Comput Optim Appl
DOI 10.1007/s10589-015-9761-5

# Exact line and plane search for tensor optimization

**Laurent Sorber**[1,3,4] · **Ignat Domanov**[2,3,4] · **Marc Van Barel**[1] · **Lieven De Lathauwer**[2,3,4]

Received: 31 May 2014
© Springer Science+Business Media New York 2015

**Abstract** Line and plane searches are used as accelerators and globalization strategies in many optimization algorithms. We introduce a class of optimization problems called tensor optimization, which comprises applications ranging from tensor decompositions to least squares support tensor machines. We develop algorithms to efficiently compute the global minimizers of their line and plane search subproblems. Furthermore, we introduce scaled line and plane search, which compute an optimal scaling of the solution simultaneously with the optimal line or plane search step, and show that this scaling can be computed at almost no additional cost. Obtaining the global minimizers of (scaled) line and plane search problems often requires solving a bivariate or polyanalytic polynomial system. We show how to compute the isolated real solutions of bivariate polynomial systems and the isolated complex solutions of polyanalytic polynomial systems using a single generalized eigenvalue decomposition. Finally,

✉ Ignat Domanov
Ignat.Domanov@kuleuven-kulak.be

Laurent Sorber
Laurent.Sorber@cs.kuleuven.be

Marc Van Barel
Marc.VanBarel@cs.kuleuven.be

Lieven De Lathauwer
Lieven.DeLathauwer@kuleuven-kulak.be

1 Department of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium

2 Group Science, Engineering and Technology, KU Leuven–Kulak, E. Sabbelaan 53, 8500 Kortrijk, Belgium

3 Department of Electrical Engineering, ESAT/STADIUS KU Leuven, Kasteelpark Arenberg 10, bus 2446, 3001 Leuven-Heverlee, Belgium

4 iMinds Medical IT, Leuven, Belgium

Published online: 12 May 2015

we apply block term decompositions to the problem of blind multi-user detection-estimation in DS-CDMA communication to demonstrate that exact line and plane search can significantly reduce computation time of the workhorse tensor decomposition algorithm alternating least squares.

**Keywords** Exact line search · Exact plane search · Tensor decomposition · Tensor optimization · Bivariate polynomial system

## 1 Introduction

Line and plane search algorithms are an integral component of many optimization algorithms, both for accelerating and guaranteeing convergence to an optimum. In this paper, we develop several types of exact line and plane search algorithms for a broad class of optimization problems, dubbed *tensor optimization* problems, in which the residual function is a polynomial tensor. In fact, these search algorithms may just as well be applied to the more general class of polynomial optimization problems, in which the objective function and constraints are (possibly polyanalytic) polynomials. However, we restrict our discussion to tensor optimization, which is interesting in and of itself and comprises applications ranging from tensor decompositions [10,33] to data fusion [2] and least squares support tensor machines [55]. Line and plane searches can often significantly reduce the length of so-called swamps [4,31,43,46], in which the optimization algorithm's convergence almost slows down to a halt.

The line and plane searches we develop are not limited to tasks which can be formulated as tensor optimization problems. In general, any nonlinear least squares problem can benefit from exact line or plane search by using a higher-order Taylor series expansion as an approximation of its residual function, effectively turning the nonlinear least squares problem into a tensor optimization problem.

Furthermore, we propose a variant of these search problems called *scaled* line and plane search which allows for an optimal scaling of the solution while simultaneously minimizing the objective function along a line or on a plane of search directions. Remarkably, we will see that for tensor optimization problems with a homogeneous residual function, the optimal scaling parameter can be obtained for almost no additional cost compared to a regular line or plane search.

Finally, we show that tensor optimization line and plane search subproblems are tantamount to solving a bivariate or polyanalytic polynomial system and present a method to compute the isolated solutions of bivariate and polyanalytic polynomial systems using a single generalized eigenvalue decomposition. In our numerical experiments, we compare a total of six line and plane search algorithms to accelerate complex tensor decomposition algorithms for blind detection-estimation of multi-user DS-CDMA communication using structured rank-$(L_r, L_r, 1)$ block term decompositions [19,42]. The search algorithms are able to reduce the CPU time of the baseline algorithm, alternating least squares, by between 50 and 75 %.

The paper is organized as follows. In the next subsection we review our notation and introduce some basic definitions. Section 2 introduces tensor optimization and how it can be used to compute various tensor decompositions. In Sect. 3 we define

the tensor optimization (scaled) line and plane search subproblems, explain what type of objective functions they correspond to and show that in the case of a scaled line or plane search, the scaling parameter can be obtained at little additional cost. In Sect. 4 we present a method to compute the isolated real solutions of bivariate polynomial systems and isolated complex solutions of polyanalytic polynomial systems using a single generalized eigenvalue decomposition. In Sect. 5 we evaluate the performance of the proposed search algorithms with Monte Carlo simulations in which we blindly separate and equalize multi-user DS-CDMA signals using structured rank-$(L_r, L_r, 1)$ block term decompositions. We conclude the paper in Sect. 6.

### 1.1 Notation and preliminaries

A tensor is an element of a tensor product of vector spaces. In this article, we refer to a tensor represented as a multidimensional array, given a choice of bases for each of these vector spaces. The order, or the number of modes, of a tensor is the number of indices associated with each element of that tensor. Vectors are denoted by boldface letters and are lower case, e.g., $\boldsymbol{a}$. Matrices are usually denoted by capital letters, e.g., $A$. Higher-order tensors are denoted by Euler script letters, e.g., $\mathscr{A}$. An entry of a vector $\boldsymbol{a}$, matrix $A$ or tensor $\mathscr{A}$ is denoted by $a_i$, $a_{ij}$ or $a_{ijk\ldots}$, depending on the number of modes. A colon is used to select all entries of a mode. For instance, $\boldsymbol{a}_{:j}$ corresponds to the $j$th column of a matrix $A$. When there is no confusion, we also use $\boldsymbol{a}_j$ to denote the $j$th column of the matrix $A$. Sequences are denoted by a superscript in parentheses, e.g., $\{A^{(n)}\}_{n=1}^{N}$. The superscripts $\cdot^{\mathrm{T}}$, $\cdot^{\mathrm{H}}$, $\cdot^{-1}$ and $\cdot^{\dagger}$ are used for the transpose, Hermitian conjugate, matrix inverse and Moore–Penrose pseudoinverse, respectively. The identity matrix of order $n$ is denoted by $\mathbb{I}_n$ and the complex conjugate is denoted by an overbar, e.g., $\overline{a}$ is the complex conjugate of the scalar $a$. We use parentheses to denote the concatenation of two or more vectors, e.g., $(\boldsymbol{a}, \boldsymbol{b})$ is equivalent to $[\boldsymbol{a}^{\mathrm{T}}\ \boldsymbol{b}^{\mathrm{T}}]^{\mathrm{T}}$.

The vec$(\cdot)$ operator represents the column-wise vectorization of matrices and tensors. Let $\mathscr{T} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ and $\mathscr{U} \in \mathbb{C}^{J_1 \times \cdots \times J_N}$ be two $N$th-order tensors. In a mode-$n$ matricization $T_{(n)} \in \mathbb{C}^{I_n \times I_1 \cdots I_{n-1} I_{n+1} \cdots I_N}$ of the tensor $\mathscr{T}$, tensor element with indices $(i_1, \ldots, i_N)$ is mapped to matrix element $(i_n, j)$ such that

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^{N} (i_k - 1) J_k \text{ with } J_k = \prod_{\substack{m=0 \\ m \neq n}}^{k-1} I_m,$$

wherein we define $I_0 := 1$. The inner product of $\mathscr{T}$ and $\mathscr{U}$ (assuming $I_n \equiv J_n$) is defined as $\langle \mathscr{T}, \mathscr{U} \rangle = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} \overline{t}_{i_1 \cdots i_N} u_{i_1 \cdots i_N}$. The Frobenius norm is then given by $\|\mathscr{T}\|_{\mathrm{F}} = \sqrt{\langle \mathscr{T}, \mathscr{T} \rangle}$. The outer product $\mathscr{T} \circ \mathscr{U}$ is the tensor defined by $(\mathscr{T} \circ \mathscr{U})_{i_1 \cdots i_P j_1 \cdots j_Q} = t_{i_1 \cdots i_P} u_{j_1 \cdots j_Q}$. The mode-$n$ product $\mathscr{T} \cdot_n A$ of the tensor $\mathscr{T}$ with a matrix $A \in \mathbb{C}^{J \times I_n}$ is the tensor defined by $(\mathscr{T} \cdot_n A)_{(n)} = A \cdot T_{(n)}$.

We will associate a coefficient vector $\boldsymbol{p} \in \mathbb{C}^{d_p+1}$ and coefficient matrices $Q \in \mathbb{C}^{(d_q^{(y)}+1) \times (d_q^{(x)}+1)}$ and $R \in \mathbb{C}^{(d_r^{(\overline{z})}+1) \times (d_r^{(z)}+1)}$ with polynomials $p(z)$, bivariate polynomials $q(x, y)$ and polyanalytic polynomials $r(z, \overline{z})$, defined by

$$p(z) := \begin{bmatrix} 1 & z & \cdots & z^{d_p} \end{bmatrix} \boldsymbol{p},$$
$$q(x, y) := \begin{bmatrix} 1 & y & \cdots & y^{d_q^{(y)}} \end{bmatrix} Q \begin{bmatrix} 1 & x & \cdots & x^{d_q^{(x)}} \end{bmatrix}^{\mathrm{T}} \text{ and}$$
$$r(z, \overline{z}) := \begin{bmatrix} 1 & \overline{z} & \cdots & \overline{z}^{d_r^{(\overline{z})}} \end{bmatrix} R \begin{bmatrix} 1 & z & \cdots & z^{d_r^{(z)}} \end{bmatrix}^{\mathrm{T}},$$

respectively. The term polyanalytic refers to the fact that the function is not analytic in its argument $z$, but is analytic when viewed as a function of its argument $z$ and its complex conjugate $\overline{z}$ as a whole [52]. Here, $p_{d_p+1} \neq 0$ and $\max_i |q_{i,d_q^{(x)}+1}| \cdot \max_j |q_{d_q^{(y)}+1,j}| \neq 0$, so that $d_p$ is the degree of $p(z)$ and $q_{d_q^{(x)}}$ and $q_{d_q^{(y)}}$ are the coordinate degrees of $Q$. An analogous condition holds for $R$, so that $d_r^{(z)}$ and $d_r^{(\overline{z})}$ are its coordinate degrees. The total degree of the polynomial $q(x, y)$ ($r(z, \overline{z})$) is defined as the largest value $i + j$ for which $q_{i+1,j+1} \neq 0$ ($r_{i+1,j+1} \neq 0$).

## 2 Tensor optimization

Tensor optimization is a special case of the more general class of polynomial optimization, in which the objective function and constraints are polynomials. Let $\mathscr{T}$ be a given tensor in $\mathbb{C}^{I_1 \times \cdots \times I_N}$ and let $\mathscr{M}(\boldsymbol{z}) : \mathbb{C}^K \to \mathbb{C}^{I_1 \times \cdots \times I_N}$ be a polynomial tensor-valued function. A polynomial tensor is a tensor whose elements are (possibly polyanalytic) multivariate polynomials. The degree $d_{\mathscr{M}}$ of a polynomial tensor $\mathscr{M}(\boldsymbol{z})$ is the maximum total degree of its elements. If each element's total degree is equal to $d_{\mathscr{M}}$, $\mathscr{M}$ is called homogeneous. Unconstrained tensor optimization problems considered here are nonlinear least squares problems with a polynomial tensor-valued residual function $\mathscr{F}$ of the form

$$\underset{\boldsymbol{z} \in \mathbb{C}^K}{\text{minimize}} \ \frac{1}{2} \|\mathscr{F}(\boldsymbol{z})\|_{\mathrm{F}}^2, \quad \text{where } \mathscr{F}(\boldsymbol{z}) := \mathscr{M}(\boldsymbol{z}) - \mathscr{T}. \tag{TO}$$

Among others, this model comprises the canonical polyadic decomposition (CPD) [8,31], low multilinear rank approximation or Tucker approximation (LMLRA) [33, 56,61], block term decompositions (BTD) [14,15,20], data fusion as coupled matrix and tensor factorizations (CMTF) [2] and least squares support tensor machines (LS-STM) [55]. Due to the inner product, the objective function in (TO) is polyanalytic in $\boldsymbol{z}$, meaning it is function of both $\boldsymbol{z}$ and $\overline{\boldsymbol{z}}$ and its Taylor series in $\boldsymbol{z}$ and $\overline{\boldsymbol{z}}$ is convergent for every point on its domain. More specifically, the objective function is a polyanalytic polynomial of total degree $2d_{\mathscr{M}}$ in $K$ complex variables.

For instance, the CPD approximates a tensor by a sum of $R$ rank-one tensors. Let $A^{(n)} \in \mathbb{C}^{I_n \times R}$, $n = 1, \ldots, N$, consist of nonzero columns, then

$$\mathscr{T} \approx \mathscr{M}_{\mathrm{CPD}}(\boldsymbol{z}) := \sum_{r=1}^{R} \boldsymbol{a}_r^{(1)} \circ \cdots \circ \boldsymbol{a}_r^{(N)} \tag{CPD}$$

is a CPD of $\mathscr{T}$ in $R$ rank-one tensors. The model $\mathscr{M}_{\mathrm{CPD}}$ has degree $N$ and its argument $\boldsymbol{z} \in \mathbb{C}^{\sum I_n R}$ concatenates the factor matrices as $(\mathrm{vec}(A^{(1)}), \ldots, \mathrm{vec}(A^{(N)}))$. In

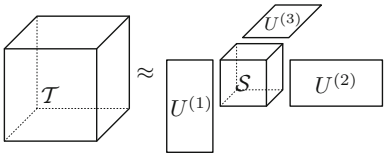



**Fig. 1** A LMLRA of a third-order tensor

principle, the decomposition is only called canonical if it is exact and $R$ is minimal. However, in the approximation sense the term can be interpreted to mean that $R$ is the smallest integer for which the decomposition approximates the tensor sufficiently well.

To a large extent, the practical importance of the CPD stems from the fact that it is unique under relatively mild conditions. A CPD is called (essentially) unique when it is subject only to a permutation and scaling ambiguity. It is clear that one can arbitrarily permute the different rank-one terms. Also, the vectors in a single rank-one term may be arbitrarily scaled, as long as their product remains the same. The most well-known sufficient condition for uniqueness is due to Kruskal [34,35] and more recent results on uniqueness can be found in [24,25]. This striking property has been exploited in a wide range of applications in signal processing [10], including blind source separation [11,12,17], blind multi-user detection-estimation in DS-CDMA communication [1,51], multiple-invariance sensor array processing [49], 3D radar clutter modeling and mitigation [44] and multi-dimensional harmonic retrieval [49].

One could also consider isolating the $r$th rank-one tensor's norm in the real scalar $\lambda_r$ and constraining the vectors $\boldsymbol{a}_r^{(n)}$ to have unit norm, leading to the modified model

$$\mathscr{M}_{\text{mCPD}} := \sum_{r=1}^{R} \lambda_r \cdot \boldsymbol{a}_r^{(1)} \circ \cdots \circ \boldsymbol{a}_r^{(N)} \tag{mCPD}$$

of degree $N+1$. This form of the CPD is useful in the context of CMTF, as we will see later. Although the scaling ambiguity is largely resolved in the modified model, it is still subject to a phase ambiguity if the decomposition is complex.

In a LMLRA, a tensor is approximated by a tensor of low multilinear rank (cf. Fig. 1). Let $\mathscr{S} \in \mathbb{C}^{R_1 \times \cdots \times R_N}$ and let $U^{(n)} \in \mathbb{C}^{I_n \times R_n}$, $n = 1, \ldots, N$, then

$$\mathscr{T} \approx \mathscr{M}_{\text{LMLRA}}(\boldsymbol{z}) := \mathscr{S} \cdot_1 U^{(1)} \cdot_2 \cdots \cdot_N U^{(N)} \tag{LMRLA}$$

is a rank-$(R_1, \ldots, R_N)$ LMLRA of $\mathscr{T}$. The model $\mathscr{M}_{\text{LMLRA}}(\boldsymbol{z})$ now has degree $N+1$ and its argument $\boldsymbol{z} \in \mathbb{C}^{\prod R_n + \sum I_n R_n}$ concatenates the core tensor $\mathscr{S}$ and factor matrices $U^{(n)}$ as $(\text{vec}(\mathscr{S}), \text{vec}(U^{(1)}), \ldots, \text{vec}(U^{(N)}))$. The LMLRA offers a way to generalize subspace problems in signal processing to multi-way data, with applications in dimensionality reduction [21], feature extraction and classification [30,38,57] and noise reduction [32].

The block term decompositions framework unifies the CPD and LMLRA by approximating tensors with a sum of low multilinear rank terms. Of particular interest is the rank-$(L_r, L_r, 1)$ BTD, which approximates a tensor by a sum of outer products of rank-$L_r$ matrices and nonzero vectors (cf. Fig. 2). The rank-$(L_r, L_r, 1)$ BTD combines the

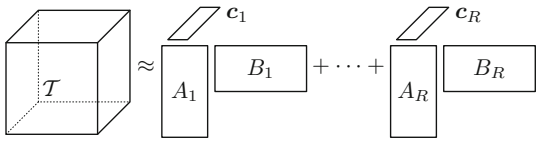



**Fig. 2** A rank-$(L_r, L_r, 1)$ BTD of a third-order tensor

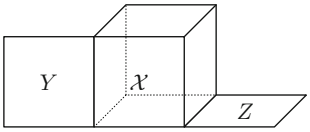



**Fig. 3** Coupled data sets of different orders. The tensor $\mathscr{X}$ has modes restaurant × meal × customer and contains ratings assigned by a customer to a meal at a certain restaurant. The matrix $Y$ has modes restaurant × category, is coupled along the first mode with $\mathscr{X}$ and contains information about the type of restaurant. The matrix $Z$ has modes customer × customer, is coupled along the third mode with $\mathscr{X}$ and contains social network information between customers

mild conditions for uniqueness of the CPD [15,16] with the more general low-rank structure of a LMLRA to form a promising new candidate for blind source separation [16,18,19,42,43,49–51]. Let $A_r \in \mathbb{C}^{I_1 \times L_r}$ and $B_r \in \mathbb{C}^{I_2 \times L_r}$, $r = 1, \ldots, R$, have rank $L_r$ and let $C \in \mathbb{C}^{I_3 \times R}$ consist of nonzero columns, then

$$\mathscr{T} \approx \mathscr{M}_{\mathrm{BTD}}(z) := \sum_{r=1}^{R} \left(A_r \cdot B_r^{\mathrm{T}}\right) \circ c_r \tag{BTD}$$

is a rank-$(L_r, L_r, 1)$ BTD of the third-order tensor $\mathscr{T}$. The model $\mathscr{M}_{\mathrm{BTD}}$ has degree three and its argument $z \in \mathbb{C}^{(I_1+I_2)\sum L_r + I_3 R}$ concatenates the factor matrices as $(\mathrm{vec}(A_1), \ldots, \mathrm{vec}(A_R), \mathrm{vec}(B_1), \ldots, \mathrm{vec}(B_R), \mathrm{vec}(C))$.

The CPD, LMLRA and rank-$(L_r, L_r, 1)$ BTD are all examples of multilinear polynomial tensors. A function is said to be multilinear in its argument $z$ if for all $i$ it is linear in $z_i$ when the remaining variables $z_j$ $(j \neq i)$ are fixed. However, this property can be lost by imposing symmetry. For example, INDSCAL [8] is a special case of the CPD for third-order tensors in which $A^{(1)} \equiv A^{(2)}$. Its model is a polynomial tensor, but is no longer multilinear.

In CMTF, the objective is to minimize a weighted sum of tensor optimization objective functions $\sum_k \lambda_k \|\mathscr{F}_k(z)\|_{\mathrm{F}}^2$, wherein $\mathscr{F}_k$ are coupled polynomial tensor-valued residual functions. Data fusion in the form of CMTF is readily cast as a tensor optimization problem by vectorizing and concatenating the residual functions $\mathscr{F}_k$ into a single polynomial tensor-valued residual function as

$$\mathscr{F}_{\mathrm{CMTF}}(z) := \left(\sqrt{\lambda_1}\,\mathrm{vec}(\mathscr{F}_1(z)), \ldots, \sqrt{\lambda_K}\,\mathrm{vec}(\mathscr{F}_K(z))\right).$$

The polynomial tensor-valued function $\mathscr{F}_{\mathrm{CMTF}}$ can then be separated into a model $\mathscr{M}_{\mathrm{CMTF}}(z)$ and a constant $\mathscr{T}$ as in (TO). The degree of the model $\mathscr{M}_{\mathrm{CMTF}}$ is then equal to the maximum degree of its constituent models.

Consider for example the coupled dataset shown in Fig. 3. One way to jointly analyze this dataset would be to compute a CMTF consisting of three modified CPD

models (mCPD) as follows. If the tensor $\mathscr{X}$ is decomposed using the factor matrices $A^{(1)}$, $A^{(2)}$, $A^{(3)}$ and scaling vector $\boldsymbol{\lambda}$, the matrices $Y$ and $Z$ can be coupled with $\mathscr{X}$ by decomposing them using the modified CPD models $B \cdot \operatorname{diag}(\boldsymbol{\mu}) \cdot A^{(1)\mathrm{T}}$ and $A^{(3)} \cdot \operatorname{diag}(\boldsymbol{\nu}) \cdot C^{\mathrm{T}}$, respectively. The absolute values of the scaling vectors $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ give an indication of how strongly certain factors are shared between the two matrices and the tensor. Note that although we have treated the three scaling vectors separately from the factor matrices, they may equally well be interpreted as factor matrices of canonical polyadic decompositions of tensors of one order higher than shown in Fig. 3. More specifically, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are the factor matrices corresponding to the singleton dimension following the last nonsingleton dimension of $\mathscr{X}$, $Y$ and $Z$, respectively.

## 3 Exact line and plane search

In optimization, two-dimensional subspace minimization is a prevalent subproblem in trust region globalization strategies. Given a vector-valued residual function $F$, a current iterate $z_k$ and two search directions $\Delta z_1$ and $\Delta z_2$, the objective is to minimize $\|F(z)\|$ in the plane spanned by the two search directions, subject to a norm constraint on the step. The corresponding optimization problem

$$\begin{aligned} \underset{\alpha,\,\beta\,\in\,\mathbb{R}}{\text{minimize}} \quad & \frac{1}{2}\,\|F(z_k + \alpha\Delta z_1 + \beta\Delta z_2)\|^2 \\ \text{subject to} \quad & \|\alpha\Delta z_1 + \beta\Delta z_2\|^2 \leq \delta^2 \end{aligned}$$

is often simplified using a linear approximation of $F$ around $z_k$ in order to make the problem tractable. The resulting Gauss–Newton two-dimensional subspace minimization problem [7]

$$\begin{aligned} \underset{\alpha,\,\beta\,\in\,\mathbb{R}}{\text{minimize}} \quad & \frac{1}{2}\left\| F(z_k) + \frac{dF}{dz}(z_k) \cdot (\alpha\Delta z_1 + \beta\Delta z_2) \right\|^2 \\ \text{subject to} \quad & \|\alpha\Delta z_1 + \beta\Delta z_2\|^2 \leq \delta^2 \end{aligned}$$

is equivalent to solving a bivariate polynomial system with polynomials of total degree two. Instead of a linear approximation, one could consider using a second-, third- or higher-order Taylor series approximation of $F$. In that case, the approximation is a polynomial tensor, and we will see that the two-dimensional subspace minimization problem still leads to a bivariate polynomial system, but where one of the two polynomials now has a total degree higher than two.

From here on, $\mathscr{F}$ is assumed to be a polynomial tensor (approximation of a residual function). Given a current iterate $z_k$ and two descent directions $\Delta z_1$ and $\Delta z_2$, we refer to the following tensor optimization subproblems

$$\underset{\alpha}{\text{minimize}} \quad \frac{1}{2}\|\mathscr{F}(z_k + \alpha\Delta z_1)\|_{\mathrm{F}}^2 \tag{LS}$$

$$\underset{\alpha,\gamma}{\text{minimize}} \ \frac{1}{2}\|\mathscr{F}(\gamma z_k + \alpha \Delta z_1)\|_{\text{F}}^2 \tag{SLS}$$

$$\underset{\alpha,\beta}{\text{minimize}} \ \frac{1}{2}\|\mathscr{F}(z_k + \alpha \Delta z_1 + \beta \Delta z_2)\|_{\text{F}}^2 \tag{PS}$$

$$\underset{\alpha,\beta,\gamma}{\text{minimize}} \ \frac{1}{2}\|\mathscr{F}(\gamma z_k + \alpha \Delta z_1 + \beta \Delta z_2)\|_{\text{F}}^2 \tag{SPS}$$

as a line search (LS), scaled line search (SLS), plane search (PS) and scaled plane search (SPS), respectively. We append $\mathbb{R}$ or $\mathbb{C}$ to the optimization problem's acronym to distinguish over which field the solution is sought, e.g., a real line search is denoted by LS-$\mathbb{R}$, while a complex scaled line search would be denoted by SLS-$\mathbb{C}$.

Line and plane search problems play an important part in algorithms that solve tensor optimization problems. Many optimization-based algorithms rely on either line or plane searches [45] and other popular algorithms for tensor decompositions such as alternating least squares have been combined with line searches to mitigate the effect of so-called swamps [4,36,40,46]. In particular, LS-$\mathbb{R}$ and LS-$\mathbb{C}$ have received some attention recently in the case of a CPD [9,43,46].

In signal processing, the recovery of data symbols transmitted through a distorting medium is a fundamental problem. Assuming that zero-mean data symbols $s_n$ are transmitted through a linear time-invariant time-dispersive channel of order $M$, the channel output measured with an antenna array of length $P$ can be modelled as

$$\boldsymbol{x}_n = \sum_{k=0}^{M} \boldsymbol{h}_k s_{n-k} + \boldsymbol{v}_n, \tag{1}$$

where $H = \begin{bmatrix}\boldsymbol{h}_0 \cdots \boldsymbol{h}_M\end{bmatrix} \in \mathbb{C}^{P\times(M+1)}$ contains the channel impulse responses for each of the $P$ antennas and $\boldsymbol{v}_n \in \mathbb{C}^{P\times 1}$ is an additive noise term. The single-input multiple-output (SIMO) model (1) reduces to the single-input single-output (SISO) model for $P = 1$. The SIMO model can also be obtained by exploiting temporal diversity (e.g., time oversampling) instead of spatial diversity (e.g., an antenna array), and can easily be extended to the multiple-input (MIMO) case. To recover the data symbols $s_n$, a linear equalizer $\boldsymbol{f} \in \mathbb{C}^{PL}$ with finite impulse response spanning $L$ measurements is used which produces the output symbols $y_n := \boldsymbol{f}^{\text{H}}\tilde{\boldsymbol{x}}_n$, where $\tilde{\boldsymbol{x}}_n := (\boldsymbol{x}_n, \ldots, \boldsymbol{x}_{n-L+1})$. One of the most widespread blind channel equalization principles is the constant modulus (CM) criterion [28], which looks for the filter $\boldsymbol{f}$ which minimizes a tensor optimization problem (TO) with a residual function $\mathscr{F}_{\text{CM}}$ defined by

$$(\mathscr{F}_{\text{CM}}(\boldsymbol{f}))_n := |y_n|^2 - 1.$$

Furthermore, it is not hard to show that the steepest ascent direction at the $k$th iterate $\boldsymbol{f}_k$ of the tensor optimization objective function is given by

$$\boldsymbol{g}_k := 2\left(\frac{d\mathscr{F}_{\text{CM}}}{d\boldsymbol{f}}(\boldsymbol{f}_k)\right)^{\text{H}} \cdot \mathscr{F}_{\text{CM}}(\boldsymbol{f}_k).$$

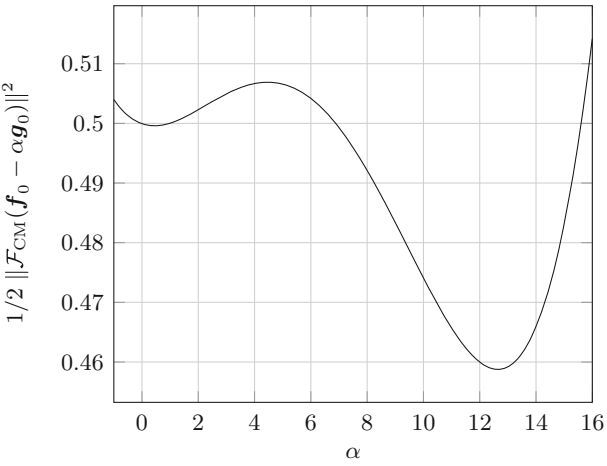



**Fig. 4** The LS-$\mathbb{R}$ objective function of the constant modulus problem along the steepest descent direction $-\boldsymbol{g}_0$ at the iterate $\boldsymbol{f}_0$ defined by the measurements $\boldsymbol{x}_2$, $\boldsymbol{x}_1$ and $\boldsymbol{x}_0$

The CM criterion often suffers from the existence of numerous local minima, usually associated with different equalization delays, in the equalizer parameter space [22]. An exact line search may be useful to help reduce the chances of converging to an undesired local minimum. By means of an example, we choose $M = 1$, $P = 2$ and $L = 2$ and consider the measurements $\boldsymbol{x}_2 := (1 + i, -i)$, $\boldsymbol{x}_1 := (1, -1)$ and $\boldsymbol{x}_0 := (1, 1-i)$ and an initial filter $\boldsymbol{f}_0 = (0.2413+0.1593i, 0.3496+0.0858i, 0.2875 + 0.3305i, 0.5979 + 0.5929i)$. The line search objective function (LS) of the CM tensor optimization problem along the steepest descent direction $-\boldsymbol{g}_0$ at $\boldsymbol{f}_0$ is the polynomial of degree four shown in Fig. 4. Whereas an approximate line search will likely choose a suboptimal line search parameter $\alpha$ near 0.4, an exact line search would find the global minimizer $\alpha^* \approx 12.6429$. Previous work has shown that the optimal line search parameter can be obtained by computing the roots of a polynomial of degree three [60]. Viewed in the context of tensor optimization, scaled line searches and even plane searches may also be applied efficiently to the problem of constant modulus equalization.

Table 1 gives an overview of the type of objective function the different kinds of search problems correspond to for the two most common choices of the solution field. From the table, it is clear that LS-$\mathbb{R}$ and SLS-$\mathbb{R}$ are easy problems to solve since their stationary points are just the roots of the derivative of their polynomial and rational objective functions, respectively. Obtaining the global minimizer of the other four search problems is less straightforward because their objective functions are either bivariate or polyanalytic. Remarkably, the scaling parameter $\gamma$ in SLS and SPS can be obtained at almost no additional cost compared to LS and PS, assuming the objective function is homogeneous and more expensive to compute than solving the underlying polynomial system.

The following two subsections explain how to arrive at the type of objective function given by Table 1. The last subsection describes how a norm constraint on the step can

**Table 1** The type of the objective function each of the search problems correspond to for two choices of the solution field

| | Field $\mathbb{R}$ | Field $\mathbb{C}$ |
|---|---|---|
| Problem LS | Degree $2d_{\mathscr{M}}$ polynomial | Coordinate degree $d_{\mathscr{M}}$ polyanalytic polynomial |
| Problem SLS[a] | Degree $2d_{\mathscr{M}}$ rational function | Coordinate degree $d_{\mathscr{M}}$ polyanalytic rational function |
| Problem PS | Total degree $2d_{\mathscr{M}}$ bivariate polynomial | —[b] |
| Problem SPS[a] | Total degree $2d_{\mathscr{M}}$ bivariate rational function | —[b] |

[a] Assuming $\mathscr{M}$ is homogeneous
[b] Not discussed in this article

be imposed on PS-$\mathbb{R}$, which generalizes two-dimensional subspace minimization from a linear approximation of the residual function to a polynomial tensor approximation such as a higher-order Taylor series expansion. Section 4 then describes how to obtain the global minimizers of (S)LS-$\mathbb{C}$ and (S)PS-$\mathbb{R}$ by recasting them as a generalized eigenvalue problem.

### 3.1 (Scaled) line search

Since LS-$\mathbb{R}$, LS-$\mathbb{C}$ and SLS-$\mathbb{R}$ are special cases of SLS-$\mathbb{C}$, we first focus on SLS-$\mathbb{C}$. Separate $\mathscr{F}$ into a model $\mathscr{M}$ and a constant $\mathscr{T}$ as in (TO) and expand the objective function in (SLS) as

$$\frac{1}{2}\big(\|\mathscr{T}\|_{\mathrm{F}}^2 + \|\mathscr{M}(\gamma \mathcal{Z}_k + \alpha \Delta \mathcal{Z}_1)\|_{\mathrm{F}}^2 - \langle \mathscr{T}, \mathscr{M}(\gamma \mathcal{Z}_k + \alpha \Delta \mathcal{Z}_1)\rangle - \langle \mathscr{M}(\gamma \mathcal{Z}_k + \alpha \Delta \mathcal{Z}_1), \mathscr{T}\rangle\big).$$

Under the assumption that $\mathscr{M}$ is homogeneous[1] with degree $d_{\mathscr{M}}$, we introduce a change of variables $c := \gamma^{d_{\mathscr{M}}}$ and $a := \alpha/\gamma$ so the objective function can be written as

$$f(a, \overline{a}, c, \overline{c}) := \frac{1}{2}\big(\|\mathscr{T}\|_{\mathrm{F}}^2 + |c|^2 q(a, \overline{a}) - cp(a) - \overline{c}\,\overline{p(a)}\big),$$

where $p(a) := \langle \mathscr{T}, \mathscr{M}(\mathcal{Z}_k + a\Delta \mathcal{Z}_1)\rangle$ and $q(a, \overline{a}) := \|\mathscr{M}(\mathcal{Z}_k + a\Delta \mathcal{Z}_1)\|_{\mathrm{F}}^2$ are two univariate polynomials. In the case of a line search, $c \equiv \gamma \equiv 1$, $a \equiv \alpha$ and (LS) is equivalent to

[1] The homogeneity requirement, which is satisfied by most tensor decompositions, is only necessary for scaled line and plane search.

$$\underset{\alpha \in \mathbb{R}}{\text{minimize}} \;\; f(\alpha) \tag{LS-$\mathbb{R}$}$$

$$\underset{\alpha \in \mathbb{C}}{\text{minimize}} \;\; f(\alpha, \overline{\alpha}) \tag{LS-$\mathbb{C}$}$$

for LS-$\mathbb{R}$ and LS-$\mathbb{C}$, respectively. In the case of a scaled line search, notice that $f$ is quadratic in $c$ so that a global optimum $(a^*, c^*)$ must satisfy $c^* = \mathrm{Re}\{p(a^*)\}/q(a^*)$ and $c^* = \overline{p(a^*)}/q(a^*, \overline{a}^*)$ for SLS-$\mathbb{R}$ and SLS-$\mathbb{C}$, respectively. Substituting back in $f$ reveals that $a^*$ is the solution of

$$\underset{a \in \mathbb{R}}{\text{minimize}} \;\; -\frac{\mathrm{Re}\{p(a)\}^2}{q(a)} \tag{SLS-$\mathbb{R}$}$$

$$\underset{a \in \mathbb{C}}{\text{minimize}} \;\; -\frac{|p(a)|^2}{q(a, \overline{a})} \tag{SLS-$\mathbb{C}$}$$

for SLS-$\mathbb{R}$ and SLS-$\mathbb{C}$, respectively.

### 3.2 (Scaled) plane search

Analogously to scaled line search, expand the objective function in (SPS) as

$$\begin{aligned}\frac{1}{2}\big(&\|\mathscr{T}\|_{\mathrm{F}}^2 + \|\mathscr{M}(\gamma \mathcal{Z} + \alpha \Delta \mathcal{Z}_1 + \beta \Delta \mathcal{Z}_2)\|_{\mathrm{F}}^2 \\ &- \langle \mathscr{T}, \mathscr{M}(\gamma \mathcal{Z} + \alpha \Delta \mathcal{Z}_1 + \beta \Delta \mathcal{Z}_2)\rangle \\ &- \langle \mathscr{M}(\gamma \mathcal{Z} + \alpha \Delta \mathcal{Z}_1 + \beta \Delta \mathcal{Z}_2), \mathscr{T}\rangle\big)\end{aligned}$$

and assume that $\mathscr{M}$ is homogeneous with degree $d_{\mathscr{M}}$ to introduce a change of variables $c := \gamma^{d_{\mathscr{M}}}$, $a := \alpha/\gamma$ and $b := \beta/\gamma$. For real $\alpha$, $\beta$ and $\gamma$, the objective function can then be written as

$$f(a, b, c) := \frac{1}{2}\big(\|\mathscr{T}\|_{\mathrm{F}}^2 + c^2 q(a, b) - c(p(a, b) + \overline{p(a, b)})\big),$$

where $p(a, b) := \langle \mathscr{T}, \mathscr{M}(\mathcal{Z} + a\Delta \mathcal{Z}_1 + b\Delta \mathcal{Z}_2)\rangle$ and $q(a, b) := \|\mathscr{M}(\mathcal{Z} + a\Delta \mathcal{Z}_1 + b\Delta \mathcal{Z}_2)\|_{\mathrm{F}}^2$ are two bivariate polynomials. In the case of a plane search, $c \equiv \gamma \equiv 1$, $a \equiv \alpha$, $b \equiv \beta$ and (PS) is equivalent to

$$\underset{\alpha, \beta \in \mathbb{R}}{\text{minimize}} \;\; f(\alpha, \beta). \tag{PS-$\mathbb{R}$}$$

In the case of a scaled plane search, notice that $f$ is quadratic in $c$ so that a global optimum $(a^*, b^*, c^*)$ must satisfy $c^* = \mathrm{Re}\{p(a^*, b^*)\}/q(a^*, b^*)$. Substituting back in $f$ reveals that $(a^*, b^*)$ is the solution of

$$\underset{a, b \in \mathbb{R}}{\text{minimize}} \;\; -\frac{\mathrm{Re}\{p(a, b)\}^2}{q(a, b)}. \tag{SPS-$\mathbb{R}$}$$

### 3.3 Higher-order two-dimensional subspace minimization

When $\mathscr{F}$ is a polynomial tensor approximation of a nonlinear tensor-valued residual function $F$, it may be desirable to impose a norm constraint on the plane search step to avoid the next iterate from straying too far from where $\mathscr{F}$ is considered to be an accurate approximation of $F$. For example, $\mathscr{F}$ could be a higher-order Taylor series expansion of $F$. Given a trust-region radius $\delta \in \mathbb{R}_+^0$, the higher-order two-dimensional subspace minimization problem is defined as

$$\begin{aligned} \underset{\alpha,\,\beta\,\in\,\mathbb{R}}{\text{minimize}} \quad & \frac{1}{2}\|\mathscr{F}(\mathcal{Z}_k + \alpha\Delta\mathcal{Z}_1 + \beta\Delta\mathcal{Z}_2)\|_{\mathrm{F}}^2 \\ \text{subject to} \quad & \|\alpha\Delta\mathcal{Z}_1 + \beta\Delta\mathcal{Z}_2\|^2 \leq \delta^2. \end{aligned}$$

The case where the norm constraint is not active corresponds to solving PS-$\mathbb{R}$. When the norm constraint is active, it can be incorporated into the objective function using the method of Lagrange multipliers. Write $\mathscr{F}(\mathcal{Z})$ as $\mathscr{M}(\mathcal{Z}) - \mathscr{T}$, let $p(\alpha, \beta) := \langle \mathscr{T}, \mathscr{M}(\mathcal{Z} + \alpha\Delta\mathcal{Z}_1 + \beta\Delta\mathcal{Z}_2)\rangle$, $q(\alpha, \beta) := \|\mathscr{M}(\mathcal{Z} + \alpha\Delta\mathcal{Z}_1 + \beta\Delta\mathcal{Z}_2)\|_{\mathrm{F}}^2$ and $r(\alpha, \beta) := \|\alpha\Delta\mathcal{Z}_1 + \beta\Delta\mathcal{Z}_2\|^2$ and introduce the Lagrange multiplier $\lambda$ to obtain the Lagrangian

$$\frac{1}{2}\big(\|\mathscr{T}\|_{\mathrm{F}}^2 + q(\alpha, \beta) - p(\alpha, \beta) - \overline{p(\alpha, \beta)}\big) + \frac{\lambda}{2}\big(r(\alpha, \beta) - \delta^2\big).$$

Let $s(\alpha, \beta) := q(\alpha, \beta) - 2\mathrm{Re}\{p(\alpha, \beta)\}$. Setting the Lagrangian's gradient equal to zero and eliminating $\lambda$, it follows that the global minimizer $(\alpha^*, \beta^*)$ is a solution of the bivariate polynomial system

$$\begin{cases} \frac{\partial r(\alpha,\beta)}{\partial \alpha}\frac{\partial s(\alpha,\beta)}{\partial \beta} - \frac{\partial r(\alpha,\beta)}{\partial \beta}\frac{\partial s(\alpha,\beta)}{\partial \alpha} = 0 \\ r(\alpha, \beta) - \delta^2 = 0 \end{cases} \tag{2}$$

in which the polynomials are of total degree $2d_{\mathscr{M}}$ and 2, respectively.

## 4 Solving systems of bivariate and polyanalytic polynomials

The two search problems LS-$\mathbb{R}$ and SLS-$\mathbb{R}$ are easily solved by computing the roots of their objective function's derivative, which is simply a univariate polynomial. On the other hand, the four search problems (S)LS-$\mathbb{C}$ and (S)PS-$\mathbb{R}$ give rise to polyanalytic and bivariate polynomial systems, respectively (cf. Table 1). For example, the LS-$\mathbb{C}$ objective function is a polyanalytic polynomial $f(\alpha, \overline{\alpha})$ of coordinate degree $d_{\mathscr{M}}$ and its stationary points are the solutions of the polyanalytic polynomial system

$$\frac{\partial f(\alpha, \overline{\alpha})}{\partial \alpha} = \frac{\partial f(\alpha, \overline{\alpha})}{\partial \overline{\alpha}} = 0,$$

where $\frac{\partial}{\partial \alpha}$ ($\frac{\partial}{\partial \overline{\alpha}}$) is a Wirtinger derivative and acts as partial derivative with respect to $\alpha$ ($\overline{\alpha}$), while treating $\overline{\alpha}$ ($\alpha$) as constant [52]. The PS-$\mathbb{R}$ objective function is a bivariate

polynomial $f(\alpha, \beta)$ of total degree $2d_{\mathscr{M}}$ and its stationary points are the solutions of the bivariate polynomial system

$$\frac{\partial f(\alpha, \beta)}{\partial \alpha} = \frac{\partial f(\alpha, \beta)}{\partial \beta} = 0.$$

Of the many techniques to solve systems of polynomials [5,27,37,41,45,58], we will focus on resultant-based methods [26], which originate from algebraic geometry and are used to eliminate variables from systems of equations. The Sylvester resultant is perhaps the most well-known resultant formulation for systems of bivariate polynomials. Typically, the technique is used to eliminate one variable, say $y$, from a bivariate polynomial system $f(x, y) = g(x, y) = 0$, leaving us with a so-called resultant polynomial $r(x)$. If $(x^*, y^*)$ is a root of the system $f(x, y) = g(x, y) = 0$, then $x^*$ is also a root of the resultant $r(x)$.

At least two approaches can be applied to recover the corresponding $y$-coordinates. The first is to eliminate both $x$ and $y$ separately to obtain two resultants $r(x)$ and $s(y)$ [3,23,48]. Let $x^*$ and $y^*$ be roots of $r(x)$ and $s(y)$, respectively. For each such pair $(x^*, y^*)$, the task is then to determine whether or not that pair constitutes a root of the system $f(x, y) = g(x, y) = 0$. The second approach is often more efficient but also less robust. The roots of the resultant $r(x)$ are computed as the solution of an eigenvalue problem, and then these roots are lifted to obtain the corresponding $y$-coordinates by solving an additional smaller eigen-problem for each $x$-coordinate [6,13,39]. However, the lifting step requires estimating the root's multiplicity. Furthermore, for both approaches, some real roots may have a small imaginary part due to numerical rounding errors. Detecting which roots are real and removing candidate roots which are not roots of the system $f(x, y) = g(x, y) = 0$ are the two most delicate steps in numerical algorithms for solving polynomial systems based on resultants.

However, we shall see that it is possible to compute the real $(x^*, y^*)$ roots of the system $f(x, y) = g(x, y) = 0$ simultaneously as the solution of a single eigen-problem [53]. In fact, the same method can be applied to solve systems of polyanalytic polynomials $f(z, \overline{z}) = g(z, \overline{z}) = 0$. The key is that Sylvester's resultant can also be used to eliminate a complex variable from a polyanalytic polynomial system, as shown by the following lemma.

**Lemma 1** *Let $f(z, \overline{z})$ and $g(z, \overline{z})$ be two nonconstant polyanalytic polynomials. If $f(z, \overline{z})$ and $g(z, \overline{z})$ have a common zero $z^*$, then there are nonzero polynomials $s(\overline{z})$ and $t(\overline{z})$ satisfying*

$$f(z^*, \overline{z})s(\overline{z}) + g(z^*, \overline{z})t(\overline{z}) \equiv 0, \tag{3}$$

*where $d_s < d_g^{(\overline{z})}$ and $d_t < d_f^{(\overline{z})}$.*

*Proof* Write $f(z, \overline{z})$ and $g(z, \overline{z})$ as

$$f(z, \overline{z}) = \sum_{i=0}^{d_f^{(\overline{z})}} f_i(z)\overline{z}^i \quad \text{and} \quad g(z, \overline{z}) = \sum_{i=0}^{d_g^{(\overline{z})}} g_i(z)\overline{z}^i$$

where $f_i(x)$ ($g_i(x)$) is the polynomial in $x$ associated with the coefficient vector $\boldsymbol{f}_{i+1,:}$ ($\boldsymbol{g}_{i+1,:}$). Since $f(z^*, \overline{z}^*) = g(z^*, \overline{z}^*) = 0$, $f(z^*, \overline{z})$ and $g(z^*, \overline{z})$ will have a common factor $k(\overline{z}) := (\overline{z} - \overline{z}^*)^n$ for some positive integer $n$. We have that $\frac{f(z^*,\overline{z})}{k(\overline{z})} g(z^*, \overline{z}) = \frac{g(z^*,\overline{z})}{k(\overline{z})} f(z^*, \overline{z})$. Set $s(\overline{z}) := g(z^*, \overline{z})/k(\overline{z})$ and $t(\overline{z}) := -f(z^*, \overline{z})/k(\overline{z})$. □

In other words, if $z^*$ is a common root of $f(z, \overline{z}) = g(z, \overline{z}) = 0$, then once $f$ and $g$ are partially evaluated in $z^*$, there exist nonzero polynomials $s(\overline{z})$ and $t(\overline{z})$ such that the polynomial $f(z^*, \overline{z})s(\overline{z}) + g(z^*, \overline{z})t(\overline{z})$ is identically zero. Equation (3) can be rewritten as a set of linear equations in terms of the coefficient vectors $\boldsymbol{s}$ and $\boldsymbol{t}$ as

$$S^{f,g}(z^*) \cdot \begin{bmatrix} \boldsymbol{s} \\ \boldsymbol{t} \end{bmatrix} = \boldsymbol{0}, \tag{4}$$

where the Sylvester matrix $S^{f,g}(z)$ is a polynomial matrix defined as

$$S^{f,g}(z) := \left[\begin{array}{ccc|ccc} f_0(z) & & & g_0(z) & & \\ \vdots & \ddots & & \vdots & \ddots & \\ f_{d_f^{(\overline{z})}}(z) & & f_0(z) & g_{d_g^{(\overline{z})}}(z) & & g_0(z) \\ & \ddots & \vdots & & \ddots & \vdots \\ & & f_{d_f^{(\overline{z})}}(z) & & & g_{d_g^{(\overline{z})}}(z) \end{array}\right]. \tag{5}$$

$d_g^{(\overline{z})}$ columns; $d_f^{(\overline{z})}$ columns

The Sylvester matrix is singular when evaluated at a common root $z^*$. It follows that by Lemma 1, the resultant $\mathrm{res}^{f,g}(z) := \det(S^{f,g}(z))$ vanishes when $z$ is a root of the system $f(z, \overline{z}) = g(z, \overline{z}) = 0$. Hence, the roots of a polyanalytic polynomial system can be recovered as (a subset of) the roots of its resultant. Exploiting this fact, both coordinates of the real roots of a bivariate polynomial system $f(x, y) = g(x, y) = 0$ can now be obtained by encoding the variables $x$ and $y$ in the complex variables $z$ and $\overline{z}$ using the transformation

$$(x, y) = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \cdot (z, \overline{z}). \tag{6}$$

It is well known that a polynomial's roots can be very sensitive to small changes in the polynomial's coefficients [59]. For this reason, explicitly building the resultant $\mathrm{res}^{f,g}(z)$ as the determinant of the polynomial matrix $S^{f,g}(z)$ [47] and then computing its roots should be avoided. Instead, the roots can be recovered directly by noticing that (4) is a polynomial eigenvalue problem (PEP) with matrix pencil

$$S^{(0)} + S^{(1)}z + \cdots + S^{(d)}z^d, \tag{7}$$

where $S^{(i)} \in \mathbb{C}^{n\times n}$ for $i = 0, \ldots, d$, $n := d_f^{(\overline{z})} + d_g^{(\overline{z})}$ and $d := \max(d_f^{(z)}, d_g^{(z)})$. The PEP (4) can be solved by linearizing it into a generalized eigenvalue problem (GEP) of the form $(A - z^*B) \cdot \boldsymbol{y} = \boldsymbol{0}$. The matrices $A$ and $B$ are chosen so that $A - z^*B$ has the same spectrum as (7). For example, one such linearization is the second companion form [29], defined by

$$A := \begin{bmatrix} -S^{(d-1)} & \mathbb{I}_n & & \\ \vdots & & \ddots & \\ -S^{(1)} & & & \mathbb{I}_n \\ -S^{(0)} & & & \end{bmatrix} \text{ and} \tag{8a}$$

$$B := \begin{bmatrix} S^{(d)} & & & \\ & \mathbb{I}_n & & \\ & & \ddots & \\ & & & \mathbb{I}_n \end{bmatrix}. \tag{8b}$$

To further improve the accuracy of the computed eigenvalues, the system $f(z, \overline{z}) = g(z, \overline{z}) = 0$ and its associated pencil $A - z^*B$ can be balanced and the eigenvalues iteratively refined and filtered by a complex Newton–Raphson algorithm. For details, we refer the interested reader to the companion paper [53].

In summary, the polyanalytic polynomial systems arising in (S)LS-$\mathbb{C}$ can be solved efficiently by building the pencil (8) associated with the polyanalytic polynomial system and then computing its eigenvalues. For bivariate polynomial systems such as those arising in (S)PS-$\mathbb{R}$, the bivariate system can first be transformed into a polyanalytic polynomial system using (6), after which its roots can be recovered as the real and imaginary parts of the eigenvalues of the associated pencil. Since we are interested in global minimizers of (S)LS-$\mathbb{C}$ or (S)PS-$\mathbb{R}$, we have only to evaluate the corresponding objective function for each of the eigenvalues to see which one corresponds to a global optimizer. Consequently, any spurious roots introduced by the resultant are of no concern, as they will simply evaluate to a poor objective function value.

We conclude this section by discussing the complexity of line and plane searches for CPD of an $N$-th order order tensor. To solve the LS-$\mathbb{R}$ subproblem, the roots of a univariate polynomial must be computed. This can be done by computing the eigenvalues of the polynomial's companion matrix. The degree of the polynomial to be minimized is $2N$, and so we need to compute the roots of its derivative (which has degree $2N - 1$). So, the computational complexity of solving LS-$\mathbb{R}$ is equal to the cost of computing the roots of a polynomial of degree $2N - 1$ as the eigenvalues of its companion matrix. This is $O((2N - 1)^3)$. Solving the GEP is the largest cost of solving the line and plane search subproblems, and so their cost can be described as $O((2d^2)^3)$, where $d$ is the total degree of the to be rooted bivariate polynomials. For example, from Table 1, we see that for PS-$\mathbb{R}$ $d = 2N$, and so the cost of PS-$\mathbb{R}$ is $O(8N^6)$, which is much more expensive than LS-$\mathbb{R}$, which is only $O(8N^3)$. On the other hand, $N$ is "small", and plane search computes better steps than line search. But in the end, both LS-$\mathbb{R}$ and PS-$\mathbb{R}$ give about the same speedup in cputime (see Sect. 5).

## 5 Numerical experiments

The following numerical experiments investigate the benefit of combining tensor optimization algorithms with exact line and plane search in the context of blind separation and equalization of convolutive DS-CDMA mixtures received by an antenna array after multipath propagation. Similarly to [42,43], we assume that the signal of the $r$th user is subject to intersymbol interference (ISI) over $L$ consecutive symbols and that this signal arrives at the antenna array via $P$ specular paths.

Let $I$ be the CDMA spreading code length and let $\mathscr{H}^{(r)} \in \mathbb{C}^{I\times L\times P}$, $r = 1, \ldots, R$, contain the convolution of the $r$th user's spreading code with the corresponding channel for each of the $P$ specular paths. Furthermore, let $S^{(r)} \in \mathbb{C}^{J\times L}$, $r = 1, \ldots, R$, be a Toeplitz matrix containing the $J$ transmitted QPSK symbols of the $r$th user (along with $L-1$ interfering symbols) and let $A^{(r)} \in \mathbb{C}^{K\times P}$, $r = 1, \ldots, R$, contain the response of the $K$ antennas according to the angles of arrival of the $P$ paths. The convolutive model [42] for the observed tensor $\mathscr{Y} \in \mathbb{C}^{I\times J\times K}$ is then a rank-$(L, L, P)$ BTD of the form

$$\mathscr{Y} = \sum_{r=1}^{R}\sum_{p=1}^{P} \left(H_{::p}^{(r)} \cdot S^{(r)\mathrm{T}}\right) \circ \boldsymbol{a}_p^{(r)}, \tag{9}$$

which can also be interpreted as a structured CPD or rank-$(L, L, 1)$ BTD.

In our experiments, we use pseudorandom spreading codes of length $I = 32$, frames of $J = 64$ QPSK symbols, $K = 8$ antennas, $L = 4$ interfering symbols, $P = 2$ major paths per user and $R = 16$ users. The signal-to-noise ratio (SNR) at the receiver is defined as $10\log_{10}(\|\mathscr{Y}\|_{\mathrm{F}}^2/\|\mathscr{N}\|_{\mathrm{F}}^2)$, where $\mathscr{Y}$ is the complex-valued noise-free tensor of observations and $\mathscr{N}$ is complex-valued zero-mean white Gaussian noise. We measure the bit error rate (BER) and the speedup in terms of iterations and CPU time compared to a baseline algorithm with two Monte Carlo experiments consisting of 500 runs for each SNR level. In the first experiment, we set the condition number of the antenna matrix $A := \begin{bmatrix} A^{(1)} & \cdots & A^{(R)} \end{bmatrix}$ equal to $\kappa(A) = 1$ and in the second, we choose $\kappa(A) = 10$.

For each run, we generate the observations $\mathscr{Y}$ according to (9) and generate a noise tensor $\mathscr{N}$. We choose the structured alternating least squares (ALS) algorithm implemented in Tensorlab [54] as the baseline algorithm to decompose the noisy tensor $\mathscr{Y}+\mathscr{N}$ as a structured CPD. The algorithm alternatingly computes least squares updates for the channel tensors $\mathscr{H}^{(r)}$, the users' symbols in $S^{(r)}$ and the antenna matrix $A$ and is in this respect similar to the ALS algorithm described in [20,42]. We also decompose the tensor using ALS combined with the (S)LS-$\mathbb{R}$, (S)LS-$\mathbb{C}$ and (S)PS-$\mathbb{R}$ search algorithms, which are applied every four iterations. In our experience, applying the search algorithms once every few iterations give the parent optimization algorithm some breathing room for the iterates to stray away from valley floors in the objective function. This allows for similar or even better gains in the number of iterations, compared to executing the search algorithm every iteration, for a fraction of the cost.

Let $z^{(k)}$ be the $k$th iterate of the ALS algorithm. For line searches, we choose to search along the line $\Delta z_1 := z_k - z_{k-1}$ and for plane searches we add a second

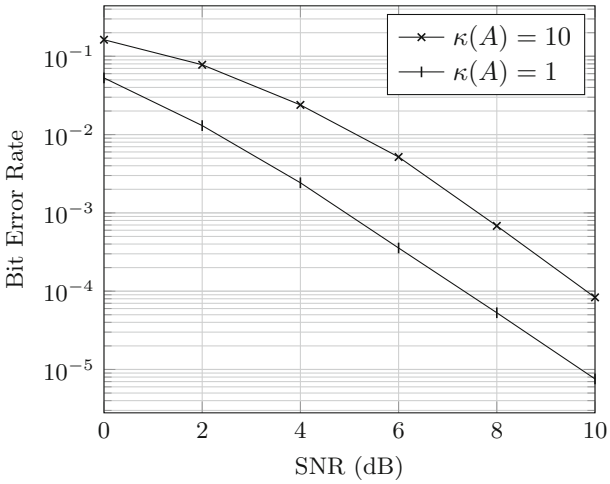



**Fig. 5** Mean BER as a function of the SNR of the observed tensor and the condition number of the antenna matrix $A$

direction $\Delta z_2 := z_{k-1} - z_{k-2}$. We refer to these two search directions as linear search directions. The bivariate systems arising in (scaled) complex line search and real plane search are solved using the generalized eigenvalue method described in the previous section. All aforementioned line and plane search algorithms were implemented as part of Tensorlab. We measure the iteration and CPU time speedup as the mean number of iterations and CPU time required by ALS, relative to that required by ALS with search algorithm, respectively. The algorithms are stopped if the maximum of 1000 iterations has been reached or if the norm of the step relative to the norm of the current iterate is smaller than $10^{-3}$. All experiments were performed in MATLAB 8.1 (R2013a) on two hexacore Intel Xeon E5645 CPUs with 48 GB RAM.

Figure 5 shows the mean BER of the two Monte Carlo experiments as a function of the SNR of the observed tensor. Increasing the condition number of the antenna matrix has an adverse effect on the BER, and combining ALS with a search algorithm does not deteriorate nor improve the BER.

In Fig. 6a and b, we see the mean iteration and CPU time speedup compared to ALS, respectively, for the case $\kappa(A) = 1$. Both line and plane searches offer a good speedup of the number of iterations and CPU time, relative to the baseline algorithm ALS. Compared to LS-$\mathbb{R}$, LS-$\mathbb{C}$ and PS-$\mathbb{R}$, the scaled searches SLS-$\mathbb{R}$, SLS-$\mathbb{C}$ and SPS-$\mathbb{R}$ offer only a small increase in iteration speedup. For real and complex line search, their scaled counterparts also translate into slightly faster CPU times. However, the improvement is almost negligible, and in the case of real scaled plane search even negative due to the additional overhead of solving a higher degree polynomial system than that of a real plane search. There also does not seem to be much difference in performance between real and complex line searches. These observations suggest that the linear search directions are already well scaled in magnitude and phase and that ALS mainly benefits from increasing the dimension of the search space.

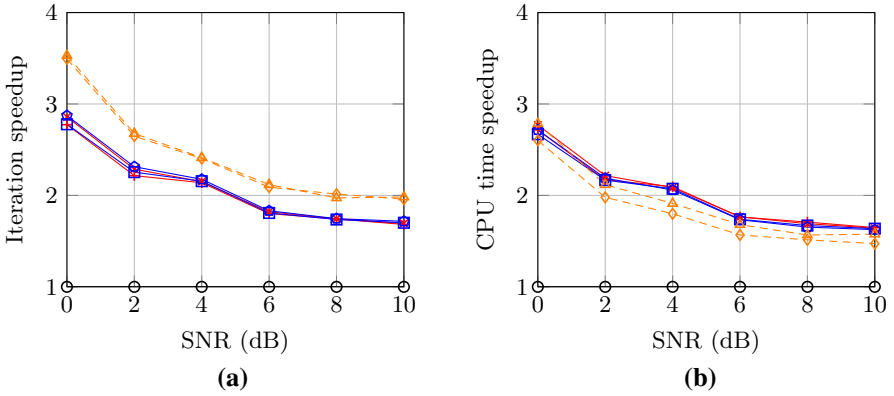



**Fig. 6** Mean iteration and CPU time speedup compared to ALS as a function of the SNR of the observed tensor. The condition number of the antenna matrix is $\kappa(A) = 1$. Both line and plane searches use linear search directions. ALS(—○—), ALS+LS-$\mathbb{R}$(—+—), ALS+SLS-$\mathbb{R}$(—×—), ALS+LS-$\mathbb{C}$(—□—), ALS+SLS-$\mathbb{C}$(—⬡—), ALS+PS-$\mathbb{R}$(- -△- -), ALS+SPS-$\mathbb{R}$(- -◇- -)

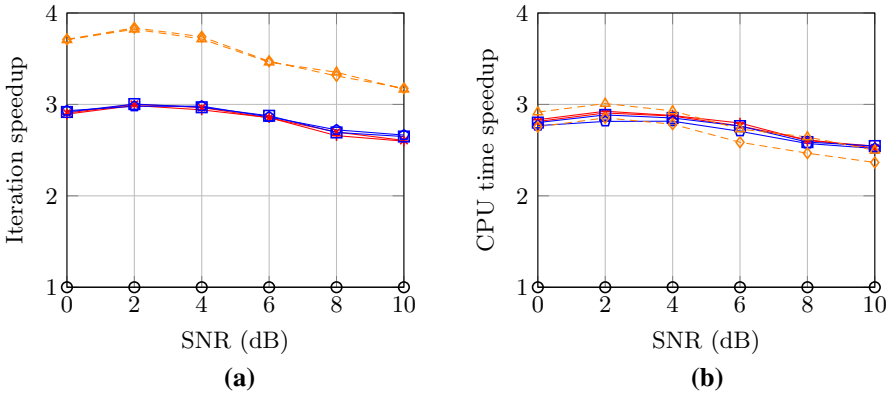



**Fig. 7** Mean iteration and CPU time speedup compared to ALS as a function of the SNR of the observed tensor. The condition number of the antenna matrix is $\kappa(A) = 10$. Both line and plane searches use linear search directions. ALS(—○—), ALS+LS-$\mathbb{R}$(—+—), ALS+SLS-$\mathbb{R}$(—×—), ALS+LS-$\mathbb{C}$(—□—), ALS+SLS-$\mathbb{C}$(—⬡—), ALS+PS-$\mathbb{R}$(- -△- -), ALS+SPS-$\mathbb{R}$(- -◇- -)

Figure 7a and b show the speedups for the case $\kappa(A) = 10$, which is expected to be more difficult than the previous situation as the terms in the decomposition tend to be more similar, and hence more difficult to separate, as the condition number of the antenna matrix increases. The search algorithms are now able to reduce the CPU time by about 60%. The speedup now seems to depend less on the SNR of the observed tensor.

For a final set of experiments, we define the search direction [9]

$$\Delta z_1 := 2z_k - 3z_{k-1} + z_{k-2} \tag{10}$$

which is based on a quadratic extrapolation of the iterates $z_k$, assuming they are spaced equally apart. Figures 8 and 9 show the results of repeating the previous two

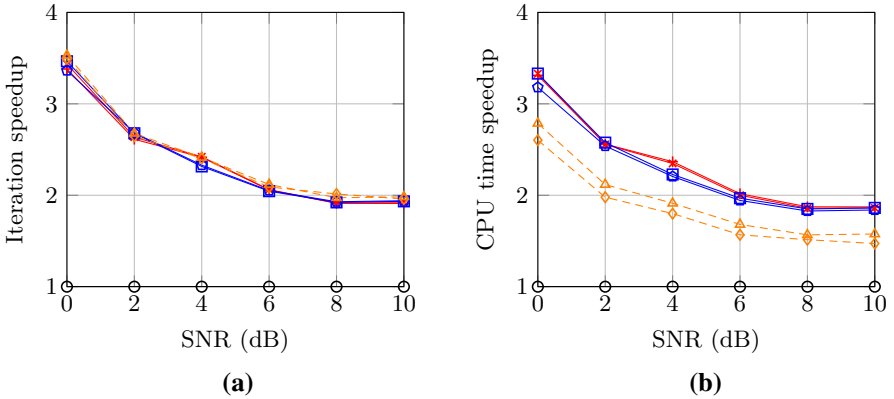



**Fig. 8** Mean iteration and CPU time speedup compared to ALS as a function of the SNR of the observed tensor. The condition number of the antenna matrix is $\kappa(A) = 1$. Line searches use the quadratically extrapolated direction (10), while plane searches use linear search directions. ALS(—o—), ALS+LS-$\mathbb{R}$(—+—), ALS+SLS-$\mathbb{R}$(—×—), ALS+LS-$\mathbb{C}$(—□—), ALS+SLS-$\mathbb{C}$(—⬠—), ALS+PS-$\mathbb{R}$(- △ -), ALS+SPS-$\mathbb{R}$(- ◇ -)

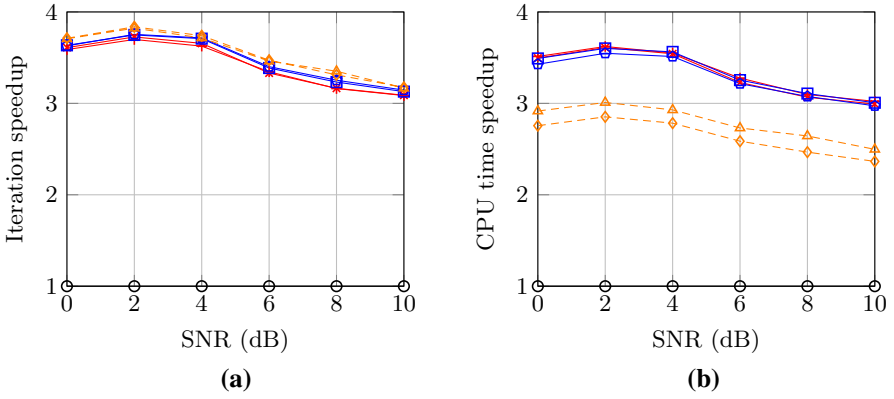



**Fig. 9** Mean iteration and CPU time speedup compared to ALS as a function of the SNR of the observed tensor. The condition number of the antenna matrix is $\kappa(A) = 10$. Line searches use the quadratically extrapolated direction (10), while plane searches use linear search directions. ALS(—o—), ALS+LS-$\mathbb{R}$(—+—), ALS+SLS-$\mathbb{R}$(—×—), ALS+LS-$\mathbb{C}$(—□—), ALS+SLS-$\mathbb{C}$(—⬠—), ALS+PS-$\mathbb{R}$(- △ -), ALS+SPS-$\mathbb{R}$(- ◇ -)

experiments using the quadratically extrapolated search direction for the line searches. It is immediately clear that line searches are able to obtain speedups close to those of the plane searches by using the improved search direction. Evidently, the quality of the search direction plays an important role in the improvement a line or plane search method can offer.

## 6 Conclusion

We introduced tensor optimization as a generic framework for computing tensor decompositions, least squares support tensor machines and jointly analysing multiple

datasets. We showed that the global minimizers of their line and plane search subproblems can be obtained by computing the roots of univariate polynomials or bivariate polynomial systems. In nonlinear least squares problems, a linear approximation of the residual function is typically used to solve the underlying two-dimensional subspace minimization problem. We showed that one may instead use a higher-order Taylor series expansion of the residual function and efficiently obtain its global minimizers in the same way as for tensor optimization problems. Furthermore, we introduced scaled line and plane searches which compute an additional scaling parameter for little additional cost. Finally, we showed how to obtain the real solutions of a bivariate polynomial system and the complex solutions of a polyanalytic polynomial system by solving a single generalized eigenvalue problem. In the numerical experiments, we applied a structured ALS algorithm in combination with (S)LS-$\mathbb{R}$, (S)LS-$\mathbb{C}$ and (S)PS-$\mathbb{R}$ to blind multi-user detection-estimation in DS-CDMA communication, in which the exact line and plane searches were able to speed up the computational time by up to a factor of 3. All algorithms were implemented in the MATLAB toolbox Tensorlab [54] and are available online.

**Acknowledgments** L. Sorber was supported by a doctoral fellowship of the Flanders agency for Innovation by Science and Technology (IWT). I. Domanov and L. De Lathauwer were supported by the Research Council KU Leuven: GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC), PDM postdoc Grant; F.W.O.: project G.0830.14N, G.0881.14N; the Belgian Federal Science Policy Office: IUAP P7 (DYSCO II, Dynamical systems, control and optimization, 2012-2017); EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no. 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. M. Van Barel was supported by the Research Council KU Leuven: OT/10/038, CoE PF/10/002 (OPTEC); F.W.O.: project G.0828.14N; by the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office.

## References

1. Abatzoglou, T.J.: Tensor-based techniques for the blind separation of DS-CDMA signals. Signal Process. **87**(2), 322–336 (2007)
2. Acar, E., Rasmussen, M.A., Savorani, F., Næs, T., Bro, R.: Understanding data fusion within the framework of coupled matrix and tensor factorizations. Chemom. Intell. Lab. Syst. **129**, 53–63 (2013)
3. Berberich, E., Emeliyanenko, P., Sagraloff, M.: An elimination method for solving bivariate polynomial systems: Eliminating the usual drawbacks. In: Müller-Hannemann, M., Werneck, R. (eds.) Proceedings of the Thirteenth Workshop on Algorithm Engineering and Experiments (ALENEX), pp. 35–47. SIAM, San Francisco, CA, USA (2011)
4. Bro, R.: Multi-way analysis in the food industry: models, algorithms, and applications. Ph.D. Thesis, University of Amsterdam (1998)
5. Buchberger, B.: Gröbner bases and systems theory. Multidim. Syst. Signal Process. **12**, 223–251 (2001)
6. Busé, L., Khalil, H., Mourrain, B.: Resultant-based methods for plane curves intersection problems. In: Proceedings of the 8th International Conference on Computer Algebra in Scientific Computing. Lecture Notes in Computer Science, vol. 3718, pp. 75–92. Springer-Verlag, Heidelberg (2005)
7. Byrd, R.H., Schnabel, R.B., Schultz, G.A.: Approximate solution of the trust region problem by minimization over two-dimensional subspaces. Math. Program. **40**(1), 247–263 (1988)
8. Carroll, J.D., Chang, J.J.: Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. Psychometrika **35**(3), 283–319 (1970)
9. Chen, Y., Han, D., Qi, L.: New ALS methods with extrapolating search directions and optimal step size for complex-valued tensor decompositions. IEEE Trans. Signal Process. **59**(12), 5888–5898 (2011)

10. Cichocki, A., Mandic, D., Caiafa, C., Phan, A.H., Zhou, G., Zhao, Q., De Lathauwer, L.: Tensor decompositions for signal processing applications. From two-way to multiway component analysis. IEEE Signal Process. Mag. **32**(2), 145–163 (2015)
11. Comon, P.: Independent component analysis, a new concept? Signal Process. **36**(3), 287–314 (1994)
12. Comon, P., Jutten, C.: Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press, New York (2010)
13. Corless, R.M., Gianni, P.M., Trager, B.M.: A reordered Schur factorization method for zero-dimensional polynomial systems with multiple roots. In: Proceedings of the International Symposium on Symbolic and Algebraic Computation, pp. 133–140. ACM Press, New York (1997)
14. De Lathauwer, L.: Decompositions of a higher-order tensor in block terms—Part I: lemmas for partitioned matrices. SIAM J. Matrix Anal. Appl. **30**(3), 1022–1032 (2008)
15. De Lathauwer, L.: Decompositions of a higher-order tensor in block terms—Part II: definitions and uniqueness. SIAM J. Matrix Anal. Appl. **30**(3), 1033–1066 (2008)
16. De Lathauwer, L.: Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms. SIAM J. Matrix Anal. Appl. **32**(4), 1451–1474 (2011)
17. De Lathauwer, L.: A short introduction to tensor-based methods for factor analysis and blind source separation. In: Proceeding of the 7th International Symposium on Image and Signal Processing and Analysis (ISPA 2011), pp. 558–563 (2011)
18. De Lathauwer, L.: Block component analysis, a new concept for source separation and factor analysis. In: Latent Variable Analysis and Signal Separation. Lecture Notes in Computer Science, vol. 7191, pp. 1–8. Proceedings of 10th International Conference, LVA/ICA 2012, Tel Aviv, Israel, March 12–15, 2012. (2012)
19. De Lathauwer, L., de Baynast, A.: Blind deconvolution of DS-CDMA signals by means of decomposition in rank-$(1, L, L)$ terms. IEEE Trans. Signal Process. **56**(4), 1562–1571 (2008)
20. De Lathauwer, L., Nion, D.: Decompositions of a higher-order tensor in block terms—Part III: alternating least squares algorithms. SIAM J. Matrix Anal. Appl. **30**(3), 1067–1083 (2008)
21. De Lathauwer, L., Vandewalle, J.: Dimensionality reduction in higher-order signal processing and rank-$(R_1, R_2, \ldots, R_N)$ reduction in multilinear algebra. Linear Algebra Appl. **391**, 31–55 (2004)
22. Ding, Z., Johnson, C.R., Kennedy, R.A.: On the (non)existence of undesirable equilibria of godard blind equalizers. IEEE Trans. Signal Process. **40**(10), 2425–2432 (1992)
23. Diochnos, D.I., Emiris, I.Z., Tsigaridas, E.P.: On the asymptotic and practical complexity of solving bivariate systems over the reals. J. Symb. Comput. **44**(7), 818–835 (2009)
24. Domanov, I., De Lathauwer, L.: On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part I: basic results and uniqueness of one factor matrix. SIAM J. Matrix Anal. Appl. **34**, 855–875 (2013)
25. Domanov, I., De Lathauwer, L.: On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part II: overall uniqueness. SIAM J. Matrix Anal. Appl. **34**, 876–903 (2013)
26. Emiris, I.Z., Mourrain, B.: Matrices in elimination theory. J. Symb. Comput. **28**(1–2), 3–43 (1999)
27. Faugère, J.C.: A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). In: Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation, ISSAC '02, pp. 75–83. ACM, New York (2002)
28. Godard, D.N.: Self-recovering equalization and carrier tracking in two-dimensional data communication systems. IEEE Trans. Commun. **28**(11), 1867–1875 (1980)
29. Gohberg, I.C., Lancaster, P., Rodman, L.: Matrix Polynomials. Academic Press, New York (1982)
30. Haardt, M., Roemer, F., Del Galdo, G.: Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems. IEEE Trans. Signal Process. **56**(7), 3198–3213 (2008)
31. Harshman, R.A.: Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis. UCLA Working Papers in Phonetics, vol. 16, pp. 1–84 (1970)
32. Karami, A., Yazdi, M., Zolghadre Asli, A.: Noise reduction of hyperspectral images using kernel non-negative Tucker decomposition. IEEE J. Sel. Top. Signal Process. **5**(3), 487–493 (2011)
33. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
34. Kruskal, J.B.: Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra Appl. **18**(2), 95–138 (1977)
35. Kruskal, J.B.: Rank, Decomposition, and Uniqueness for 3-way and $N$-way Arrays, pp. 7–18. Elsevier Science Publishers B.V., Amsterdam (1989)

36. Kruskal, J.B., Harshman, R.A., Lundy, M.E.: How 3-MFA data can cause degenerate PARAFAC solutions, among other relationships. In: Coppi, R., Bolasco, S. (eds.) Multiway Data Analysis, pp. 115–121. North-Holland Publishing Co., Amsterdam (1989)
37. Li, T.Y.: Numerical solution of multivariate polynomial systems by homotopy continuation methods. Acta Numer. **6**, 399–436 (1997)
38. Lu, H., Plataniotis, K.N., Venetsanopoulos, A.N.: A survey of multilinear subspace learning for tensor data. Pattern Recognit. **44**(7), 1540–1551 (2011)
39. Manocha, D., Demmel, J.W.: Algorithms for intersecting parametric and algebraic curves II: multiple intersections. Graph. Models Image Process. **57**(2), 81–100 (1995)
40. Mitchell, B.C., Burdick, D.S.: Slowly converging PARAFAC sequences: swamps and two-factor degeneracies. J. Chem. **8**(2), 155–168 (1994)
41. Moore, R.E.: Methods and Applications of Interval Analysis. SIAM, Philadelphia (1987)
42. Nion, D., De Lathauwer, L.: Block component model based blind DS-CDMA receivers. IEEE Trans. Signal Process. **56**(11), 5567–5579 (2008)
43. Nion, D., De Lathauwer, L.: An enhanced line search scheme for complex-valued tensor decompositions. Appl. DS-CDMA Signal Process. **88**(3), 749–755 (2008)
44. Nion, D., De Lathauwer, L.: Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar. IEEE Trans. Signal Process. **58**(11), 5693–5705 (2010)
45. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operations Research, 2nd edn. Springer, New York (2006)
46. Rajih, M., Comon, P., Harshman, R.A.: Enhanced line search: a novel method to accelerate PARAFAC. SIAM J. Matrix Anal. Appl. **30**(3), 1128–1147 (2008)
47. Šebek, M., Pejchová, S., Henrion, D., Kwakernaak, H.: Numerical methods for zeros and determinant of polynomial matrix. In: Proceedings of the IEEE Mediterranean Symposium on New Directions in Control and Automation, pp. 488–491. Chania, Crete, Greece (1996)
48. Seidel, R., Wolpert, N.: On the exact computation of the topology of real algebraic curves. In: Proceedings of the Twenty-first Annual Symposium on Computational Geometry, SCG 2005, pp. 107–115. ACM, New York (2005)
49. Sidiropoulos, N.D., Bro, R., Giannakis, G.B.: Parallel factor analysis in sensor array processing. IEEE Trans. Signal Process. **48**(8), 2377–2388 (2000)
50. Sidiropoulos, N.D., Dimić, G.Z.: Blind multiuser detection in W-CDMA systems with large delay spread. IEEE Signal Process. Lett. **8**(3), 87–89 (2001)
51. Sidiropoulos, N.D., Giannakis, G.B., Bro, R.: Blind PARAFAC receivers for DS-CDMA systems. IEEE Trans. Signal Process. **48**(3), 810–823 (2000)
52. Sorber, L., Van Barel, M., De Lathauwer, L.: Unconstrained optimization of real functions in complex variables. SIAM J. Optim. **22**(3), 879–898 (2012)
53. Sorber, L., Van Barel, M., De Lathauwer, L.: Numerical solution of bivariate and polyanalytic polynomial systems. SIAM J. Numer. Anal. **52**(4), 1551–1572 (2014)
54. Sorber, L., Van Barel, M., De Lathauwer, L.: Tensorlab v2.0. http://www.tensorlab.net/. Accessed January 2014
55. Tao, D., Li, X., Wu, X., Hu, W., Maybank, J.: Supervised tensor learning. Knowl. Inform. Syst. **13**(1), 1–42 (2007)
56. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika **31**(3), 279–311 (1966)
57. Vasilescu, M.A.O., Terzopoulos, D.: Multilinear analysis of image ensembles: tensorfaces. In: Proceedings 7th European Conference on Computer Vision. Lecture Notes in Computer Science, vol. 2350, pp. 447–460. Springer-Verlag, Heidelberg (2002)
58. Verschelde, J., Verlinden, J., Cools, R.: Homotopies exploiting newton polytopes for solving sparse polynomial systems. SIAM J. Numer. Anal. **31**, 915–930 (1994)
59. Wilkinson, J.H.: The evaluation of the zeros of ill-conditioned polynomials. Part I. Numer. Math. **1**(1), 150–166 (1959)
60. Zarzoso, V., Comon, P.: Optimal step-size constant modulus algorithm. IEEE Trans. Commun. **56**(1), 10–13 (2008)
61. Zhang, X., Ling, C., Qi, L.: The best rank-1 approximation of a symmetric tensor and related spherical optimization problems. SIAM J. Matrix Anal. Appl. **33**(3), 806–821 (2012)