[2] M. Castella and P. Comon, “Blind separation of instantaneous mixtures of dependent sources,” in Independent Component Analysis and Signal Separation, pp. 9–16, Springer, 2007.

[3] A. L. F. de Almeida, G. Favier, and J. C. M. Mota, “Para4fac-based unified tensor modeling for wireless communication systems with application to blind multilayer equalization,” Signal Processing, vol. 87, no. 2, pp. 337–351, Feb. 2007.

[4] A. Smilde, R. Bro, and P. Geladi, Multi-way analysis: applications in the chemical sciences, John Wiley & Sons, 2005.

[5] H. Becker, L. Albera, P. Comon, M. Haardt, G. Birot, F. Wendling, M. Gavaret, C.-G. Bénar, and I. Merlet, “Eeg extended source localization: tensor-based vs. conventional methods,” NeuroImage, vol. 96, pp. 143–157, 2014.

[6] S. Sahnow and P. Comon, “Joint subspace estimation and localization,” IEEE Trans. on Signal Processing, vol. 63, no. 10, 2015.

[7] B. Savas, Algorithms in Data Mining using Matrix and Tensor Metho ds, Ph.D. thesis, Linköping Univ. Tech., 2008.

[8] H. A. L. Kiers, “Towards a standardized notation and terminology in multiway analysis,” J. Chemometrics, pp. 105–122, 2000.
[9] J. M. Ten Berge, “Kruskal’s polynomial for 2 × 2 × 2 arrays and a generalization to 2 × n × n arrays,” Psychometrika, vol. 56, no. 4, pp. 631–636, 1991.

[10] J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas, “Symmetric tensor decomposition,” Linear Algebra and its Applications, vol. 433, no. 11, pp. 1851–1872, 2010.

[11] L. De Lathauwer, “A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization,” SIAM Journal on Matrix Analysis and Applications, vol. 28, no. 3, pp. 642–666, 2006.

[12] L. Oeding and G. Ottaviani, “Eigenvectors of tensors and algorithms for waring decomposition,” Journal of Symbolic Computation, vol. 54, pp. 9–35, 2013.

[13] C. J. Hillar and L. H. Lim, “Most tensor problems are NP-hard,” Journal of the ACM (JACM), vol. 60, no. 6, pp. 45, 2013.

[14] 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, vol. 30, no. 3, pp. 1084–1127, 2008.

[15] L. Oeding and G. Ottaviani, “Tensor decompositions, alternating least squares and other tales,” Journal of Chemometrics, vol. 23, no. 7–8, pp. 393–405, 2009.

[16] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” SIAM Review, vol. 51, no. 3, pp. 455–500, 2009.

[17] M. Rajih, P. Comon, and R. A. Harshman, “Enhanced line search: A novel method to accelerate parafac,” SIAM Journal on Matrix Analysis and Applications, vol. 30, no. 3, pp. 1128–1147, 2008.

[18] L. Sorber, M. Van Barel, and L. De Lathauwer, “Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-(L × L × 1) terms, and a new generalization,” SIAM Journal on Optimization, vol. 23, no. 2, pp. 695–720, 2013.

[19] C. J. Hillar and L.-H. Lim, “Fast local algorithms for large scale nonnegative matrix and tensor factorizations,” IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. E92-A, no. 3, pp. 708–721, 2009.

[20] A. P. da Silva, P. Comon, and A. L. F. de Almeida, “An iterative deflation algorithm for exact CP tensor decomposition,” IEEE conference on Acoustics, Speech and Signal Processing, 2015.

[21] L. De Lathauwer, B. De Moor, and J. Vandewalle, “A multilinear singular value decomposition,” SIAM Journal on Matrix Analysis and Applications, vol. 21, no. 4, pp. 1253–1278, 2000.

[22] L. Wang and T. Chu, “On the global convergence of the alternating least squares method for rank-one approximation to generic tensors,” SIAM Journal on Optimization, vol. 11, no. 3, pp. 796–817, 2001.

[23] M. A. Bucero and B. Mourrain, “Border basis relaxation for polynomial optimization,” arXiv preprint arXiv:1404.5489, 2014.

[24] J. Nie and L. Wang, “Semidefinite relaxations for best rank-1 tensor approximations,” SIAM Journal on Matrix Analysis and Applications, vol. 35, no. 3, pp. 1155–1179, 2014.

[25] G. H. Golub and C. F. Van Loan, Matrix computations, vol. 3, JHU Press, 2012.

[26] E. Polak, Optimization: algorithms and consistent approximations, Springer-Verlag New York, Inc., 1997.

[27] P. Comon and G. H. Golub, “Tracking a few extreme singular values and vectors in signal processing,” Proceedings of the IEEE, vol. 78, no. 8, pp. 1327–1343, 1990.

[28] C. F. Van Loan and N. Pitsianis, Approximation with Kronecker products, Springer, 1993.