

An algebraic solution for the Candecomp/PARAFAC decomposition with circulant factors

José Henrique de Morais Goulart, Gérard Favier

► **To cite this version:**

José Henrique de Morais Goulart, Gérard Favier. An algebraic solution for the Candecomp/PARAFAC decomposition with circulant factors. 2014. hal-00967263v1

HAL Id: hal-00967263

<https://hal.archives-ouvertes.fr/hal-00967263v1>

Submitted on 28 Mar 2014 (v1), last revised 17 Mar 2015 (v2)

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.



INFORMATIQUE, SIGNAUX ET SYSTÈMES
DE SOPHIA ANTIPOLIS
UMR7271

An algebraic solution for the Candecomp/PARAFAC decomposition with circulant factors

José Henrique De Morais Goulart, Gérard Favier

Équipe SIGNAL

Rapport de Recherche
ISRN I3S/RR-2014-02-FR

Février 2014

An algebraic solution for the Candecomp/PARAFAC decomposition with circulant factors

José Henrique De Morais Goulart¹, Gérard Favier²

Équipe SIGNAL

ISRN I3S/RR-2014-02-FR

Février 2014 - 20 pages

Abstract: The Candecomp/PARAFAC decomposition (CPD) is an important mathematical tool used in several fields of application. Yet, its computation is usually performed with iterative methods which are subject to reaching local minima and to exhibiting slow convergence. In some practical contexts, the data tensors of interest admit decompositions constituted by matrix factors with particular structure. Often, such structure can be exploited for devising specialized algorithms with superior properties in comparison with general iterative methods. In this paper, we propose a novel approach for computing a CPD of a tensor whose factors are (possibly tall) circulant matrices. To this end, we exploit the algebraic structure of such tensor, showing that the elements of its frequency-domain counterpart satisfy homogeneous monomial equations in the eigenvalues of square circulant matrices associated with its factors, which we can therefore estimate by solving these equations. Then, we characterize the sets of solutions admitted by such equations under Kruskal's uniqueness condition. Simulation results are presented, validating our approach and showing that it can help avoiding typical disadvantages of iterative methods.

Key-words: Candecomp/PARAFAC decomposition ; canonical polyadic decomposition ; tensor decomposition ; circulant matrices ; homogeneous monomial equats.

¹ Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France – goulart@unice.fr

² Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France – favier@i3s.unice.fr

AN ALGEBRAIC SOLUTION FOR THE CANDECOMP/PARAFAC DECOMPOSITION WITH CIRCULANT FACTORS

J. H. DE M. GOULART*[†] AND G. FAVIER*

Abstract. The Candecomp/PARAFAC decomposition (CPD) is an important mathematical tool used in several fields of application. Yet, its computation is usually performed with iterative methods which are subject to reaching local minima and to exhibiting slow convergence. In some practical contexts, the data tensors of interest admit decompositions constituted by matrix factors with particular structure. Often, such structure can be exploited for devising specialized algorithms with superior properties in comparison with general iterative methods. In this paper, we propose a novel approach for computing a CPD of a tensor whose factors are (possibly tall) circulant matrices. To this end, we exploit the algebraic structure of such tensor, showing that the elements of its frequency-domain counterpart satisfy homogeneous monomial equations in the eigenvalues of square circulant matrices associated with its factors, which we can therefore estimate by solving these equations. Then, we characterize the sets of solutions admitted by such equations under Kruskal's uniqueness condition. Simulation results are presented, validating our approach and showing that it can help avoiding typical disadvantages of iterative methods.

Key words. Candecomp/PARAFAC decomposition, canonical polyadic decomposition, tensor decomposition, circulant matrices, homogeneous monomial equations

AMS subject classifications. 15A69

1. Introduction. The Candecomp/PARAFAC or canonical polyadic decomposition (CPD) [17, 5] is a mathematical tool of great utility in many practical applications [4, 20]. Basically, it consists of a generalization of bilinear decompositions of matrices to higher-order tensors. In addition, unlike its bilinear counterparts, it is essentially unique under mild conditions. Therefore, it is naturally well suited for the analysis of data sets constituted by observations of a function of multiple discrete indices, as encountered in signal processing [29, 14, 13, 15], data mining [2, 26] and biomedical engineering [1]; see [4, 20] for other examples.

The attractive properties of the CPD are, however, accompanied by difficulties. Namely, its computation is generally a difficult non-convex optimization problem and, furthermore, determining the minimal number of terms that are necessary to decompose a tensor, which corresponds to its rank, is NP-hard [18]. Notwithstanding these difficulties, there are several methods which try to fit a best rank- R approximate CPD model to the data tensor of interest [33, 20].

In the general case, when no assumption is made about the matrix factors involved in the decomposition, one usually resorts to iterative methods for estimating them. Among these methods, the so-called alternating least-squares (ALS), which was originally and independently proposed in [5, 17], is the most used. Another popular algorithm is the classical Levenberg-Marquardt optimization method for nonlinear least-squares problems [23, 25]. However, given the nature of the CPD computation problem, which is generally formulated as the minimization of a least-squares objective function, iterative methods are evidently subject to convergence towards local minima. Moreover, they often suffer from slow convergence speed, and therefore may require a large amount of computing time.

*Laboratoire I3S, Université de Nice Sophia-Antipolis, CNRS, 2000 route des Lucioles, Les Algorithmes/Euclide B, 06900 Sophia-Antipolis, France (goulart@i3s.unice.fr, favier@i3s.unice.fr).

[†]Sponsored by CNPq–Brazil (individual grant 245358/2012-9).

Conceivably, when the matrix factors of a CPD are known to have special structure, as, e.g., Toeplitz, circulant or Vandermonde, one could take this into account to formulate specialized algorithms which avoid the disadvantages of iterative methods. Indeed, such strategy has been followed in some works, motivated by the pertinence of structural constraints in several practical contexts. For instance, [19] proposes a non-iterative method for the estimation of a third-order CPD in which a banded circulant factor is involved. Similarly, [30] develops methods which can recover in a non-iterative fashion banded (and possibly Toeplitz or Hankel) factors of tensors having order up to four. Another recent example of a structurally constrained CPD estimation method is [31], where Vandermonde matrix factors are considered. As examples of applications where some structurally constrained CPD arises, we can mention the blind identification of single-input single-output (SISO) finite impulse response (FIR) channels via high-order statistics [15] and the estimation of parameters of Wiener-Hammerstein nonlinear systems from the kernels of equivalent Volterra systems [19]. Constrained structures have also been exploited for estimating other tensor models. For instance, in [12] the third-order core tensors of block Tucker models are characterized by matrix slices having Hankel and Vandermonde forms, which is taken into account for deriving a specialized parameter estimation algorithm.

In this paper, we develop a novel algebraic approach for computing the CPD of a hypercubic tensor whose matrix factors are (Toeplitz and) circulant. This approach relies on the resolution of a system of homogeneous monomial equations, which are directly obtained from the multidimensional Fourier transform of the tensor to be decomposed and follow from the well known Fourier eigenstructure property of circulant matrices. In the (ideal) noiseless setting, this yields an exact solution for the decomposition. In addition, our approach does not limit the number of degrees of freedom that each factor can possess, as long as they are circulant. This is in contrast to the strategy followed in [19, 30], which can estimate a circulant factor by linearly transforming a basis obtained via subspace analysis so that a matrix with the desired structure results, but applies only if additional constraints are imposed over the value of some of its elements or over the decomposition rank. Our approach, in its turn, does not need such additional assumptions for yielding an (essentially) unique solution. Instead, it requires only the fulfillment of the standard Kruskal's uniqueness condition [21, 28]. We note that, due to permutation ambiguity which is inherent to the CPD, our approach is as well valid for estimating factors with any structure that can be transformed into a Toeplitz-circulant one via a postmultiplication by a permutation matrix, as, e.g., a "Hankel-circulant" matrix. As another contribution, we formulate also a version of the ALS algorithm which constrains the estimated factors of a CPD to be circulant, and thus is more appropriate as a reference for comparison with our algebraic approach.

The contents of this paper are organized as follows. Section 2 briefly reviews important concepts and definitions related to tensors, the CPD, its basic properties and the main existing methods for its computation. Then, the proposed approach is developed in §3, after recalling the fundamental property of circulant matrices on which it is based. Some simple illustrative examples of the application of our approach are provided in §4, for enhancing the understanding of the developed analysis. Next, §5 presents and discusses some simulation results through which the proposed approach is compared with several other methods. Finally, concluding remarks are given in §6.

Notational conventions: Matrices and column vectors are denoted by boldface capital and boldface lowercase letters, respectively. The columns of a $M \times R$ matrix \mathbf{A}

are denoted by $\mathbf{a}_1, \dots, \mathbf{a}_R$, and it is said to be tall if $M > R$. We denote the transpose, Hermitian transpose and Moore-Penrose pseudoinverse of \mathbf{A} by, respectively, \mathbf{A}^T , \mathbf{A}^H and \mathbf{A}^\dagger , and the $R \times R$ identity matrix by \mathbf{I}_R . The operator $\text{vec}(\cdot)$ maps $\mathbf{A} \in \mathbb{C}^{M \times R}$ into $\mathbf{a} = [\mathbf{a}_1^T \ \dots \ \mathbf{a}_R^T]^T \in \mathbb{C}^{RM}$, and $\text{diag}(\cdot)$ maps a vector $\mathbf{v} \in \mathbb{C}^M$ into an $M \times M$ matrix which contains its elements on the main diagonal and zeros everywhere else. Given $\mathbf{v}, \mathbf{w} \in \mathbb{C}^M$, the Hadamard product $\mathbf{v} \odot \mathbf{w} \in \mathbb{C}^M$ yields their element-wise multiplication. The Kronecker product between matrices $\mathbf{A} \in \mathbb{C}^{M_1 \times R_1}$ and $\mathbf{B} \in \mathbb{C}^{M_2 \times R_2}$ is denoted by $\mathbf{A} \otimes \mathbf{B} \in \mathbb{C}^{M_1 M_2 \times R_1 R_2}$, while their Khatri-Rao (i.e., columnwise Kronecker) product, which is only defined if $R_1 = R_2 = R$, is denoted by $\mathbf{A} \diamond \mathbf{B} \in \mathbb{C}^{M_1 M_2 \times R}$. We use δ_{i_1, \dots, i_P} to denote the generalized Kronecker delta, i.e., $\delta_{i_1, \dots, i_P} = 1$ if $i_1 = i_2 = \dots = i_P$ and $\delta_{i_1, \dots, i_P} = 0$ otherwise.

2. Review of the CP decomposition.

2.1. Basic definitions. Formally, a tensor of order P is an element of a tensor product among P vector spaces [24]. However, once a certain basis is chosen for each of these spaces, a tensor can be represented by a multidimensional array of numbers from a field (usually \mathbb{R} or \mathbb{C}). We denote a P th order tensor of dimensions $M_1 \times \dots \times M_P$ over \mathbb{C} by a capital calligraphic letter \mathcal{T} , and its corresponding space by $\mathbb{C}^{M_1 \times \dots \times M_P}$. Its elements are denoted by lowercase letters, t_{m_1, \dots, m_P} , having indices m_1, \dots, m_P which satisfy $m_j \in \{1, \dots, M_j\}$, or, alternatively, by the notation $[\mathcal{T}]_{m_1, \dots, m_P}$.

Each geometric dimension of a tensor is called a mode. Given a tensor $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$, for each mode p we can arrange its elements in a matrix unfolding $\mathbf{T}_p \in \mathbb{C}^{M_p \times M_{p+1} \dots M_{p-1} \dots M_1}$ with elements $[\mathbf{T}_p]_{m_p, j} = t_{m_1, \dots, m_P}$, where [20]

$$j = 1 + \sum_{\substack{q=1 \\ q \neq p}}^P (m_q - 1) \prod_{\substack{r=1 \\ r \neq p}}^{q-1} M_r.$$

A P th order tensor \mathcal{T} whose all modes have the same dimension M is called hypercubic and is denoted by $\mathcal{T} \in \mathbb{C}^{M^{(P)}}$, where $M^{(P)}$ is a shorthand for $M \times \dots \times M$, with P occurrences of M . A hypercubic tensor $\mathcal{T} \in \mathbb{C}^{M^{(P)}}$ is said to be symmetric whenever its elements are invariant to any permutation $\pi(m) = (\pi_1, \dots, \pi_P)$ of their indices $m = (m_1, \dots, m_P)$, i.e., when $t_{m_1, \dots, m_P} = t_{\pi_1, \dots, \pi_P}$. The n -mode product [6] of $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$ by a matrix $\mathbf{A} \in \mathbb{C}^{K \times M_n}$, denoted by $\mathcal{T} \times_n \mathbf{A}$, yields a $M_1 \times \dots \times M_{n-1} \times K \times M_{n+1} \times \dots \times M_P$ tensor such that

$$(2.1) \quad [\mathcal{T} \times_n \mathbf{A}]_{m_1, \dots, m_{n-1}, k, m_{n+1}, \dots, m_P} = \sum_{m_n=1}^{M_n} a_{k, m_n} t_{m_1, \dots, m_n, \dots, m_P}.$$

Another important operator is the outer product (or tensor product [24]), which is defined as follows. Let $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$ and $\mathcal{S} \in \mathbb{C}^{K_1 \times \dots \times K_Q}$ be tensors of orders P and Q , respectively. The outer product between them, denoted by $\mathcal{T} \circ \mathcal{S}$, yields $\mathcal{V} \in \mathbb{C}^{M_1 \times \dots \times M_P \times K_1 \times \dots \times K_Q}$ (of order $P + Q$) and can be expressed in scalar form as

$$v_{m_1, \dots, m_P, k_1, \dots, k_Q} = t_{m_1, \dots, m_P} s_{k_1, \dots, k_Q}.$$

Note that this definition includes as a particular case the well known outer product between two vectors, which yields a matrix.

Let $\omega_M \triangleq \exp(j2\pi/M)$ and \mathbf{F}_M be the $M \times M$ Fourier matrix, given by

$$(2.2) \quad \mathbf{F}_M = \frac{1}{\sqrt{M}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega_M & \omega_M^2 & \dots & \omega_M^{M-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \omega_M^{M-2} & \omega_M^{M-4} & \dots & \omega_M^2 \\ 1 & \omega_M^{M-1} & \omega_M^{M-2} & \dots & \omega_M \end{bmatrix}.$$

The multidimensional discrete Fourier transform (MDFT) of $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$, which we denote by $\mathcal{Y} = \text{MDFT}\{\mathcal{T}\}$, is defined as

$$\mathcal{Y} = \mathcal{T} \times_1 \mathbf{F}_{M_1}^H \times_2 \mathbf{F}_{M_2}^H \cdots \times_P \mathbf{F}_{M_P}^H \in \mathbb{C}^{M_1 \times \dots \times M_P}.$$

In scalar form, we have

$$y_{k_1, \dots, k_P} = \left(\prod_{p=1}^P \frac{1}{\sqrt{M_p}} \right) \sum_{m_1=1}^{M_1} \cdots \sum_{m_P=1}^{M_P} t_{m_1, \dots, m_P} \omega_{M_1}^{-(m_1-1)(k_1-1)} \cdots \omega_{M_P}^{-(m_P-1)(k_P-1)},$$

for $k_p \in \{1, \dots, M_p\}$, $p \in \{1, \dots, P\}$. Analogously, we define the inverse multidimensional discrete Fourier transform (IMDFT) of $\mathcal{Y} \in \mathbb{C}^{M_1 \times \dots \times M_P}$ as

$$\mathcal{T} = \mathcal{Y} \times_1 \mathbf{F}_{M_1} \times_2 \mathbf{F}_{M_2} \cdots \times_P \mathbf{F}_{M_P} \in \mathbb{C}^{M_1 \times \dots \times M_P},$$

and denote it by $\mathcal{T} = \text{IMDFT}\{\mathcal{Y}\}$. In scalar form, this relation can be written as

$$t_{m_1, \dots, m_P} = \left(\prod_{p=1}^P \frac{1}{\sqrt{M_p}} \right) \sum_{k_1=1}^{M_1} \cdots \sum_{k_P=1}^{M_P} y_{k_1, \dots, k_P} \omega_{M_1}^{(m_1-1)(k_1-1)} \cdots \omega_{M_P}^{(m_P-1)(k_P-1)},$$

with $m_p \in \{1, \dots, M_p\}$, $p \in \{1, \dots, P\}$.

2.2. The CP decomposition. The CP decomposition (CPD) or CP model of a tensor $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$ is defined by

$$(2.3) \quad \mathcal{T} = \sum_{r=1}^R \mathbf{a}_r^{(1)} \circ \cdots \circ \mathbf{a}_r^{(P)},$$

where $\mathbf{a}_r^{(p)}$ is the r -th column of the matrix factor $\mathbf{A}^{(p)} \in \mathbb{C}^{M_p \times R}$. The minimal value of R such that \mathcal{T} can be written as in (2.3) is called the rank of \mathcal{T} . It is clear from (2.3) that the CPD can be seen as a generalization of the decomposition of a matrix into a sum of rank-one matrices; in fact, a P th order tensor is said to have rank one precisely when it can be written as an outer product of P vectors. Also, as in the matrix case, every tensor admits a decomposition of the form (2.3). Using the definition of the outer product, we conclude that the CPD can be expressed in scalar form as

$$(2.4) \quad t_{m_1, \dots, m_P} = \sum_{r=1}^R \prod_{p=1}^P a_{m_p, r}^{(p)},$$

where $a_{m, r}^{(p)}$ is the m -th element of $\mathbf{a}_r^{(p)}$. By defining the P th order identity tensor $\mathcal{I} \in \mathbb{C}^{R \times \dots \times R}$, whose elements satisfy $i_{r_1, \dots, r_P} = \delta_{r_1, \dots, r_P}$, from (2.4) we can deduce that

$$(2.5) \quad \mathcal{T} = \mathcal{I} \times_1 \mathbf{A}^{(1)} \times_2 \cdots \times_P \mathbf{A}^{(P)}.$$

Here, we will also use the convenient Kruskal's notation [21]

$$(2.6) \quad \mathcal{T} = \llbracket \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(P)} \rrbracket.$$

In relation with the n -mode product, an important property of the CPD is that for any $\mathbf{B} \in \mathbb{C}^{K \times M_n}$ we have

$$(2.7) \quad \mathcal{T} \times_n \mathbf{B} = \llbracket \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(n-1)}, \mathbf{B}\mathbf{A}^{(n)}, \mathbf{A}^{(n+1)}, \dots, \mathbf{A}^{(P)} \rrbracket,$$

which can be deduced from (2.5) and the definition of the n -mode product. Another useful fact is that the unfolding $\mathbf{T}_p \in \mathbb{C}^{M_p \times M_P \dots M_{p+1} M_{p-1} \dots M_1}$ of \mathcal{T} can be written as

$$(2.8) \quad \mathbf{T}_p = \mathbf{A}^{(p)} \left(\mathbf{A}^{(P)} \diamond \dots \diamond \mathbf{A}^{(p+1)} \diamond \mathbf{A}^{(p-1)} \diamond \dots \diamond \mathbf{A}^{(1)} \right)^T.$$

Finally, if a CPD of a symmetric tensor \mathcal{S} is constituted by identical factors, i.e.,

$$(2.9) \quad \mathcal{S} = \llbracket \mathbf{A}, \dots, \mathbf{A} \rrbracket \in \mathbb{C}^{M^{(P)}},$$

where $\mathbf{A} \in \mathbb{C}^{M \times R_s}$, it is called a symmetric CPD of \mathcal{S} [7, 3]. If R_s is the minimal value for which (2.9) holds, it is called the symmetric rank of \mathcal{S} .

2.3. Uniqueness of the CPD. The CPD (2.6) is said to be essentially unique if any other CPD of \mathcal{T} with factors $\mathbf{B}^{(p)} \in \mathbb{C}^{M_p \times R}$, i.e., $\mathcal{T} = \llbracket \mathbf{B}^{(1)}, \dots, \mathbf{B}^{(P)} \rrbracket$, is such that $\mathbf{B}^{(p)} = \mathbf{A}^{(p)} \mathbf{\Pi} \mathbf{\Delta}_p$, where $\mathbf{\Pi} \in \mathbb{R}^{R \times R}$ is a permutation matrix and the matrices $\mathbf{\Delta}_p \in \mathbb{C}^{R \times R}$, $p \in \{1, \dots, P\}$, are diagonal and satisfy $\mathbf{\Delta}_1 \dots \mathbf{\Delta}_P = \mathbf{I}_R$. Basically, this definition allows us to disregard the trivial column permutation and scaling ambiguities that are inherent to the model. With it at hand, we can state a fundamental result on the uniqueness of the CPD, which was first established by Kruskal [21] for third-order tensors and then generalized by Sidiropoulos and Bro [28] for any order. Before, however, another concept must be introduced: the k -rank (named after Kruskal) of a matrix \mathbf{A} , denoted by $k_{\mathbf{A}}$, is the largest integer such that every set containing $k_{\mathbf{A}}$ columns of \mathbf{A} is linearly independent [21].

THEOREM 2.1 ([28]). *Let $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$ be given by (2.6), where $\mathbf{A}^{(p)} \in \mathbb{C}^{M_p \times R}$, with $P > 2$ and $R > 1$. If the matrix factors $\mathbf{A}^{(p)}$ are such that their k -ranks satisfy $k_{\mathbf{A}^{(1)}} + \dots + k_{\mathbf{A}^{(P)}} \geq 2R + P - 1$, then the CPD is essentially unique.*

It should be noted that the condition involved in Theorem 2.1 is sufficient but not necessary. Note also that, unlike matrix decompositions, whose uniqueness can only be established by imposing special constraints—such as, e.g., orthogonality—on their factors, the CPD requires much less constraining assumptions.

2.4. CPD computation. Given a data tensor $\mathcal{D} \in \mathbb{C}^{M_1 \times \dots \times M_P}$, the computation of a CPD of \mathcal{D} is generally stated as searching for factors $\mathbf{A}^{(p)} \in \mathbb{C}^{M_p \times R}$, $p \in \{1, \dots, P\}$, that jointly minimize an error criterion, which is often formulated as

$$(2.10) \quad J(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(P)}) = \left\| \mathcal{D} - \mathcal{T}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(P)}) \right\|_F^2,$$

where $\mathcal{T}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(P)})$ is given by (2.6) and $\|\cdot\|_F$ stands for the Frobenius norm, defined as $\|\mathcal{T}\|_F \triangleq \sqrt{\text{vec}(\mathbf{T}_1)^H \text{vec}(\mathbf{T}_1)}$. The choice of R is clearly critical, but in practice it is usually unknown. Furthermore, computing R is NP-hard [18]. Consequently, a CPD is often computed by searching for the best rank- R approximation of \mathcal{T} , given a choice of R . Yet, this problem is ill-posed in general, because a tensor

may not have a best rank- R approximation [10]. In that case, one can find factors which yield an arbitrarily low approximation error, although their columns contain meaningless elements with very large magnitude [27, 32].

The cost function (2.10) is clearly nonlinear in the parameters, and thus the solution is, in general, computed iteratively. To perform this task, the most commonly used algorithm is the alternating least-squares (ALS) [17]. Basically, it consists in sequentially updating the estimate of each factor, fixing the other ones to their previous estimates. Hence, the original nonlinear problem is tackled by considering a sequence of linear ones, which can be solved using the classical least-squares estimator. Although this strategy is conceptually simple and has led to useful results in many applications, there is no proof of its convergence. Nonlinear optimization schemes, such as the Levenberg-Marquardt (LM) method, are also often employed [8, 33]. Unlike the ALS approach, techniques of this nature rely on a solid mathematical framework, but may also lead to local minima and/or converge very slowly.

2.4.1. Previous approaches for estimating circulant factors. In [19], a non-iterative algorithm which decomposes third-order tensors having one circulant factor is proposed, whose extension to higher order tensors is trivial.¹ Basically, a matrix whose columns form a basis for the column space of such factor is obtained from a SVD, and then a linear transformation capable of mapping this matrix to another with the desired structure is found through the resolution of a system of linear equations. Then, the other factors can be estimated by solving a number of best rank-one tensor approximations,² which for $P = 3$ can be performed non-iteratively by computing SVDs. A drawback of this method, however, is that the circulant factor is constrained to be tall and banded in a particular way.

In [9, 30], that strategy is generalized to compute a CPD constituted by banded and/or structured factors. In particular, both papers offer procedures for estimating Toeplitz factors, which can thus be used when they are circulant. These procedures are non-iterative for $P = 3$ and also in some cases where $P = 4$. Yet, the factors must satisfy additional constraints, as having a banded structure and/or being tall, which essentially stems from the subspace-based strategy employed. For instance, one cannot estimate square and non-banded circulant factors with such procedures.

It should be noted that, when the additional constraints imposed in [19, 9, 30] apply, these methods are attractive for having good numerical properties, since they rely on the resolution of (generally overdetermined) systems of linear equations. Indeed, the simulation results shown in these works indicate that they can outperform the standard ALS both in terms of efficiency and of precision. Furthermore, when all factors are circulant, a repeated application of those methods, once for each mode, enables us to estimate the factors non-iteratively, regardless of the order P .

3. Algebraic solution for a circulant-constrained CPD. In the following, we provide a formulation of the problem addressed in this work, namely, the estimation of a CP model involving only circulant matrix factors, which will be referred to as circulant-constrained CPD (CCPD). Then, we describe the proposed approach, which is based on the resolution of a system of monomial equations that can be obtained through a multilinear transformation of the tensor to be decomposed, provided that its

¹For $P > 3$, the recovery of the other factors requires computing the best rank-one approximation of R tensors, because of the Khatri-Rao structure of the matrix from which they are obtained [30].

²It should be noted that the rank-one approximation problem is well posed, as opposed to finding the best rank- R approximation of a tensor [10].

rank is known *a priori*. After that, we study the solutions admitted by such equations. Finally, we specialize our analysis to the case where the CCPD is symmetric.

First of all, however, it is important to briefly review some properties of circulant matrices which play a major role in our approach.

DEFINITION 3.1. A matrix $\mathbf{C} \in \mathbb{C}^{M \times R}$ is said to be circulant if it can be written as

$$(3.1) \quad \mathbf{C} = [\mathbf{c} \quad \mathbf{\Pi}_M \mathbf{c} \quad \dots \quad \mathbf{\Pi}_M^{R-1} \mathbf{c}],$$

where $\mathbf{c} = [c_0 \quad c_1 \quad \dots \quad c_{M-1}]^T \in \mathbb{C}^M$ is called its generating vector and $\mathbf{\Pi}_M$ is the following $M \times M$ permutation matrix:

$$(3.2) \quad \mathbf{\Pi}_M = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix}.$$

Although no restriction has been placed over the dimensions M and R , in the sequel we consider only square and tall matrices, as will be explained in §3.1. Square circulant matrices possess a very special property that is well known: they all share the same eigenvectors, which are simply the columns of the Fourier matrix [16]. Since these columns form an orthogonal basis for \mathbb{C}^M , it follows that any circulant matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$ can be diagonalized by the Fourier matrix, or, equivalently,

$$(3.3) \quad \mathbf{C} = \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H,$$

where \mathbf{F}_M is defined in (2.2) and $\mathbf{\Lambda}$ is a $M \times M$ diagonal matrix containing the eigenvalues of \mathbf{C} . Another direct consequence of this particular eigenstructure is that, if we rewrite (3.3) as $\mathbf{F}_M^H \mathbf{C} = \mathbf{\Lambda} \mathbf{F}_M^H$, then from the first column of the latter identity we have that $\mathbf{F}_M^H \mathbf{c} = \boldsymbol{\lambda} / \sqrt{M}$, where $\boldsymbol{\lambda}$ is such that $\text{diag}(\boldsymbol{\lambda}) = \mathbf{\Lambda}$. In other words, \mathbf{c} and $\boldsymbol{\lambda} / \sqrt{M}$ form a DFT pair. As property (3.3) applies only for square matrices, we now introduce a definition which will facilitate the treatment of tall circulant factors.

DEFINITION 3.2. Let $\mathbf{C} \in \mathbb{C}^{M \times R}$ be a circulant matrix with $M \geq R$, and \mathbf{c} be its generating vector. We define the circulant completion of \mathbf{C} as the square matrix $\check{\mathbf{C}} \in \mathbb{C}^{M \times M}$ given by

$$\check{\mathbf{C}} = \begin{cases} [\mathbf{C} \quad \mathbf{\Pi}_M^R \mathbf{c} \quad \mathbf{\Pi}_M^{R+1} \mathbf{c} \quad \dots \quad \mathbf{\Pi}_M^{M-1} \mathbf{c}], & \text{if } M > R \\ \mathbf{C}, & \text{if } M = R, \end{cases}$$

where $\mathbf{\Pi}_M$ is as defined by (3.2).

From the above definition, it follows that any $M \times R$ circulant matrix with $M \geq R$ is such that its circulant completion can be decomposed as in (3.3).

By noting that $\mathbf{\Pi}_M$ itself is circulant, we can also state the following useful property, whose proof will be omitted due to space reasons.

PROPOSITION 3.3. Let $\mathbf{\Pi}_M$ be as defined by (3.2). Then, $\mathbf{\Pi}_M^k = \mathbf{\Pi}_M^{(k)_M}$ for all $k \in \mathbb{Z}$, where $(\cdot)_M$ denotes the modulo M operator.³ Consequently, there exist only M distinct matrices of the form $\mathbf{\Pi}_M^k$, namely, $\mathbf{\Pi}_M^0, \mathbf{\Pi}_M, \dots, \mathbf{\Pi}_M^{M-1}$. Moreover, $\mathbf{\Pi}_M^k$ admits the eigendecomposition

$$(3.4) \quad \mathbf{\Pi}_M^k = \mathbf{F}_M \text{diag} \left(1, \omega_M^{-k}, \dots, \omega_M^{-k(M-1)} \right) \mathbf{F}_M^H.$$

³ $\forall k \in \mathbb{Z}, M \in \mathbb{N}, (k)_M = m \Leftrightarrow k = lM + m$ for some $l \in \mathbb{Z}$, with $0 \leq m < M$.

3.1. General CCPD. Let \mathcal{T} be a hypercubic P th order tensor of dimensions $M \times \cdots \times M$. Suppose that \mathcal{T} admits a CCPD, i.e., that there exist P complex circulant matrices $\mathbf{C}^{(p)} \in \mathbb{C}^{M \times R}$, $p \in \{1, \dots, P\}$, such that \mathcal{T} can be written as

$$(3.5) \quad \mathcal{T} = \llbracket \mathbf{C}^{(1)}, \dots, \mathbf{C}^{(P)} \rrbracket \in \mathbb{C}^{M^{(P)}}.$$

Our goal is then to determine the circulant factors of (3.5), up to trivial ambiguities.

A distinguishing feature of the CCPD (3.5) is that its rank is upper bounded by M , as opposed to the general CPD (2.6), whose rank can exceed the dimensions of the tensor [20]. Such conclusion can be reached by noting that, if the factors $\mathbf{C}^{(p)} \in \mathbb{C}^{M \times R}$ in (3.5) are such that $R > M$, then from Definition 3.1 and Proposition 3.3 we have $\mathbf{c}_{M+m}^{(p)} = \mathbf{c}_m^{(p)}$ for all p and $m \geq 1$. Clearly, this ‘‘joint repetition’’ would imply the existence of linearly dependent terms in the decomposition, and therefore R would not be minimal. Because of this fact, in what follows we assume that $M \geq R$, which covers the range of all admissible ranks of the CCPD (3.5).

3.1.1. Derivation of the monomial equations. The first step in the derivation of our approach consists in showing how property (3.3) can lead to a set of monomial equations from a hypercubic tensor \mathcal{T} which is known to admit a CCPD. Taking the MDFT of \mathcal{T} , we have

$$(3.6) \quad \mathcal{Y} = \mathcal{T} \times_1 \mathbf{F}_M^H \times_2 \mathbf{F}_M^H \cdots \times_P \mathbf{F}_M^H \in \mathbb{C}^{M^{(P)}}.$$

Applying property (2.7), it can be deduced from (3.5) and (3.6) that

$$(3.7) \quad \mathcal{Y} = \llbracket \mathbf{F}_M^H \mathbf{C}^{(1)}, \dots, \mathbf{F}_M^H \mathbf{C}^{(P)} \rrbracket \in \mathbb{C}^{M^{(P)}}.$$

But, since each $\mathbf{C}^{(p)} \in \mathbb{C}^{M \times R}$ is a circulant matrix with $M \geq R$, we can decompose its circulant completion as $\check{\mathbf{C}}^{(p)} = \mathbf{F}_M \mathbf{\Lambda}_p \mathbf{F}_M^H$, where $\mathbf{\Lambda}_p = \text{diag}(\boldsymbol{\lambda}_p)$ is the diagonal matrix of eigenvalues of $\check{\mathbf{C}}^{(p)}$. Therefore,

$$(3.8) \quad \begin{aligned} \mathbf{F}_M^H \check{\mathbf{C}}^{(p)} &= \text{diag}(\boldsymbol{\lambda}_p) \mathbf{F}_M^H \\ &= [\boldsymbol{\lambda}_p \odot \mathbf{f}_{M,1}^* \quad \boldsymbol{\lambda}_p \odot \mathbf{f}_{M,2}^* \quad \cdots \quad \boldsymbol{\lambda}_p \odot \mathbf{f}_{M,M}^*], \end{aligned}$$

where $\mathbf{f}_{M,m}$ is the m -th column of $\mathbf{F}_M = \mathbf{F}_M^T$ and the $*$ superscript stands for complex conjugation. Thus, considering the first R columns of (3.8), we have

$$(3.9) \quad \mathbf{F}_M^H \mathbf{C}^{(p)} = [\boldsymbol{\lambda}_p \odot \mathbf{f}_{M,1}^* \quad \boldsymbol{\lambda}_p \odot \mathbf{f}_{M,2}^* \quad \cdots \quad \boldsymbol{\lambda}_p \odot \mathbf{f}_{M,R}^*].$$

In particular, as already mentioned, the first column of the identity (3.9) means that

$$(3.10) \quad \text{DFT} \left\{ \mathbf{c}^{(p)} \right\} = \frac{\boldsymbol{\lambda}_p}{\sqrt{M}} \iff \text{IDFT} \left\{ \frac{\boldsymbol{\lambda}_p}{\sqrt{M}} \right\} = \mathbf{c}^{(p)},$$

where $\mathbf{c}^{(p)}$ is the generating vector of $\mathbf{C}^{(p)}$.

Relations (3.7) and (3.9) provide the key for deriving the equations that we shall use to estimate the CCPD (3.5). More specifically, by substituting (3.9) into (3.7) and resorting to the scalar form (2.4) of the CPD, the elements of \mathcal{Y} are seen to satisfy

$$(3.11) \quad y_{m_1, \dots, m_P} = \left(\frac{1}{\sqrt{M}} \right)^P \prod_{p=1}^P \lambda_{p, m_p} \sum_{r=1}^R \omega_M^{-(r-1)(m_1 + \dots + m_P - P)},$$

where $\lambda_{p,m}$ is the m -th element of $\boldsymbol{\lambda}_p$. Hence, the elements of the frequency-domain counterpart of \mathcal{T} satisfy P th order monomial equations whose unknowns are the eigenvalues of the circulant completions of its matrix factors.

3.1.2. Characterization of the derived equations and of its solutions.

The set of equations (3.11) for $m_p \in \{1, \dots, M\}$, $p \in \{1, \dots, P\}$, might lead to a solution for the computation of the CCPD (3.5), if we can compute the eigenvalues of the circulant completion associated with each factor. To this end, it is important to first determine which equations of the form (3.11) are relevant, since the summation of complex exponentials in the right-hand side can vanish for some indices m_1, \dots, m_P , which clearly renders the corresponding equation irrelevant. Thus, the next result establishes a necessary and sufficient condition for the non-nullity of that summation.

PROPOSITION 3.4. *Let $M, R \in \mathbb{N}$ such that $M \geq R$ and $M > 1$, $m_1, \dots, m_P \in \{1, \dots, M\}$ and Q be defined as*

$$(3.12) \quad Q \triangleq \frac{M}{\gcd(M, m_1 + \dots + m_P - P)},$$

where $\gcd(\cdot, \cdot)$ yields the greatest common divisor of its arguments. Then, we have

$$(3.13) \quad \sum_{r=1}^R \omega_M^{-(r-1)(m_1 + \dots + m_P - P)} \neq 0$$

if and only if one of the following (mutually exclusive) conditions are met:

- (i) $Q = 1$;
- (ii) Q does not divide R .

Proof. Defining

$$v \triangleq \frac{m_1 + \dots + m_P - P}{\gcd(M, m_1 + \dots + m_P - P)} \in \mathbb{N}$$

and using (3.12), we can obtain from (3.13) the equivalent relation

$$(3.14) \quad \sum_{r=1}^R \omega_Q^{-(r-1)v} \neq 0.$$

From (3.14), it is now evident that (i) implies (3.13), since $\omega_1 = 1$. On the other hand, if (ii) is true, then we necessarily have $Q > 1$. In this case, (3.14) corresponds to a sum of R Q th roots of unity raised to a power v that is co-prime with Q by definition, which yields zero if and only if $R = lQ$ for some positive integer l . Hence, (ii) also implies (3.13). Note that the last argument also establishes the only if part of the proof, since in this part we have to show that the sum in (3.13) vanishes if $Q > 1$ and Q divides R . We observe also that, when these conditions hold, we have $l \in \{1, \dots, d\}$ where d is the greatest integer that divides M and satisfies $d < M$, since Q divides M and $R \leq M$. \square

PROPOSITION 3.5. *In the special case where $M = R$ (i.e., the factors are square), condition (i) is necessary and sufficient for (3.13) of Proposition 3.4. Moreover, we can write it alternatively as*

$$(3.15) \quad \exists l \in \mathbb{N} \text{ such that } m_1 + \dots + m_P - P = lM.$$

Proof. The condition (ii) of Proposition 3.4 is immediately ruled out when $M = R$, since Q divides M by definition. The fact that the condition (i) and (3.15) are equivalent is trivial, and thus the proof is complete. \square

We now focus on the characterization of the solutions to equations (3.11). First, we claim that there can be no “spurious” solution, in the sense that every solution to (3.11) provides a CPD of the desired form. This is assured by the following lemma.

LEMMA 3.6. *Let \mathcal{T} be a tensor such that the elements of $\mathcal{Y} = \text{MDFT}\{\mathcal{T}\}$ are given by (3.11), for some M, R satisfying $M \geq R$ and some set of M -tuples $\{(\lambda_{p,1}, \lambda_{p,2}, \dots, \lambda_{p,M})\}_{p=1}^P$. Then, another set of M -tuples $\{(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M})\}_{p=1}^P$ denotes a solution to the equations of the form (3.11) if and only if it is associated with P circulant $M \times R$ matrix factors constituting a CCPD of \mathcal{T} .*

Proof. The if part is clear from the analysis conducted to obtain the system of equations. For the only if part, we note that every solution $\{(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M})\}_{p=1}^P$ can be injectively associated with P circulant $M \times R$ matrices $\mathbf{G}^{(1)}, \dots, \mathbf{G}^{(P)}$ whose circulant completions are given by

$$\check{\mathbf{G}}^{(p)} = \mathbf{F}_M \text{diag}(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M}) \mathbf{F}_M^H.$$

By construction, those matrices are such that

$$\text{MDFT} \left\{ \llbracket \mathbf{G}^{(1)}, \dots, \mathbf{G}^{(P)} \rrbracket \right\} = \mathcal{Y},$$

because the eigenvalues of $\check{\mathbf{G}}^{(1)}, \dots, \check{\mathbf{G}}^{(P)}$ jointly satisfy the system of equations (3.11). Now, taking the IMDFT of both sides, we obtain $\llbracket \mathbf{G}^{(1)}, \dots, \mathbf{G}^{(P)} \rrbracket = \mathcal{T}$, as claimed. \square

It should be noted that, in general, (3.11) admits infinitely many solutions. In light of the above result, this is of course expected, because of the ambiguities that are inherent to the CP model. In particular, if the factors $\mathbf{C}^{(p)}$ in (3.5) satisfy Kruskal’s uniqueness condition, Theorem 2.1 tells us that the CPD is essentially unique, i.e., with only column scaling and permutation ambiguities on its factor matrices. This implies that the different solutions of (3.11) should also be related with each other accordingly. Our next result shows that such relation is rather simple.

THEOREM 3.7. *Let $\mathcal{T} \in \mathbb{C}^{M^{(P)}}$, with $P \geq 3$ and $M > 1$, be a hypercubic tensor of the form (3.5), whose factors $\mathbf{C}^{(p)}$ are circulant $M \times R$ matrices, with $R > 1$, that satisfy Kruskal’s condition $\sum_{p=1}^P k_{\mathbf{C}^{(p)}} \geq 2R + P - 1$. Let $\{(\lambda_{p,1}, \lambda_{p,2}, \dots, \lambda_{p,M})\}_{p=1}^P$ denote the solution of (3.11) corresponding to the eigenvalues of the circulant completions of those factors and assume that $\{(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M})\}_{p=1}^P$ is another possible solution of (3.11). Then, there are P complex scalars α_p satisfying $\prod_{p=1}^P \alpha_p = 1$ and an integer $r \in \{0, \dots, R - 1\}$ such that, for all $p \in \{1, \dots, P\}$ and $m \in \{1, \dots, M\}$, we have $\mu_{p,m} = \alpha_p \omega_M^{-r(m-1)} \lambda_{p,m}$.*

Proof. From Lemma 3.6, we have that the M -tuples $\{(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M})\}_{p=1}^P$ can be associated with circulant factors $\mathbf{G}^{(1)}, \dots, \mathbf{G}^{(P)}$ such that $\mathcal{T} = \llbracket \mathbf{G}^{(1)}, \dots, \mathbf{G}^{(P)} \rrbracket$. More precisely, $\{(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M})\}_{p=1}^P$ are the eigenvalues of $\check{\mathbf{G}}^{(1)}, \dots, \check{\mathbf{G}}^{(P)}$. But, since we assume that the factors satisfy Kruskal’s uniqueness condition, we have

$$(3.16) \quad \mathbf{G}^{(p)} = \mathbf{C}^{(p)} \mathbf{\Pi} \mathbf{\Delta}_p$$

for some permutation matrix $\mathbf{\Pi} \in \mathbb{R}^{R \times R}$ and some diagonal matrix $\mathbf{\Delta}_p \in \mathbb{C}^{R \times R}$, with the constraint $\mathbf{\Delta}_1 \dots \mathbf{\Delta}_P = \mathbf{I}_R$. Since by definition every column of $\mathbf{C}^{(p)}$ is of the form $\mathbf{\Pi}_M^r \mathbf{c}^{(p)}$, where $\mathbf{\Pi}_M$ is given by (3.2) and $\mathbf{c}^{(p)}$ is the generating vector of $\mathbf{C}^{(p)}$, we have from (3.16) that there exists $r \in \{0, \dots, R - 1\}$ such that

$$(3.17) \quad \mathbf{g}^{(p)} = \alpha_p \mathbf{\Pi}_M^r \mathbf{c}^{(p)},$$

where $\mathbf{g}^{(p)}$ is the generating vector of $\mathbf{G}^{(p)}$ and $\alpha_p = [\Delta_p]_{1,1}$. Due to the circulant structure of $\mathbf{G}^{(p)}$, we can also write

$$(3.18) \quad \mathbf{G}^{(p)} = \alpha_p \mathbf{\Pi}_M^r \mathbf{C}^{(p)}.$$

Now, substituting (3.4) in (3.17) and premultiplying both sides by $\sqrt{M} \mathbf{F}_M^H$, we obtain

$$\sqrt{M} \mathbf{F}_M^H \mathbf{g}^{(p)} = \alpha_p \sqrt{M} \mathbf{F}_M^H \mathbf{F}_M \text{diag} \left(1, \omega_M^{-r}, \dots, \omega_M^{-r(M-1)} \right) \mathbf{F}_M^H \mathbf{c}^{(p)},$$

But, since $\mathbf{G}^{(p)}$ and $\mathbf{C}^{(p)}$ are circulant, we have $\sqrt{M} \mathbf{F}_M^H \mathbf{g}^{(p)} = \boldsymbol{\mu}_p$ and $\sqrt{M} \mathbf{F}_M^H \mathbf{c}^{(p)} = \boldsymbol{\lambda}_p$, where $\boldsymbol{\mu}_p = [\mu_{p,1} \ \dots \ \mu_{p,M}]^T$ and $\boldsymbol{\lambda}_p = [\lambda_{p,1} \ \dots \ \lambda_{p,M}]^T$. Using this property and recalling that \mathbf{F}_M is unitary, we can rewrite the above equation as

$$\boldsymbol{\mu}_p = \alpha_p \text{diag} \left(1, \omega_M^{-r}, \dots, \omega_M^{-r(M-1)} \right) \boldsymbol{\lambda}_p,$$

which, in scalar form, corresponds to $\mu_{p,m} = \alpha_p \omega_M^{-r(m-1)} \lambda_{p,m}$. To complete the proof, we observe that the constraint $\Delta_1 \dots \Delta_P = \mathbf{I}_R$ implies $\prod_{p=1}^P \alpha_p = 1$ and that the constant $r \in \{0, \dots, R-1\}$ is the same for all $p \in \{1, \dots, P\}$, since the same permutation $\mathbf{\Pi}$ applies to all factors in (3.16). \square

COROLLARY 3.8. *If $M = R$, then $\mathbf{\Pi} = \mathbf{\Pi}_M^r$, where r is the same integer as in (3.18), and $\Delta_p = \alpha_p \mathbf{I}_R$ for all p .*

Proof. From Kruskal's condition, $k_{\mathbf{C}^{(q)}} > 1$ must hold for some $q \in \{1, \dots, P\}$. Let us consider two distinct columns of $\mathbf{G}^{(q)}$, which can be written as $\mathbf{g}_k^{(q)} = \alpha_{q,k} \mathbf{\Pi}_M^m \mathbf{c}^{(q)}$ and $\mathbf{g}_l^{(q)} = \alpha_{q,l} \mathbf{\Pi}_M^n \mathbf{c}^{(q)}$ for some $m, n \in \{0, \dots, M-1\}$, with $l > k$ and $\alpha_{q,k}, \alpha_{q,l} \neq 0$. As $\mathbf{G}^{(q)}$ is circulant, we also have $\mathbf{g}_l^{(q)} = \mathbf{\Pi}_M^{l-k} \mathbf{g}_k^{(q)}$, which implies $\mathbf{\Pi}_M^n \mathbf{c}^{(q)} = \beta \mathbf{\Pi}_M^{l-k+m} \mathbf{c}^{(q)}$, with $\beta = \alpha_{q,k} / \alpha_{q,l}$. Now, if $\beta \neq 1$, then $k_{\mathbf{C}^{(q)}} \leq 1$, contradicting our hypothesis. Using the same reasoning for all couples of columns of $\mathbf{G}^{(q)}$, we deduce that $\Delta_q = \alpha_q \mathbf{I}_R$. Moreover, since all circulant matrices commute, we have from (3.18) that $\mathbf{G}^{(q)} = \alpha_q \mathbf{C}^{(q)} \mathbf{\Pi}_M^r$. Substituting these results in (3.16) with $p = q$, it follows that $\mathbf{\Pi} = \mathbf{\Pi}_M^r$, which in turn implies $\Delta_p = \alpha_p \mathbf{I}_R$ for all p . \square

REMARK 3.9. *As seen above, when $M = R$ the permutation matrix of (3.16) must be circulant. Conversely, all the M distinct circulant permutations of the original factors yield an equivalent CCPD. In other words, if $\{(\lambda_{p,1}, \lambda_{p,2}, \dots, \lambda_{p,M})\}_{p=1}^P$ is a solution for (3.11), then so is any set $\{(\omega_M^0 \lambda_{p,1}, \omega_M^{-r} \lambda_{p,2}, \dots, \omega_M^{-r(M-1)} \lambda_{p,M})\}_{p=1}^P$ with $r \in \{0, \dots, M-1\}$. Indeed, from Proposition 3.5 we know that for $M = R$ every nonzero equation of the form (3.11) is associated with indices that satisfy $m_1 + \dots + m_P - P = lM$ for some $l \in \mathbb{N}$, and hence*

$$\prod_{p=1}^P \lambda_{p,m_p} = \omega_M^{-r(m_1 + \dots + m_P - P)} \prod_{p=1}^P \lambda_{p,m_p} = \prod_{p=1}^P \omega_M^{-r(m_p - 1)} \lambda_{p,m_p}.$$

3.2. Symmetric CCPD. The symmetric case can be derived from the previous subsection by considering that $\mathbf{C}^{(1)} = \dots = \mathbf{C}^{(P)} = \mathbf{C} \in \mathbb{C}^{M \times R}$, which yields

$$(3.19) \quad \mathcal{T} = \llbracket \mathbf{C}, \dots, \mathbf{C} \rrbracket \in \mathbb{C}^{M^{(P)}}.$$

As we shall see, this simplifies the set of solutions of equations (3.11). First, let us rewrite those equations as

$$(3.20) \quad y_{m_1, \dots, m_P} = \left(\frac{1}{\sqrt{M}} \right)^P \prod_{p=1}^P \lambda_{m_p} \sum_{r=1}^R \omega_M^{-(r-1)(m_1 + \dots + m_P - P)},$$

TABLE 1
Summary of the AS method.

Inputs: Hypercubic tensor $\mathcal{T} \in \mathbb{C}^{M^{(P)}}$ to be decomposed and the rank R of the CCPD.

Outputs: Circulant $M \times R$ complex matrix factors $\mathbf{C}^{(1)}, \dots, \mathbf{C}^{(P)}$ (or \mathbf{C} , in the symmetric case).

1. Compute $\mathcal{Y} = \text{MDFT}\{\mathcal{T}\}$.
2. Find the relevant equations using Proposition 3.4 (or Proposition 3.5, if $M = R$).
3. Solve the relevant equations (3.11) (resp., (3.20)) for $\lambda_{p,m}$ (resp., λ_m).
4. Reconstruct the factors $\mathbf{C}^{(1)}, \dots, \mathbf{C}^{(P)}$ (resp., \mathbf{C}) by resorting to property (3.3).

where the first subscript of the eigenvalues has been dropped, since they all refer now to the same circulant completion, $\check{\mathbf{C}}$. This particularization clearly renders some of the equations (3.20) redundant, because the MDFT of a symmetric tensor is also symmetric. Moreover, assuming that Kruskal's condition holds, i.e., that $k_{\mathbf{C}} \geq (2R - 1)/P + 1$, we can show that the set of solutions of (3.20) is finite, having at most PM elements. Once again, this is demonstrated by exploiting the link between these solutions and the set of factors which can yield a symmetric CCPD of \mathcal{T} , as follows.

THEOREM 3.10. *Let \mathcal{T} be given by (3.19), where $\mathbf{C} \in \mathbb{C}^{M \times R}$, with $M \geq R > 1$, is circulant. If $k_{\mathbf{C}} \geq (2R - 1)/P + 1$, then the equations (3.20) admit at most PM different solutions.*

Proof. The proof is similar to that of Theorem 3.7, with the difference that we must now have $\alpha_1 = \dots = \alpha_P$, since the CCPD (3.19) is symmetric. Due to the constraint $\mathbf{\Delta}_1 \dots \mathbf{\Delta}_P = \mathbf{I}_R$, the only possible distinct values for α are the P th roots of unity, i.e., $\alpha = \omega_P^p$ for $p \in \{0, \dots, P-1\}$. Combining this with the existence of only M distinct values for the integer r in (3.17), we are left with at most PM distinct solutions for (3.20). \square

REMARK 3.11. *From the result of Corollary 3.8 we conclude that, if a symmetric CCPD with square factors is essentially unique, then (3.20) admits exactly PM solutions, since every combination of one of the P distinct matrices $\omega_P^p \mathbf{I}_R$ with one of the M distinct matrices $\mathbf{\Pi}_M^r$ yields an equivalent symmetric CCPD.*

3.3. The AS method. As we have shown, *any* solution of the equations (3.11) provides a set of P factors which jointly fulfill (3.5), as desired. Therefore, we conclude that the resolution of such equations provides an *exact* solution for the CCPD in the noiseless case. In the presence of noise, evidently, only an approximate solution of (3.11) can be sought.

Table 1 summarizes the approach proposed in this paper, which we shall refer to as an algebraic solution (AS) for the CCPD. With regard to the first step, we point out that \mathcal{Y} can be computed in an efficient way, through a multidimensional FFT algorithm [11]. The second step can be performed by verifying for all combinations of indices $m_1, \dots, m_P \in \{1, \dots, M\}$ which of them satisfy one of the conditions of Proposition 3.4, if $M > R$, or the condition of Proposition 3.5, if $M = R$. It is evident that, since for each combination of indices these conditions depend only on P , R and M , their assessment can be done *a priori* and reused for several different hypercubic tensors of same order, rank and dimension. Step 3 is the core of the method, being discussed below. Finally, step 4 consists in simply reconstructing the desired factors from the estimated eigenvalues of their circulant completions.

The resolution of the set of monomial equations (i.e., step 3 of Table 1) is a central matter in our approach. However, discussing computational methods for the resolution of polynomial systems, which have been employed to solve practical problems

(see, e.g., [22] and references therein), is outside the scope of this paper. Here, we consider only some simple examples of the systems (3.11) and (3.20), where the solution of the monomial equations can be easily derived in a direct, non-optimal, manner. This derivation is done in §4, which provides an illustration of our theoretical results. Moreover, the derived procedures will be used for numerical evaluation purposes in §5. This kind of *ad hoc* approach, which we shall refer to as AAS (*ad-hoc* algebraic solution), is of very low computational cost and can always be derived, as long as some assumptions about the non-nullity of certain eigenvalues hold. These assumptions hold generically for hypercubic tensors \mathcal{T} of the form (3.5) having tall or square factors, and therefore little generality is lost. Although the accuracy of the AAS estimate is degraded in the presence of noise, it can be effectively and efficiently refined with iterative algorithms. Indeed, when initialized with the AAS estimate, these algorithms can avoid local minima and converge relatively fast, performing substantially better than when randomly initialized, as will be illustrated in §5.

4. Illustrative examples.

EXAMPLE 4.1. Let $\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \mathbf{C}^{(3)} \in \mathbb{C}^{3 \times 2}$ be circulant matrices with generating vectors $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}$ and $\mathbf{c}^{(3)}$, respectively, and consider the CCPD $\mathcal{T} = \llbracket \mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \mathbf{C}^{(3)} \rrbracket$. We can obtain from the nonzero elements of $\mathcal{Y} = \text{MDFT}\{\mathcal{T}\}$ 27 equations in the 9 unknowns of interest, which are the eigenvalues of $\check{\mathbf{C}}^{(1)}, \check{\mathbf{C}}^{(2)}$ and $\check{\mathbf{C}}^{(3)}$. In particular, we have the equation

$$(4.1) \quad \lambda_{1,1} \lambda_{2,1} \lambda_{3,1} = \frac{\tilde{y}_{1,1,1}}{2},$$

where $\tilde{y}_{m_1, m_2, m_3} \triangleq 3\sqrt{3} y_{m_1, m_2, m_3}$. We know also that an infinite number of solutions exist, which is a consequence of the scaling ambiguity. Yet, since $\boldsymbol{\lambda}_p / \sqrt{3}$ is the DFT of $\mathbf{c}^{(p)}$, under the assumption that $\lambda_{1,1}, \lambda_{2,1}, \lambda_{3,1} \neq 0$ we can conclude that the value of each $\lambda_{p,1}$ determines the scaling of its associated factor $\mathbf{C}^{(p)}$. In this case, we can eliminate that ambiguity by imposing the values of all but one $\lambda_{p,1}$. For instance, if we choose $\lambda_{2,1} = \lambda_{3,1} = 1$, then (4.1) can be directly solved for $\lambda_{1,1}$. Next, we can use, for example, the equations

$$\begin{aligned} \lambda_{1,1} \lambda_{2,1} \lambda_{3,2} &= \frac{\tilde{y}_{1,1,2}}{1 + \omega_3^{-1}} \rightarrow \lambda_{3,2}, & \lambda_{1,2} \lambda_{2,1} \lambda_{3,1} &= \frac{\tilde{y}_{2,1,1}}{1 + \omega_3^{-1}} \rightarrow \lambda_{1,2}, \\ \lambda_{1,1} \lambda_{2,3} \lambda_{3,1} &= \frac{\tilde{y}_{1,3,1}}{1 + \omega_3^{-2}} \rightarrow \lambda_{2,3}, & \lambda_{1,1} \lambda_{2,2} \lambda_{3,1} &= \frac{\tilde{y}_{1,2,1}}{1 + \omega_3^{-1}} \rightarrow \lambda_{2,2}, \\ \lambda_{1,1} \lambda_{2,1} \lambda_{3,3} &= \frac{\tilde{y}_{1,1,3}}{1 + \omega_3^{-2}} \rightarrow \lambda_{3,3}, & \lambda_{1,3} \lambda_{2,1} \lambda_{3,1} &= \frac{\tilde{y}_{3,1,1}}{1 + \omega_3^{-2}} \rightarrow \lambda_{1,3} \end{aligned}$$

to find the values of the other eigenvalues, as indicated. With those eigenvalues at hand, the reconstruction of the desired factors is trivial via relation (3.3). We emphasize that other choices of equations are possible for obtaining a solution and that the procedure described above is by no means guaranteed to be optimal in any sense.

After resolving the scaling indeterminacy, obtaining the corresponding solution is straightforward in the Example 4.1, provided that $\lambda_{1,1}, \lambda_{2,1}, \lambda_{3,1} \neq 0$. This simple approach can be employed in general, but other instances may involve more complicated procedures and may require the non-nullity of other eigenvalues. Nonetheless, when applicable, the resulting procedure is computationally inexpensive, and thus may be used to initialize more complicated methods, as shown in the next section.

EXAMPLE 4.2. For the case where $P = M = R = 3$, defining $\tilde{y}_{m_1, m_2, m_3} \triangleq \sqrt{3} y_{m_1, m_2, m_3}$ and disregarding redundant equations due to symmetries, we have the

following set of equations:

$$\lambda_1^3 = \tilde{y}_{1,1,1}, \quad \lambda_2^3 = \tilde{y}_{2,2,2}, \quad \lambda_3^3 = \tilde{y}_{3,3,3}, \quad \lambda_1 \lambda_2 \lambda_3 = \tilde{y}_{1,2,3}.$$

One can see immediately that we now have three uncoupled equations for calculating the three unknowns $\lambda_1, \lambda_2, \lambda_3$. However, the coupled equation in $\lambda_1 \lambda_2 \lambda_3$ constrains the possible solutions, so that there can be at most 9 instead of 27 solutions, which correspond precisely to the maximum number of PM solutions stated by Theorem 3.10.

REMARK 4.3. If it is known a priori that $\mathbf{C} \in \mathbb{R}^{M \times R}$ and $\lambda_1 \neq 0$, we can choose a real solution for the equation involving λ_1^P , which is guaranteed to exist due to (3.10). This allows suppressing the scaling ambiguity whilst obtaining a real factor \mathbf{C} .

5. Numerical simulations. In order to evaluate the proposed approach and compare it with existing estimation methods, we present now some Monte Carlo simulations. The performance of each method is quantified by the error in the estimated parameters, which constitute the generating vectors of the circulant factors, and by the reconstruction error, which corresponds to the difference between the data tensor and the estimated model. In all simulations, white Gaussian noise is added to the signal tensor at different signal-to-noise ratio (SNR) levels to yield the data tensor.

Before, however, we formulate a modified ALS algorithm for third order tensors which constrains the factors to be circulant. Since we are interested in factors with such structure, this version, which we shall call CALS (for circulant-constrained ALS), is a more appropriate reference for our evaluation than the standard ALS.

5.1. CALS algorithm. Let $\mathcal{C}^{M \times R} \subset \mathbb{C}^{M \times R}$ be the subspace of circulant $M \times R$ matrices, and consider the basis $\{\mathbf{E}_m\}_{m=1}^M$, where the generating vector of \mathbf{E}_m is the canonical basis vector \mathbf{e}_m of \mathbb{C}^M . Clearly, any matrix $\mathbf{C}^{(p)} \in \mathcal{C}^{M \times R}$ with generating vector $\mathbf{c}^{(p)} = [c_1^{(p)} \ \dots \ c_M^{(p)}]^T$ can be written as

$$(5.1) \quad \mathbf{C}^{(p)} = \sum_{m=1}^M c_m^{(p)} \mathbf{E}_m = \text{unvec}_R(\mathbf{E} \mathbf{c}^{(p)}),$$

where $\mathbf{E} = [\text{vec}(\mathbf{E}_1) \ \dots \ \text{vec}(\mathbf{E}_M)] \in \mathbb{C}^{RM \times M}$ and the unvec_R operator is such that $\text{unvec}_R(\mathbf{a}) = [\mathbf{a}_1 \ \dots \ \mathbf{a}_R] \in \mathbb{C}^{N \times R}$ for every vector $\mathbf{a} = [\mathbf{a}_1^T \ \dots \ \mathbf{a}_R^T]^T \in \mathbb{C}^{RN}$ with $\mathbf{a}_1, \dots, \mathbf{a}_R \in \mathbb{C}^N$. Hence, if a third-order tensor \mathcal{T} is given by (3.5), from (2.8) and (5.1) we deduce that its unfolding \mathbf{T}_1 is given by

$$\begin{aligned} \mathbf{T}_1 &= \mathbf{C}^{(1)} \left(\mathbf{C}^{(3)} \diamond \mathbf{C}^{(2)} \right)^T \\ &= \text{unvec}_R(\mathbf{E} \mathbf{c}^{(1)}) \left(\mathbf{C}^{(3)} \diamond \mathbf{C}^{(2)} \right)^T. \end{aligned}$$

Now, if we assume $\mathbf{C}^{(3)} \diamond \mathbf{C}^{(2)}$ is full-column rank (which always holds for full column-rank $\mathbf{C}^{(2)}$ and $\mathbf{C}^{(3)}$), we have

$$(5.2) \quad \text{unvec}_R(\mathbf{E} \mathbf{c}^{(1)}) = \mathbf{T}_1 \left[\left(\mathbf{C}^{(3)} \diamond \mathbf{C}^{(2)} \right)^T \right]^\dagger.$$

Then, applying vec on both sides and multiplying them by $\mathbf{E}^\dagger = \frac{1}{R} \mathbf{E}^T$ we obtain

$$(5.3) \quad \mathbf{c}^{(1)} = \frac{1}{R} \mathbf{E}^T (\mathbf{I}_R \otimes \mathbf{T}_1) \text{vec} \left\{ \left[\left(\mathbf{C}^{(3)} \diamond \mathbf{C}^{(2)} \right)^T \right]^\dagger \right\},$$

TABLE 2
The CALS algorithm.

Inputs: $\mathcal{T} \in \mathbb{C}^{M \times M \times M}$, rank R of the CCPD and initial generating vectors $\mathbf{c}_0^{(1)}, \mathbf{c}_0^{(2)}, \mathbf{c}_0^{(3)}$.

Outputs: Circulant $M \times R$ complex matrix factors $\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \mathbf{C}^{(3)}$.

$i = 1$

repeat

$$\begin{aligned} \mathbf{c}_i^{(1)} &= \frac{1}{R} \mathbf{E}^T (\mathbf{I}_R \otimes \mathbf{T}_1) \text{vec} \left\{ \left[\left(\mathbf{C}_{i-1}^{(3)} \diamond \mathbf{C}_{i-1}^{(2)} \right)^T \right]^\dagger \right\} \\ \mathbf{c}_i^{(2)} &= \frac{1}{R} \mathbf{E}^T (\mathbf{I}_R \otimes \mathbf{T}_2) \text{vec} \left\{ \left[\left(\mathbf{C}_{i-1}^{(3)} \diamond \mathbf{C}_i^{(1)} \right)^T \right]^\dagger \right\} \\ \mathbf{c}_i^{(3)} &= \frac{1}{R} \mathbf{E}^T (\mathbf{I}_R \otimes \mathbf{T}_3) \text{vec} \left\{ \left[\left(\mathbf{C}_i^{(2)} \diamond \mathbf{C}_i^{(1)} \right)^T \right]^\dagger \right\} \end{aligned}$$

$i = i + 1$

until convergence

return $\mathbf{C}_i^{(1)}, \mathbf{C}_i^{(2)}, \mathbf{C}_i^{(3)}$

where we have used the property $\text{vec}(\mathbf{ABC}^T) = (\mathbf{C} \otimes \mathbf{A})\text{vec}(\mathbf{B})$ with $\mathbf{C} = \mathbf{I}_R$. Expression (5.3) provides an update rule for the generating vector $\mathbf{c}^{(1)}$, instead of a general factor matrix. This allows us to exploit the structure of $\mathbf{C}^{(1)}$ for obtaining better-estimated parameters. The resulting algorithm is summarized in Table 2.

5.2. Methodology. Initially, we draw the entries of the $P = 3$ generating vectors $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \mathbf{c}^{(3)} \in \mathbb{C}^M$ from a complex Gaussian distribution with zero mean and unit variance. However, their first elements are set to unity with the purpose of eliminating the scale ambiguity, an assumption which is usual in some applications, such as those which require the estimation of FIR systems (see, e.g., [15]). Next, $\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \mathbf{C}^{(3)} \in \mathbb{C}^{M \times R}$ are constructed according to (3.1), and the signal tensor \mathcal{T} is obtained with (3.5). The data tensor is then generated via $\mathcal{D} = \mathcal{T} + \sigma_N \mathcal{N}$, where $\mathcal{N} \in \mathbb{C}^{M \times M \times M}$ is a noise tensor, whose elements are drawn from a complex Gaussian distribution and normalized to ensure that $\|\mathcal{N}\|_F = 1$, and σ_N is a positive parameter used to impose the desired level of SNR (in dB), which is defined as

$$\text{SNR} = 10 \log_{10} \frac{\|\mathcal{T}\|_F^2}{\sigma_N^2}.$$

Once the data tensor is available, we employ each estimation method to obtain estimates $\hat{\mathbf{C}}^{(1)}, \hat{\mathbf{C}}^{(2)}$ and $\hat{\mathbf{C}}^{(3)}$ such that $\hat{\mathcal{T}} = \llbracket \hat{\mathbf{C}}^{(1)}, \hat{\mathbf{C}}^{(2)}, \hat{\mathbf{C}}^{(3)} \rrbracket$ yields the reconstructed tensor. The corresponding reconstruction error (in dB) is then calculated with

$$\text{NMSE} = 10 \log_{10} \frac{\|\mathcal{D} - \hat{\mathcal{T}}\|_F^2}{\|\mathcal{D}\|_F^2}$$

and, for the iterative algorithms, we denote the reconstruction error at iteration i by NMSE_i . Finally, the error in each estimated generating vector $\hat{\mathbf{c}}^{(p)}$ is given by

$$(5.4) \quad \varepsilon_p = \min_{r \in \{1, \dots, R\}} \frac{\left\| \mathbf{c}^{(p)} - \hat{\mathbf{c}}_r^{(p)} / \hat{c}_{r,1}^{(p)} \right\|_2^2}{\|\mathbf{c}^{(p)}\|_2^2},$$

where $\hat{c}_{r,1}^{(p)}$ is the first element of $\hat{\mathbf{c}}_r^{(p)}$. Observe that this criterion amounts to choosing among the columns of $\hat{\mathbf{C}}^{(p)}$ that one which, after being appropriately normalized,

TABLE 3
Summary of the evaluated methods.

Acronym	Algorithm	Initialization	Curve symbol
AAS	<i>ad-hoc</i> AS solution	—	○ . . .
ALS	ALS	random circulant factors	* —
AAS-ALS	ALS	given by AAS	◇ —
CALS	CALS	random circulant factors	+ - -
AAS-CALS	CALS	given by AAS	▽ - -
LMAS	LM applied to AS equations	random eigenvalues	△ - .
AAS-LMAS	LM applied to AS equations	given by AAS	□ - .

matches $\mathbf{c}^{(p)}$ with the lowest normalized error. With this criterion, we are able to circumvent both scaling and permutation ambiguities, which is necessary when measuring the error in the parameters of the estimated model (i.e., the generating vectors).

Two different scenarios were considered: the first one refers to the general CCPD (§5.3); the second one deals with the symmetric CCPD (§5.4). In both scenarios, we compared seven different methods, which are summarized in Table 3. Note that the acronym LMAS stands for Levenberg-Marquardt algorithm applied to the Algebraic Solution, since such method is employed to solve the monomial equations. We evaluate this algorithm with both random (denoted LMAS) and AAS (denoted AAS-LMAS) initializations, for comparison purposes. For a fair comparison, all the iterative algorithms were assumed to converge whenever the reconstruction errors obtained for two consecutive iterations satisfied $\text{NMSE}_i - \text{NMSE}_{i-1} < 10^{-6}$ or when the maximum number of 2000 iterations was reached. For each given choice of M and R the procedure above described was repeated for $N = 2000$ Monte Carlo runs, with the SNR varying in the interval $[20, 80]$.

5.3. General CCPD. Since the factors estimated by the standard ALS algorithm are not guaranteed to be circulant, it is necessary to post-process them for imposing such structure. To this end, we select from each factor the column that minimizes (5.4) and take its normalized version (i.e., such column is divided by its first element) as the estimate of the generating vector, $\hat{\mathbf{c}}^{(p)}$, from which the corresponding estimated circulant factor is obtained to compute the reconstruction error.

Since the optimization variables of LMAS are the eigenvalues λ_{p,m_p} , the constraint of circulant structure is already taken into account, with no need of post-processing. It should be noted, however, that the scaling indeterminacy implies the existence of degenerate critical points in the parameters space, and thus a near-singularity of the Jacobian matrix is observed in some iterations unless such ambiguity is eliminated. We avoid it by imposing $\lambda_{2,1} = \lambda_{3,1} = 1$, similarly to what was done in Example 4.1, and perform the optimization with respect to the other eigenvalues. Note that this does not imply any loss of generality as long as $\lambda_{2,1} \neq 0$ and $\lambda_{3,1} \neq 0$, which holds with probability one in our simulations. The adopted cost function is given by $J(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3) = \|\mathcal{R}\|_F^2 / \|\mathcal{Y}\|_F^2$, where $\mathcal{R} = \mathcal{R}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3) \in \mathbb{C}^{M^{(P)}}$ is a residue tensor whose elements satisfy

$$(5.5) \quad [\mathcal{R}]_{m_1, \dots, m_P} = [\mathcal{Y}]_{m_1, \dots, m_P} - M^{-\frac{3}{2}} \lambda_{1,m_1} \lambda_{2,m_2} \lambda_{3,m_3} \sum_{r=1}^R \omega_3^{-(r-1)(m_1+m_2+m_3-3)}.$$

The legend for the curves of Figures 1 and 2, which display the results for $M = 3, R = 2$ and for $M = R = 3$, respectively, is given in Table 3. We can see that the average errors associated with AAS steadily drop with the increase of SNR. The

same is not true for the randomly initialized iterative methods, which have shown a poor performance due to their frequent convergence to local minima. When they are initialized with the AAS estimate, nonetheless, they are able to refine it and attain low average error levels, which indicates that local minima are avoided. In particular, AAS-CALS and AAS-LMAS exhibit the best results. Note that no graph of ε_2 and ε_3 is shown in Figures 1 and 2, since all the ε_p are extremely close in both cases.

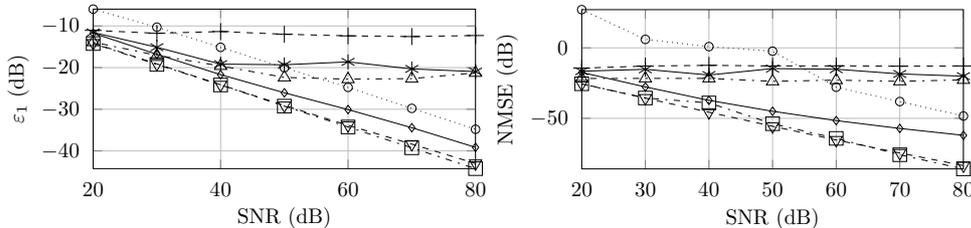


FIGURE 1. Simulation results for the general CCPD scenario with $M = 3$, $R = 2$: average errors in the estimated parameters of the first factor (left) and average reconstruction error (right).

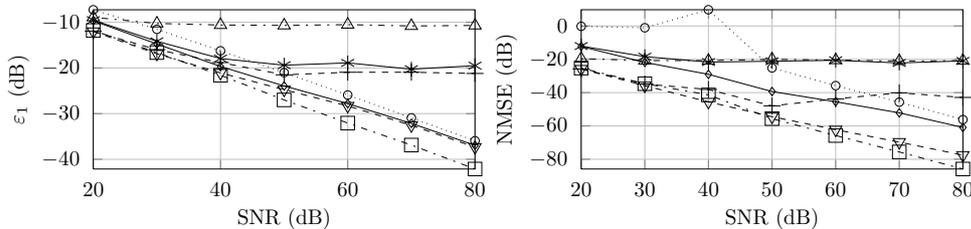


FIGURE 2. Simulation results for the general CCPD scenario with $M = R = 3$: average errors in the estimated parameters of the first factor (left) and average reconstruction error (right).

5.4. Symmetric CCPD. In the symmetric CCPD case, we have only one generating vector and one associated factor, \mathbf{c} and \mathbf{C} , respectively, with \mathcal{T} being constructed as in (3.19). Hence, $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon$. Furthermore, the outputs of both ALS and CALS need to be post-processed, since these algorithms do not constrain the decomposition to be symmetric. Thus, for each one of them, we search among the columns of all estimated factors which one minimizes ε , and take it (with proper normalization) as the generating vector of the factor estimate.

With regard to the LMAS and AAS-LMAS methods, we use again the cost function (5.5), but with $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$, since now there is only one vector of eigenvalues to be optimized. Hence, once more no post-processing is necessary. Furthermore, unlike the previous scenario, there is no need to fix the value of any optimization variable, since, under the assumptions of Theorem 3.10, which hold with probability one in our simulations, the symmetric CCPD admits only a finite number of solutions.

The results for $M = 3, R = 2$ and for $M = R = 3$ are shown, respectively, in Figures 3 and 4. The legend for the curves is also that one contained in Table 3. Again, it is seen that AAS outperforms the randomly initialized iterative methods in the sense that its errors consistently decrease with the increase of SNR. Its estimates are then notably improved when used as initial point by the iterative methods. The general conclusions are thus basically the same as in the general CCPD scenario, indicating that the AAS initialization helps avoiding local minima. In this scenario,

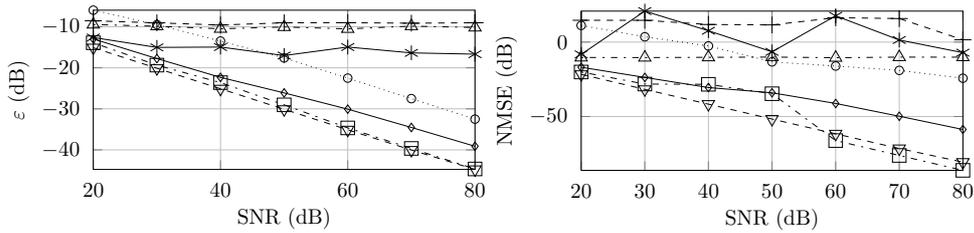


FIGURE 3. Simulation results for the symmetric CCPD scenario with $M = 3$, $R = 2$: average errors in the estimated parameters of the first factor (left) and average reconstruction error (right).

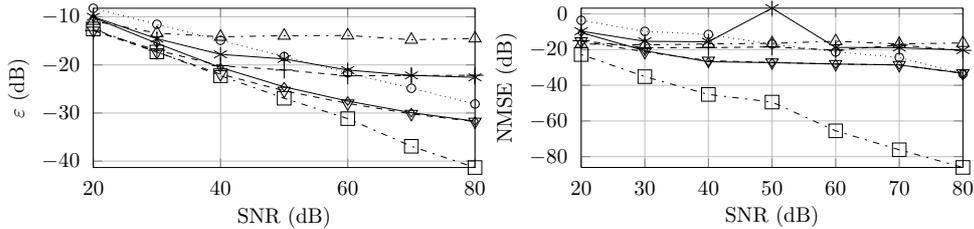


FIGURE 4. Simulation results for the symmetric CCPD scenario with $M = R = 3$: average errors in the estimated parameters of the first factor (left) and average reconstruction error (right).

however, AAS-LMAS shows a performance which is notably better than those of the other methods for $M = R = 3$, while ALS and CALS cannot improve much further the reconstruction error attained by AAS.

To evaluate the impact of the initialization in the computational cost of the iterative algorithms, we compare in Table 5.4, for two values of SNR (20 and 60 dB), the average number of iterations required for their convergence in both considered scenarios. As we can see, these numbers are significantly reduced when the AAS estimate is used as initial point, especially for 60 dB.

5.5. Discussion. The AAS estimates are degraded in the presence of noise, since such method solves the monomial equations via non-optimal *ad hoc* procedures that (i) exploit only the minimum number of equations required and (ii) resort to divisions and multiplications in order to eliminate variables, which can produce large estimation errors. Nonetheless, by refining these estimates with iterative methods, a much better performance is attained in comparison with that provided by those methods when randomly initialized. This suggests that the AAS estimate often lies in some neighborhood of an optimal one, from where an iterative algorithm can quickly converge, thus avoiding local minima.

6. Concluding remarks. In this paper, we have proposed a novel approach for the problem of estimating circulant factors which constitute a CPD. This approach is based on the resolution of a system of homogeneous monomial equations which were derived by taking the multidimensional discrete Fourier transform of the tensor to be decomposed. Our results show that, even though such system admits in general an infinite number of solutions, any of them provides factors with the desired form, which is consistent with the existence of inherent ambiguities in the CP model. Moreover, under the standard Kruskal's uniqueness condition, we have shown that all solutions are related with each other in a simple way. We have also considered the symmetric

TABLE 4
Average number of iterations required for the convergence of each iterative algorithm.

Algorithm	General CCPD				Symmetric CCPD			
	$M = 3, R = 2$		$M = 3, R = 3$		$M = 3, R = 2$		$M = 3, R = 3$	
	SNR (dB)		SNR (dB)		SNR (dB)		SNR (dB)	
	20	60	20	60	20	60	20	60
ALS	38	41	46	39	40	41	46	20
AAS-ALS	15	2	27	2	18	2	35	3
CALS	9	9	37	37	9	8	20	19
AAS-CALS	6	2	23	2	5	2	10	3
LMAS	77	78	179	182	12	12	13	12
AAS-LMAS	16	3	20	3	6	3	5	3

CPD with circulant factors, whose resulting monomial equations are simpler and have been shown to admit a finite number of solutions under the same uniqueness condition.

Numerical simulations were presented, where the estimate provided by a simple *ad-hoc* procedure used for solving the monomial equations is compared with those given by iterative algorithms. Such estimate is inexpensive to calculate but is degraded in the presence of noise. Nonetheless, it is generally better in average than those given by iterative methods when randomly initialized, and can be subsequently refined by them, which leads to a noteworthy performance. It was empirically observed that this combined strategy diminishes the probability of convergence of the iterative methods towards local minima and also significantly accelerates their convergence. It should be emphasized, however, that our use of such *ad-hoc* procedure was mainly intended for validating the proposed approach. Conceivably, more sophisticated methods for the resolution of the monomial equations could lead to better results.

REFERENCES

- [1] E. ACAR, C. AYKUT-BINGOL, H. BINGOL, R. BRO, AND B. YENER, *Multiway analysis of epilepsy tensors*, *Bioinformatics*, 23 (2007), pp. i10–i18.
- [2] B. W. BADER, M. W. BERRY, AND M. BROWNE, *Discussion tracking in Enron email using PARAFAC*, in *Survey of Text Mining II*, Michael W. Berry and Malu Castellanos, eds., Springer London, 2008, pp. 147–163.
- [3] J. BRACHAT, P. COMON, B. MOURRAIN, AND E. TSIGARIDAS, *Symmetric tensor decomposition*, *Linear Algebra and its Applications*, 433 (2010), pp. 1851–1872.
- [4] R. BRO, *PARAFAC. Tutorial and applications*, *Chemometrics and Intelligent Laboratory Systems*, 38 (1997), pp. 149–171.
- [5] J. D. CARROLL AND J.-J. CHANG, *Analysis of individual differences in multidimensional scaling via an n -way generalization of “Eckart-Young” decomposition*, *Psychometrika*, 35 (1970), pp. 283–319.
- [6] J. D. CARROLL, S. PRUZANSKY, AND J. B. KRUSKAL, *Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters*, *Psychometrika*, 45 (1980), pp. 3–24.
- [7] P. COMON, G. GOLUB, L.-H. LIM, AND B. MOURRAIN, *Symmetric tensors and symmetric tensor rank*, *SIAM Journal on Matrix Analysis and Applications*, 30 (2008), pp. 1254–1279.
- [8] P. COMON, X. LUCIANI, AND A. L. F. DE ALMEIDA, *Tensor decompositions, alternating least squares and other tales*, *Journal of Chemometrics*, 23 (2009), pp. 393–405.
- [9] P. COMON, M. SORENSEN, AND E. TSIGARIDAS, *Decomposing tensors with structured matrix factors reduces to rank-1 approximations*, in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Dallas, Texas, USA, 2010, pp. 3858–3861.
- [10] V. DE SILVA AND L.-H. LIM, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, *SIAM Journal on Matrix Analysis and Applications*, 30 (2008), pp. 1084–1127.
- [11] D. E. DUDGEON AND R. M. MERSEREAU, *Multidimensional Digital Signal Processing*, Prentice Hall, NJ, USA, 1984.

- [12] G. FAVIER, T. BOUILLOC, AND A. L. F. DE ALMEIDA, *Blind constrained block-Tucker2 receiver for multiuser SIMO NL-CDMA communication systems*, Signal Processing, 92 (2012), pp. 1624–1636.
- [13] G. FAVIER, A. Y. KIBANGOU, AND T. BOUILLOC, *Nonlinear system modeling and identification using Volterra-PARAFAC models*, International Journal of Adaptive Control and Signal Processing, 26 (2012), pp. 30–53.
- [14] C. A. R. FERNANDES, G. FAVIER, AND J. C. M. MOTA, *Parafac-based channel estimation and data recovery in nonlinear MIMO spread spectrum communication systems*, Signal Processing, 91 (2011), pp. 311–322.
- [15] C. E. R. FERNANDES, G. FAVIER, AND J. C. M. MOTA, *Blind channel identification algorithms based on the Parafac decomposition of cumulant tensors: the single and multiuser cases*, Signal Processing, 88 (2008), pp. 1382–1401.
- [16] R. M. GRAY, *Toeplitz and circulant matrices: A review*, Foundations and Trends in Communications and Information Theory, 2 (2006), pp. 155–239.
- [17] R. A. HARSHMAN, *Foundations of the PARAFAC procedure: models and conditions for an “explanatory” multi-modal factor analysis*, UCLA Working Papers in Phonetics, 16 (1970), pp. 1–84.
- [18] C. J. HILLAR AND L.-H. LIM, *Most tensor problems are NP hard*, Preprint arXiv:0911.1393v5, (2009).
- [19] A. Y. KIBANGOU AND G. FAVIER, *Non-iterative solution for PARAFAC with a Toeplitz matrix factor*, in EUSIPCO, Glasgow, UK, 2009, pp. 691–695.
- [20] T. G. KOLDA AND B. W. BADER, *Tensor decompositions and applications*, SIAM review, 51 (2009), pp. 455–500.
- [21] J. B. KRUSKAL, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra and its Applications, 18 (1977), pp. 95–138.
- [22] J. LEBRUN, *Normal forms in statistical signal processing*, in Gröbner Bases in Control Theory and Signal Processing, de Gruyter, 2007, pp. 107–126.
- [23] K. LEVENBERG, *A method for the solution of certain non-linear problems in least squares*, Quarterly of Applied Mathematics, 2 (1944), pp. 164–168.
- [24] M. MARCUS, *Finite dimensional multilinear algebra, Part 1*, Pure and applied mathematics, M. Dekker, New York, NY, 1973.
- [25] D. W. MARQUARDT, *An algorithm for least squares estimation of non-linear parameters*, SIAM Journal, 11 (1963), pp. 431–441.
- [26] M. MØRUP, *Applications of tensor (multiway array) factorizations and decompositions in data mining*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1 (2011), pp. 24–40.
- [27] P. PAATERO, *Construction and analysis of degenerate PARAFAC models*, Journal of Chemometrics, 14 (2000), pp. 285–299.
- [28] N. D. SIDIROPOULOS AND R. BRO, *On the uniqueness of multilinear decomposition of N-way arrays*, Journal of Chemometrics, 14 (2000), pp. 229–239.
- [29] N. D. SIDIROPOULOS, G. B. GIANNAKIS, AND R. BRO, *Blind PARAFAC receivers for DS-CDMA systems*, IEEE Transactions on Signal Processing, 48 (2000), pp. 810–823.
- [30] M. SORENSEN AND P. COMON, *Tensor decompositions with banded matrix factors*, Linear Algebra and its Applications, 438 (2013), pp. 919–941.
- [31] M. SORENSEN AND L. DE LATHAUWER, *Blind signal separation via tensor decomposition with Vandermonde factor: canonical polyadic decomposition*, IEEE Transactions on Signal Processing, 61 (2013), pp. 5507–5519.
- [32] A. STEGEMAN, *Low-rank approximation of generic $p \times q \times 2$ arrays and diverging components in the Candecomp/Parafac model*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 988–1007.
- [33] G. TOMASI AND R. BRO, *A comparison of algorithms for fitting the PARAFAC model*, Computational Statistics and Data Analysis, 50 (2006), pp. 1700–1734.