ON THE BEST RANK-1 APPROXIMATION OF HIGHER-ORDER SUPERSYMMETRIC TENSORS[*]

ELEFTHERIOS KOFIDIS[†] AND PHILLIP A. REGALIA[‡]

Abstract. Recently the problem of determining the best, in the least-squares sense, rank-1 approximation to a higher-order tensor was studied and an iterative method that extends the well-known power method for matrices was proposed for its solution. This higher-order power method is also proposed for the special but important class of supersymmetric tensors, with no change. A simplified version, adapted to the special structure of the supersymmetric problem, is deemed unreliable, as its convergence is not guaranteed. The aim of this paper is to show that a symmetric version of the above method converges under assumptions of convexity (or concavity) for the functional induced by the tensor in question, assumptions that are very often satisfied in practical applications. The use of this version entails significant savings in computational complexity as compared to the unconstrained higher-order power method. Furthermore, a novel method for initializing the iterative process is developed which has been observed to yield an estimate that lies closer to the global optimum than the initialization suggested before. Moreover, its proximity to the global optimum is a priori quantifiable. In the course of the analysis, some important properties that the supersymmetry of a tensor implies for its square matrix unfolding are also studied.

Key words. supersymmetric tensors, rank-1 approximation, higher-order power method, higher-order singular value decomposition

AMS subject classifications. 15A18, 15A57, 15A69

PII. S0895479801387413

1. Introduction. A *tensor of order N* is an *N*-way array, i.e., its entries are accessed via N indices.[1] For example, a scalar is a tensor of order 0, a vector is a tensor of order 1, and a matrix is a second-order tensor. Tensors find applications in such diverse fields as physics, signal processing, data analysis, chemometrics, and psychology [4].

The notion of *rank* can also be defined for tensors of order higher than 2. The way this is done is via an extension of the well-known expansion of a matrix in a sum of rank-1 terms. Thus, the rank, R, of an Nth-order tensor $\mathcal{T}$ is the *minimum* number of rank-1 tensors that sum up to $\mathcal{T}$. A rank-1 tensor of order N is given by the generalized outer product of N vectors, $\boldsymbol{u}^{(i)}$, $i = 1, 2, \ldots, N$, i.e., its $(i_1, i_2, \ldots, i_N)$ entry is $\mathcal{T}_{i_1, i_2, \ldots, i_N} = u_{i_1}^{(1)} u_{i_2}^{(2)} \cdots u_{i_N}^{(N)}$. Despite the similarity in their definitions, the ranks of lower- ($N \leq 2$) and higher-order tensors exhibit important differences. For example, the rank of a higher-order tensor is not necessarily upper bounded by the tensor dimensions [9]. Furthermore, there is not a unique way of extending to higher orders the singular value decomposition (SVD) and its connection with least-squares low rank approximation. A multilinear generalization of the matrix SVD, called *higher-order singular value decomposition* (HOSVD), was recently proposed and studied [9] and may be understood as an extension of the so-called Tucker3 model for 3-way

[†]Department of Informatics and Telecommunications, Division of Communications and Signal Processing, University of Athens, Panepistimioupoli, 157 84 Athens, Greece (kofidis@di.uoa.gr).
[‡]Département Communications, Image et Traitement de l' Information, Institut National des Télécommunications, 9 rue Charles Fourier, 91011 Évry cedex, France (Phillip.Regalia@int-evry.fr).
[1]Although the tensor admits a more rigorous definition in terms of a tensor product induced by a multilinear mapping [12], the above definition suffices for the purposes of this paper.

arrays (see, e.g., [18]). Despite the many similarities of this decomposition with the second-order SVD (e.g., orthogonality between the vectors u at different terms), its truncation does not provide the best (in the least-squares (LS) sense) low rank approximation to the tensor. Nevertheless, it has been used to initialize a higher-order equivalent of the power method, recently proposed [8, 10] for determining the best rank-1 approximation of Nth-order tensors.

Though an important problem per se, the LS reduced rank tensor approximation plays a central role in the context of blind source separation (BSS) based on higher-order statistics (HOS) [5]. The problem there is to separate and recover statistically independent random processes, $x_1, \ldots, x_K$, with the aid of observations of their linear mixture of the form $y = Hx + n$, where the $M \times K$ mixing matrix H, the source vector $x = [x_1, \ldots, x_K]^T$, and the disturbance vector n are assumed unknown (and real, for simplicity). The noise vector is also commonly assumed to be Gaussian and independent of x. A common method for recovering one of the sources is to project the observation vector y onto an $M \times 1$ vector u, chosen so that the normalized kurtosis of the source estimate $z = u^T y$, given by $\frac{\text{cum}_4(z)}{(\text{cum}_2(z))^2}$, is absolutely maximized. It has been shown [8] that this maximization problem is equivalent to that of best approximating the fourth-order cumulant tensor of y by another of rank 1. Hence the higher-order power method (HOPM) of [8, 10] can be employed. However, it should be noticed that the above tensor is *supersymmetric*, i.e., its entries remain unchanged under any permutation of their indices [4]. Such a rich symmetry would be expected to permit a simplified version of the HOPM to be applicable to this kind of tensor. Unfortunately such a symmetric HOPM (S-HOPM) is not convergent for general supersymmetric tensors as claimed in [10] and demonstrated via an example in this paper. Yet, in many cases of practical interest, namely, when the function $g(u) = \sum_{i_1, i_2, \ldots, i_N} \mathcal{T}_{i_1, i_2, \ldots, i_N} u_{i_1} u_{i_2} \cdots u_{i_N}$ is convex (or concave), this symmetric version of the HOPM can be shown to converge to a (local) maximum of the restriction of $|g|$ to the unit sphere. The gain from using this symmetric version comes mainly from the consequent reduction in computational complexity. Though more iterations are usually required for the S-HOPM to converge, the fact that they are N times cheaper than in the general HOPM more than compensates for that, resulting in significant computational complexity savings. The requirements for convexity (concavity) of the function g are always met in the BSS context when the source kurtoses are all of the same sign. For example in communications these kurtoses are negative and g turns out to be concave.[2]

The aim of this paper is to present the S-HOPM and prove that it converges for supersymmetric tensors whose induced polynomial forms enjoy the property of convexity or concavity. A novel scheme for initializing the S-HOPM for fourth-order tensors is also proposed, which has been observed to almost always outperform the HOSVD-based scheme of [10]. Moreover, it allows for an a priori quantification of its proximity to the globally optimal solution.

The rest of this paper is organized as follows. Section 2 introduces definitions of basic quantities and tensor operations. The problem is stated in section 3, where the equivalence of the rank-1 approximation problem with that of maximizing an associated functional and the characterization of the stationary points are recalled from [10, 16] for the symmetric case. The HOPM in both its general and symmetric

---

versions is given in section 4, and the S-HOPM is shown to converge under convexity/concavity assumptions. Some important properties that the supersymmetry of a tensor implies for its square matrix version are proved in section 5. These properties, to be used in the subsequent analysis, are important in their own right since they hold for *any* supersymmetric tensor, regardless of whether the associated function is convex/concave or not. Section 6 develops the new initialization method and derives bounds on its performance. The problem of computing a rank-R approximation $(R > 1)$ is briefly discussed in section 7. Section 8 concludes the paper.

1.1. Notation. Vectors will be denoted by bold lowercase letters (e.g., u) while bold uppercase letters (e.g., T) will denote tensors of second order (i.e., matrices). Higher-order tensors will be denoted by bold, calligraphic, uppercase letters (e.g., $\mathcal{T}$). The symbol I designates the identity matrix, where the dimensions will be understood from the context. The superscript ^{T} will be employed for transposition. The $(i, j, k, \ldots, l)$ element of a tensor $\mathcal{T}$ is denoted by $\mathcal{T}_{i,j,k,\ldots,l}$. All indices are assumed to start from one. The symbol $\otimes$ will be used to denote the (right) Kronecker product. If A is an $m \times n$ matrix, vec(A) will signify the $mn \times 1$ vector that is built from the columns of A stacked one below another. The inverse operator that builds a matrix from a vector is called unvec. Finally, for the sake of simplicity, only real tensors will be considered. The extension of the results to tensors with Hermitian symmetry [9] is straightforward.

2. Basic definitions. This section contains some definitions that will be useful in what follows. Since they have been presented in detail in earlier works [9, 10, 16], they are only briefly recalled here.

DEFINITION 1 (supersymmetric tensor). *A tensor is called supersymmetric if its entries are invariant under any permutation of their indices.*

The notions of scalar product and norm are easily extended to higher orders, as follows.

DEFINITION 2 (tensor scalar product). *The scalar product of two tensors $\mathcal{S}$ and $\mathcal{T}$, of the same order, N, and same dimensions, is given by*

$$\langle \mathcal{S}, \mathcal{T} \rangle = \sum_{i_1, i_2, \ldots, i_N} \mathcal{S}_{i_1, i_2, \ldots, i_N} \mathcal{T}_{i_1, i_2, \ldots, i_N}.$$

DEFINITION 3 (Frobenius norm). *The Frobenius norm of a tensor $\mathcal{T}$ of order N is defined as*

$$\|\mathcal{T}\| = \sqrt{\langle \mathcal{T}, \mathcal{T} \rangle} = \left(\sum_{i_1, i_2, \ldots, i_N} \mathcal{T}_{i_1, i_2, \ldots, i_N}^2 \right)^{1/2}.$$

DEFINITION 4 (matrix unfoldings). *The n-mode matrix unfolding, $T_{(n)}$, of an $M_1 \times M_2 \times \cdots \times M_N$ tensor $\mathcal{T}$ of order N with entries $\mathcal{T}_{i_1, i_2, \ldots, i_N}$ is defined as the $M_n \times M_1 M_2 \cdots M_{n-1} M_{n+1} \cdots M_N$ matrix whose columns are the M_n-dimensional vectors obtained from $\mathcal{T}$ by varying the index i_n and keeping the other indices fixed.*[3]

It is readily verified that, for a supersymmetric tensor, $\mathcal{T}$, all n-mode matrix unfoldings are equal, that is, $T_{(1)} = T_{(2)} = \cdots = T_{(N)}$. A square matrix version for supersymmetric tensors will also be used, as follows.

---

[3]The order of appearance of the n-mode vectors in $T_{(n)}$ is irrelevant in our context.

DEFINITION 5 (square matrix unfolding). *The square matrix unfolding, $\boldsymbol{T}$, of an $M \times M \times \cdots \times M$ supersymmetric tensor $\boldsymbol{\mathcal{T}}$ of even order $N = 2L$ is given by*

$$(2.1) \qquad\qquad T_{m,n} = \mathcal{T}_{i_1,i_2,\ldots,i_L,j_1,j_2,\ldots,j_L},$$

where

$$(2.2) \quad m = M^{L-1}(i_1 - 1) + \cdots + M(i_{L-1} - 1) + i_L, \quad 1 \le i_1, \ldots, i_L \le M,$$

$$(2.3) \quad n = M^{L-1}(j_1 - 1) + \cdots + M(j_{L-1} - 1) + j_L, \quad 1 \le j_1, \ldots, j_L \le M.$$

The outer product can be generalized to higher-orders, as follows.

DEFINITION 6 (Tucker product). *The Tucker product of N matrices $\{\boldsymbol{U}^{(n)}\}_{n=1}^{N}$, each of dimension $M_n \times L$, yields an Nth-order tensor $\boldsymbol{\mathcal{T}}$ of dimensions $M_1 \times M_2 \times \cdots \times M_N$ as*

$$\mathcal{T}_{i_1,i_2,\ldots,i_N} = \sum_{l=1}^{L} U_{i_1,l}^{(1)} U_{i_2,l}^{(2)} \cdots U_{i_N,l}^{(N)}$$

and is denoted by

$$\boldsymbol{\mathcal{T}} = \boldsymbol{U}^{(1)} \star \boldsymbol{U}^{(2)} \star \cdots \star \boldsymbol{U}^{(N)}.$$

A weighted outer product will be defined as follows.[4]

DEFINITION 7 (weighted Tucker product). *The weighted Tucker product (or $\boldsymbol{\mathcal{S}}$-product), with core an $L_1 \times L_2 \times \cdots \times L_N$ tensor $\boldsymbol{\mathcal{S}}$, of N matrices $\{\boldsymbol{U}^{(n)}\}_{n=1}^{N}$ of dimensions $M_n \times L_n$ yields an Nth-order $M_1 \times M_2 \times \cdots \times M_N$ tensor as*

$$\mathcal{T}_{i_1,i_2,\ldots,i_N} = \sum_{l_1=1}^{L_1} \sum_{l_2=1}^{L_2} \cdots \sum_{l_N=1}^{L_N} \mathcal{S}_{l_1,l_2,\ldots,l_N} U_{i_1,l_1}^{(1)} U_{i_2,l_2}^{(2)} \cdots U_{i_N,l_N}^{(N)}$$

and is denoted by

$$\boldsymbol{\mathcal{T}} = \boldsymbol{U}^{(1)} \overset{\mathcal{S}}{\star} \boldsymbol{U}^{(2)} \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \boldsymbol{U}^{(N)}.$$

It can easily be seen that the standard Tucker product is a weighted $\boldsymbol{\mathcal{I}}$-product, where $\boldsymbol{\mathcal{I}}$ is the identity tensor ($\mathcal{I}_{i_1,i_2,\ldots,i_N} = \delta(i_1, i_2, \ldots, i_N)$). The Tucker product with all $\boldsymbol{U}^{(n)}$ being vectors results in a rank-1 tensor, as follows.

DEFINITION 8 (tensor rank). *The rank, R, of an $M_1 \times M_2 \times \cdots \times M_N$ tensor $\boldsymbol{\mathcal{T}}$ is the minimal number of terms in a finite decomposition of $\boldsymbol{\mathcal{T}}$ of the form*

$$\boldsymbol{\mathcal{T}} = \sum_{r=1}^{R} \boldsymbol{u}_r^{(1)} \star \boldsymbol{u}_r^{(2)} \star \cdots \star \boldsymbol{u}_r^{(N)},$$

where $\boldsymbol{u}_r^{(i)}$ are M_i-dimensional column vectors.

A way of extending the SVD from matrices to higher-order tensors is given in the following (see [9]).

THEOREM 1 (HOSVD). *Any $M_1 \times M_2 \times \cdots \times M_N$ tensor $\boldsymbol{\mathcal{T}}$ can be expressed as*

$$\boldsymbol{\mathcal{T}} = \boldsymbol{U}^{(1)} \overset{\mathcal{S}}{\star} \boldsymbol{U}^{(2)} \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \boldsymbol{U}^{(N)},$$

---

[4]This product is denoted in [9, 10] as $\boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \cdots \times_N \boldsymbol{U}^{(N)}$, whereas it takes the form $\boldsymbol{\mathcal{S}} \bullet \boldsymbol{U}^{(1)} \bullet \cdots \bullet \boldsymbol{U}^{(N)}$ in [4].

where
- $\boldsymbol{U}^{(n)}$, $n = 1, 2, \ldots, N$, *are orthogonal* $M_n \times M_n$ *matrices, and*
- *the core tensor* $\boldsymbol{S}$ *is of the same size as* $\boldsymbol{T}$, *and its subtensors* $\boldsymbol{S}_{i_n = \alpha}$, *obtained by fixing the kth index to* α, *have the properties of*
  - *all-orthogonality: two subtensors* $\boldsymbol{S}_{i_n = \alpha}$ *and* $\boldsymbol{S}_{i_n = \beta}$ *are orthogonal for any possible values of* n *and* $\alpha \neq \beta$, *in the sense that*

$$\langle \boldsymbol{S}_{i_n = \alpha}, \boldsymbol{S}_{i_n = \beta} \rangle = 0;$$

  - *ordering: for all* n,

$$\|\boldsymbol{S}_{i_n = 1}\| \geq \|\boldsymbol{S}_{i_n = 2}\| \geq \cdots \geq \|\boldsymbol{S}_{i_n = M_n}\|.$$

The matrix $\boldsymbol{U}^{(n)}$ is computed as the matrix of the left singular vectors of the n-mode unfolding of $\boldsymbol{T}$, $\boldsymbol{T}_{(n)}$ [9]. The core tensor is then determined as $\boldsymbol{S} = (\boldsymbol{U}^{(1)})^T \overset{\mathcal{T}}{\star} (\boldsymbol{U}^{(2)})^T \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} (\boldsymbol{U}^{(N)})^T$. In the supersymmetric case, all $\boldsymbol{U}^{(n)}$ are equal and, of course, the core tensor is supersymmetric as well. The above multilinear SVD then reduces to the so-called higher-order eigenvalue decomposition (HOEVD) [9]. We will sometimes use the notation $\boldsymbol{U}^{\overset{\mathcal{S}}{\star} N}$ to denote a symmetric $\boldsymbol{S}$-product.

3. Problem statement. The best LS rank-1 tensor approximation problem is stated below along with its close connection to the maximization of the associated polynomial form on the unit sphere. This property is the higher-order equivalent of an analogous property holding for matrices [11] and is going to play a central role in the subsequent developments. Proofs can be found (for the more general, nonsymmetric case) in [7, 10, 16].

THEOREM 2 (*tensor rank-1 approximation*). *Given an* N*th-order supersymmetric* $M \times M \times \cdots \times M$ *tensor* $\boldsymbol{T}$, *consider the problem of determining a scalar* λ *and a vector* $\boldsymbol{u} \in \mathbb{R}^M$ *such that the rank-1 tensor* $\hat{\boldsymbol{T}} = \lambda \boldsymbol{u}^{\star N}$ *minimizes the function*

$$(3.1) \qquad\qquad f(\hat{\boldsymbol{T}}) = \|\boldsymbol{T} - \hat{\boldsymbol{T}}\|^2$$

subject to $\boldsymbol{u}$ *having unit norm. Then the unit-norm vector* $\boldsymbol{u}$ *corresponds to a (local) minimum of* (3.1) *if and only if it yields a (local) maximum of* $|g(\boldsymbol{u})|$, *with*

$$(3.2) \qquad\qquad g(\boldsymbol{u}) = \sum_{i_1, i_2, \ldots, i_N} \mathcal{T}_{i_1, i_2, \ldots, i_N} u_{i_1} u_{i_2} \cdots u_{i_N} = \langle \boldsymbol{T}, \boldsymbol{u}^{\star N} \rangle.$$

The corresponding value of λ *is* $\lambda = g(\boldsymbol{u})$.

To make the connection with similar results in the second-order case clearer, let us define the functional

$$(3.3) \qquad\qquad h(\boldsymbol{u}) = \frac{\langle \boldsymbol{T}, \boldsymbol{u}^{\star N} \rangle^2}{\langle \boldsymbol{u}, \boldsymbol{u} \rangle^N},$$

whose maximization corresponds to maximizing $g(\boldsymbol{u})$ for $\|\boldsymbol{u}\| = 1$. The above can be viewed as an Nth-order Rayleigh quotient squared. Recall that the corresponding maximization problem for matrices is solved by the dominant eigenpair $(\lambda, \boldsymbol{u})$, where λ is the eigenvalue with the largest absolute value [28, 11]. The stationary points of the corresponding functional $\frac{\boldsymbol{u}^T \boldsymbol{T} \boldsymbol{u}}{\boldsymbol{u}^T \boldsymbol{u}}$ are the solutions to

$$(3.4) \qquad\qquad \boldsymbol{T}\boldsymbol{u} = \frac{\boldsymbol{u}^T \boldsymbol{T} \boldsymbol{u}}{\boldsymbol{u}^T \boldsymbol{u}} \boldsymbol{u}.$$

The analogous result for the Nth-order case is as follows [7, 10, 16].

THEOREM 3 (characterization of stationary points). *The unit-norm vector $\boldsymbol{u}$ is a stationary point of the functional h of (3.3) if and only if*

$$(3.5) \qquad \sum_{i_2,\ldots,i_N} \mathcal{T}_{i_1,i_2,\ldots,i_N} u_{i_2} u_{i_3} \cdots u_{i_N} = \lambda u_{i_1} \text{ for all } i_1$$

or, equivalently,

$$(3.6) \qquad \boldsymbol{I} \overset{\mathcal{T}}{\star} \underbrace{(\boldsymbol{u}^T) \overset{\mathcal{T}}{\star} (\boldsymbol{u}^T) \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} (\boldsymbol{u}^T)}_{N-1 \ times} = \lambda \boldsymbol{u},$$

with $\lambda = \langle \boldsymbol{\mathcal{T}}, \boldsymbol{u}^{\star N} \rangle$.

4. The HOPM.

4.1. General case. An iterative use of (3.4) leads to the well-known power method for determining the dominant eigenpair of a matrix $\boldsymbol{T}$ [28, 11]. A tensorial equivalent is suggested by (3.6) and will be analyzed in the next subsection. Let us first have a look at the HOPM as given in [8, 10] for a general, not necessarily supersymmetric, $M_1 \times M_2 \times \cdots \times M_N$ tensor $\boldsymbol{\mathcal{T}}$.

ALGORITHM 1. HIGHER-ORDER POWER METHOD (HOPM).

Initialization: $\boldsymbol{u}_0^{(n)} = $ a unit-norm $M_n -$ vector, $1 \le n \le N$

    Iteration: for $k = 1, 2, \ldots$

$$\tilde{\boldsymbol{u}}_k^{(1)} = \boldsymbol{I} \overset{\mathcal{T}}{\star} (\boldsymbol{u}_{k-1}^{(2)})^T \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} (\boldsymbol{u}_{k-1}^{(N)})^T,$$

$$\lambda_k^{(1)} = \|\tilde{\boldsymbol{u}}_k^{(1)}\|,$$

$$\boldsymbol{u}_k^{(1)} = \frac{\tilde{\boldsymbol{u}}_k^{(1)}}{\lambda_k^{(1)}},$$

$$\boldsymbol{u}_k^{(2)} = (\boldsymbol{u}_k^{(1)})^T \overset{\mathcal{T}}{\star} \boldsymbol{I} \overset{\mathcal{T}}{\star} (\boldsymbol{u}_{k-1}^{(3)})^T \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} (\boldsymbol{u}_{k-1}^{(N)})^T,$$

$$\lambda_k^{(2)} = \|\tilde{\boldsymbol{u}}_k^{(2)}\|,$$

$$\boldsymbol{u}_k^{(2)} = \frac{\tilde{\boldsymbol{u}}_k^{(2)}}{\lambda_k^{(2)}},$$

$$\vdots$$

$$\tilde{\boldsymbol{u}}_k^{(n)} = (\boldsymbol{u}_k^{(1)})^T \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} (\boldsymbol{u}_k^{(n-1)})^T \overset{\mathcal{T}}{\star} \boldsymbol{I} \overset{\mathcal{T}}{\star} (\boldsymbol{u}_{k-1}^{(n+1)})^T \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} (\boldsymbol{u}_{k-1}^{(N)})^T,$$

$$\lambda_k^{(n)} = \|\tilde{\boldsymbol{u}}_k^{(n)}\|,$$

$$\boldsymbol{u}_k^{(n)} = \frac{\tilde{\boldsymbol{u}}_k^{(n)}}{\lambda_k^{(n)}},$$

$$\vdots$$

$$\tilde{\boldsymbol{u}}_k^{(N)} = (\boldsymbol{u}_k^{(1)})^T \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} (\boldsymbol{u}_k^{(N-1)})^T \overset{\mathcal{T}}{\star} \boldsymbol{I},$$

$$\lambda_k^{(N)} = \|\tilde{\boldsymbol{u}}_k^{(N)}\|,$$

$$\boldsymbol{u}_k^{(N)} = \frac{\tilde{\boldsymbol{u}}_k^{(N)}}{\lambda_k^{(N)}}$$

end

Output: $\hat{\boldsymbol{T}} = \lambda \boldsymbol{u}^{(1)} \overset{\mathcal{T}}{\star} \boldsymbol{u}^{(2)} \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} \boldsymbol{u}^{(N)}.$

The $\boldsymbol{T}$-product $\tilde{\boldsymbol{u}}_k^{(n)} = (\boldsymbol{u}_k^{(1)})^T \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} (\boldsymbol{u}_k^{(n-1)})^T \overset{\mathcal{T}}{\star} \boldsymbol{I} \overset{\mathcal{T}}{\star} (\boldsymbol{u}_{k-1}^{(n+1)})^T \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} (\boldsymbol{u}_{k-1}^{(N)})^T$ can be implemented as[5]

$$\tilde{\boldsymbol{u}}_k^{(n)} = \boldsymbol{T}_{(1)}(\boldsymbol{u}_k^{(1)} \otimes \cdots \otimes \boldsymbol{u}_k^{(n-1)} \otimes \boldsymbol{u}_{k-1}^{(n+1)} \otimes \cdots \otimes \boldsymbol{u}_{k-1}^{(N)}).$$

N such products have to be computed per iteration.

The above algorithm can be shown to always converge to a (local) maximum of (3.3), with the corresponding value of h given by λ^2. The convergence proof [22] relies on the fact that $\boldsymbol{u}^{(1)} \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} \boldsymbol{u}^{(N)}$ is a multilinear function of $\boldsymbol{u}^{(n)}$'s, that is, it is linear with respect to each of them.

An initial estimate of $\boldsymbol{u}^{(n)}$, $1 \leq n \leq N$, which has been observed in [8, 10] to very often lie in the basin of attraction of a globally optimal solution, is given by setting $\boldsymbol{u}_0^{(n)}$ equal to the dominant left singular vector of the n-mode matrix unfolding, $\boldsymbol{T}_{(n)}$. This is the first column of the matrix $\boldsymbol{U}^{(n)}$ in the HOSVD of $\boldsymbol{T}$. This initialization method is inspired from what holds in the matrix case, where the best rank-1 approximant is provided by the dominant singular triple [11]. As shown in [10], however, this property does not hold anymore for higher-order arrays, and only some bounds can be derived on the approximation error. The similarities, though, of HOSVD with its second-order counterpart suggest its use to compute an initial estimate for HOPM.

4.2. Symmetric case. As pointed out in [8, 10], for a supersymmetric $\boldsymbol{T}$, the convergence of Algorithm 1 is to a supersymmetric estimator $\hat{\boldsymbol{T}}$, with all $\boldsymbol{u}^{(n)}$'s being equal to each other. However, the intermediate results are not necessarily symmetric. It is shown in [10] that, with $\boldsymbol{T}$ being a supersymmetric $2 \times 2 \times \cdots \times 2$ tensor, the stationary points of the HOPM (solutions to (3.6)) can be determined as the roots of an appropriate Nth-order polynomial. For larger supersymmetric tensors the above algorithm is also proposed, as a constrained version suggested by (3.6) is deemed unreliable since it is not guaranteed to monotonically increase $|g|$. The algorithm suggested by (3.6) is as follows.

ALGORITHM 2. SYMMETRIC HIGHER-ORDER POWER METHOD (S-HOPM).

Initialization: $\boldsymbol{u}_0 = $ *a unit-norm* $M - $ *vector*

Iteration: for $k = 1, 2, \ldots$

$$\tilde{\boldsymbol{u}}_k = \boldsymbol{I} \overset{\mathcal{T}}{\star} (\boldsymbol{u}_{k-1}^T)^{\overset{\mathcal{T}}{\star}(N-1)},$$

$$\boldsymbol{u}_k = \frac{\tilde{\boldsymbol{u}}_k}{\|\tilde{\boldsymbol{u}}_k\|}$$

end

Output: $\hat{\boldsymbol{T}} = g(\boldsymbol{u})\boldsymbol{u}^{\overset{\mathcal{T}}{\star}N}$

The expression $\tilde{\boldsymbol{u}}_k = \boldsymbol{I} \overset{\mathcal{T}}{\star} (\boldsymbol{u}_{k-1}^T)^{\overset{\mathcal{T}}{\star}(N-1)}$ can as before be rewritten as

$$\tilde{\boldsymbol{u}}_k = \boldsymbol{T}_{(1)} \underbrace{(\boldsymbol{u}_{k-1} \otimes \cdots \otimes \boldsymbol{u}_{k-1})}_{N-1 \text{ times}}.$$

---

[5]Recall that all matrices $\boldsymbol{T}_{(n)}$, $n = 1, \ldots, N$, are equal for a supersymmetric $\boldsymbol{T}$.
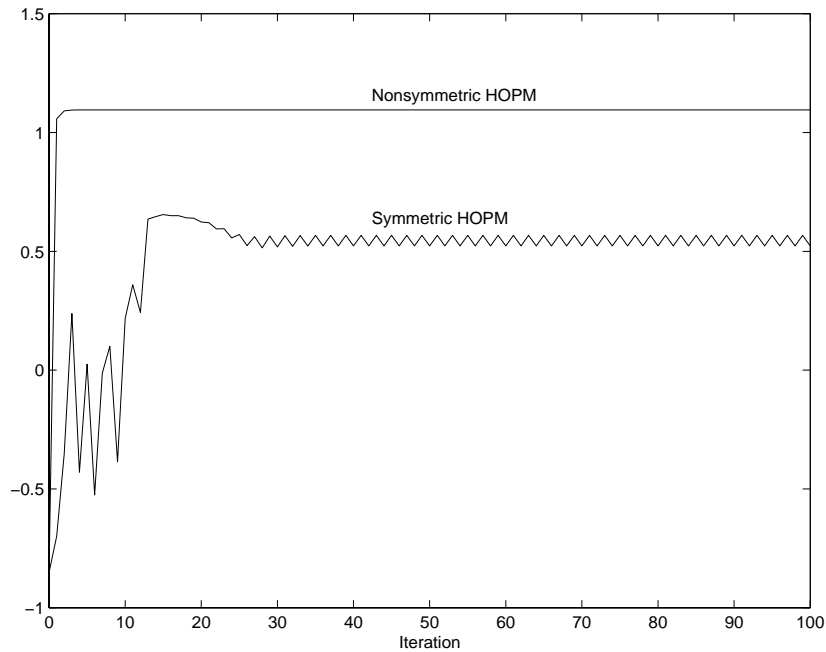
FIG. 4.1. *Results of the HOPM and S-HOPM for the supersymmetric tensor given in Example* 1.

For the special case of $N = 4$, which is the most common one in the BSS problem, the above can also take the following alternative form, in terms of the square matrix unfolding:

$$\tilde{\boldsymbol{u}}_k = \mathrm{unvec}(\boldsymbol{T}(\boldsymbol{u}_{k-1} \otimes \boldsymbol{u}_{k-1}))\boldsymbol{u}_{k-1}.$$

Note that only one such product needs to be computed per iteration, as compared to N for the general HOPM. Thus, if this constrained version is applicable, an N-fold reduction in computational complexity results. However, the form to be optimized is now nonlinear with respect to the sought for vector, $\boldsymbol{u}$, thus rendering the convergence proof for Algorithm 1 not applicable. In fact, as the following example demonstrates, the S-HOPM does not converge for all supersymmetric tensors $\boldsymbol{\mathcal{T}}$.

Example 1. Consider a supersymmetric $3 \times 3 \times 3 \times 3$ tensor with entries

$$\mathcal{T}_{1111} = 0.2883, \ \mathcal{T}_{1112} = -0.0031, \ \mathcal{T}_{1113} = 0.1973, \ \mathcal{T}_{1122} = -0.2485, \ \mathcal{T}_{1123} = -0.2939,$$

$$\mathcal{T}_{1133} = 0.3847, \ \mathcal{T}_{1222} = 0.2972, \ \mathcal{T}_{1223} = 0.1862, \ \mathcal{T}_{1233} = 0.0919, \ \mathcal{T}_{1333} = -0.3619,$$

$$\mathcal{T}_{2222} = 0.1241, \ \mathcal{T}_{2223} = -0.3420, \ \mathcal{T}_{2233} = 0.2127, \ \mathcal{T}_{2333} = 0.2727, \ \mathcal{T}_{3333} = -0.3054.$$

The results from the application of the general HOPM and its symmetric version, S-HOPM, are depicted in Figure 4.1. The curve for the general HOPM depicts the values of $\langle \boldsymbol{\mathcal{T}}, \boldsymbol{u}_k^{(1)} \star \boldsymbol{u}_k^{(2)} \star \boldsymbol{u}_k^{(3)} \star \boldsymbol{u}_k^{(4)} \rangle$. $g(\boldsymbol{u}_k)$ is plotted for the S-HOPM. Both algorithms are initialized via HOSVD. It is seen that the S-HOPM iterations do not converge.

We will show, however, that Algorithm 2 is convergent if $N = 2L$ is even and $g(\cdot)$ is a *convex (or concave)* function of $\boldsymbol{u}$. Let us recall the meaning of this property [25].

DEFINITION 9 (convex (concave) function). *Let g be a function whose values are real or $\pm\infty$ and whose domain is a convex subset S of $\mathbb{R}^M$. Then g is said to be*

convex on S if its epigraph,

$$\mathrm{epi}g \stackrel{\triangle}{=} \{(\boldsymbol{u}, \nu) | \boldsymbol{u} \in S, \nu \in \mathbb{R}, \nu \geq g(\boldsymbol{u})\},$$

is a convex subset of $\mathbb{R}^{M+1}$. *A concave function on S is a function whose negative is convex.*

From the definition of g, and using the square matrix unfolding of $\boldsymbol{\mathcal{T}}$, it is readily verified that $g(\boldsymbol{u})$ can be written as a polynomial matrix form:

$$g(\boldsymbol{u}) = \underbrace{(\boldsymbol{u} \otimes \boldsymbol{u} \otimes \cdots \otimes \boldsymbol{u})^T}_{L \text{ times}} \boldsymbol{T} \underbrace{(\boldsymbol{u} \otimes \boldsymbol{u} \otimes \cdots \otimes \boldsymbol{u})}_{L \text{ times}}.$$

A necessary and sufficient condition for a twice continuously differentiable function $g(\boldsymbol{u})$ to be convex (concave) on an open convex subset C of $\mathbb{R}^M$ is that its Hessian (i.e., second derivative) matrix be positive (negative) semidefinite on C [25]. Hence, g above is convex (concave) on $\mathbb{R}^M$ if and only if the matrix

$$(4.1) \qquad (\boldsymbol{I} \otimes \underbrace{\boldsymbol{u} \otimes \boldsymbol{u} \otimes \cdots \otimes \boldsymbol{u}}_{L-1 \text{ times}})^T \boldsymbol{T} (\boldsymbol{I} \otimes \underbrace{\boldsymbol{u} \otimes \boldsymbol{u} \otimes \cdots \otimes \boldsymbol{u}}_{L-1 \text{ times}})$$

is positive (negative) semidefinite for all $\boldsymbol{u} \in \mathbb{R}^M$. This implies that $g(\boldsymbol{u})$ has to be nonnegative (nonpositive) for all $\boldsymbol{u}$.

The above condition on the matrix (4.1) will be satisfied *if* $\boldsymbol{T}$ is positive (negative) semidefinite.[6] For example, this holds for the fourth-order cumulant tensor of the output of a linear mixing system, $\boldsymbol{y} = \boldsymbol{Hx} + \boldsymbol{n}$, whose sources have kurtoses of the same sign. In that case, $\boldsymbol{\mathcal{T}}$ is given by [16]

$$(4.2) \qquad \boldsymbol{\mathcal{T}} = \boldsymbol{H} \stackrel{\mathcal{S}}{\star} \boldsymbol{H} \stackrel{\mathcal{S}}{\star} \boldsymbol{H} \stackrel{\mathcal{S}}{\star} \boldsymbol{H}$$

with $\boldsymbol{\mathcal{S}}$ denoting the (diagonal) tensor of the fourth-order cumulants of $\boldsymbol{x}$. The corresponding matrix $\boldsymbol{T}$ is given by

$$(4.3) \qquad \boldsymbol{T} = (\boldsymbol{H} \odot \boldsymbol{H})\mathrm{diag}(\mathrm{cum}_4(x_i))(\boldsymbol{H} \odot \boldsymbol{H})^T,$$

where $\odot$ is the Khatri–Rao (columnwise Kronecker) product [27].

Notice that even if g is convex (concave), this does not hold for the quotient

$$\frac{g(\boldsymbol{u})}{\|\boldsymbol{u}\|^N}$$

since the unit sphere

$$\Sigma \stackrel{\triangle}{=} \{\boldsymbol{u} \in \mathbb{R}^M | \|\boldsymbol{u}\| = 1\}$$

is not a convex set.

Example 2. Consider the supersymmetric $3 \times 3 \times 3 \times 3$ tensor $\boldsymbol{\mathcal{T}}$ that contains the fourth-order cumulants of the mixture observations for the 3×7 mixing matrix

$$\boldsymbol{H} = \begin{bmatrix} -0.3912 & 0.1427 & 0.3087 & 0.2511 & -0.5408 & 0.3692 & 0.4894 \\ -0.6743 & -0.3816 & -0.5317 & -0.1942 & -0.2120 & -0.0770 & -0.1687 \\ 0.4947 & -0.0364 & -0.3621 & 0.2594 & -0.6336 & 0.1911 & -0.3430 \end{bmatrix}$$

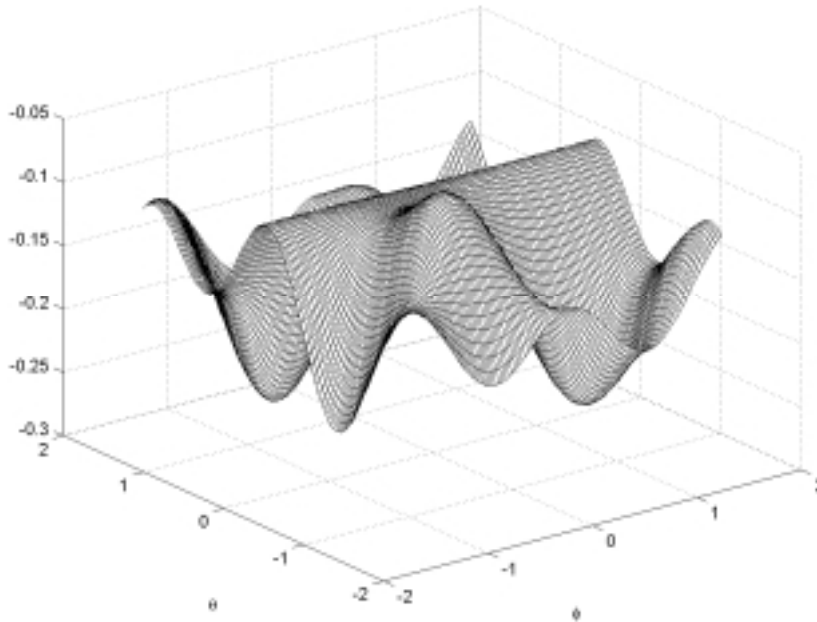[6]For second-order tensors (matrices) this is also a necessary condition.

FIG. 4.2. *The function $g(\boldsymbol{u})$ for the tensor of Example 2, restricted to the unit sphere.*

and source fourth-order cumulants

$$(\mathrm{cum}_4(x_i))_{i=1}^7 = (-0.3753, -0.3087, -0.7600, -0.0227, -0.4633, -0.0143, -0.5470).$$

The resulting function g is concave and nonpositive. Its restriction to the unit sphere is shown in Figure 4.2. To allow a three-variate function to be plotted, we have parameterized the unit-norm vector $\boldsymbol{u}$ as $\boldsymbol{u} = \begin{bmatrix} \cos(\theta) & \sin(\theta)\cos(\phi) & \sin(\theta)\sin(\phi) \end{bmatrix}^T$ where the angles θ, ϕ were normalized to the interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right]$, i.e., the cosines were constrained to be nonnegative. This can be done since the value assumed by $g(\boldsymbol{u})$ is invariant to sign changes. It is clearly seen that this is no longer concave.

THEOREM 4 (convergence of S-HOPM). *For any supersymmetric Nth-order $M \times M \times \cdots \times M$ tensor $\boldsymbol{\mathcal{T}}$ such that N is even and the associated function g is convex (concave) on $\mathbb{R}^M$, Algorithm 2 converges to a local maximum (minimum) of the restriction of g to the unit sphere, Σ, for any initialization, except for saddle points and crest lines leading to such saddle points.*

Proof. Consider first the case that g is convex. This assumption implies that the set epi$g$ is a convex subset of $\mathbb{R}^{M+1}$, and hence a tangent hyperplane at any point $(\boldsymbol{v}, g(\boldsymbol{v}))$ is in fact a supporting hyperplane of epi$g$. This fact is expressed by the so-called *(sub)gradient inequality* [25]

$$g(\boldsymbol{v}_2) - g(\boldsymbol{v}_1) \geq \langle \boldsymbol{v}_2 - \boldsymbol{v}_1, \nabla g(\boldsymbol{v}_1) \rangle$$

holding *for any* vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^M$ (regardless of how distant they may be). To apply this to the problem at hand, set $\boldsymbol{v}_2 = \boldsymbol{u}_k$ and $\boldsymbol{v}_1 = \boldsymbol{u}_{k-1}$ to obtain

$$(4.4) \qquad g(\boldsymbol{u}_k) - g(\boldsymbol{u}_{k-1}) \geq \langle \boldsymbol{u}_k, \nabla g(\boldsymbol{u}_{k-1}) \rangle - \langle \boldsymbol{u}_{k-1}, \nabla g(\boldsymbol{u}_{k-1}) \rangle.$$

What we want to show is that if we are not at a stationary point, g is increasing monotonically. It suffices then to show that the right-hand side of (4.4) is positive if

$\boldsymbol{u}_k \neq \boldsymbol{u}_{k-1}$. Note that, for any unit-norm vector $\boldsymbol{u}$, the Cauchy–Schwarz inequality yields

$$(4.5) \qquad \langle \boldsymbol{u}, \nabla g(\boldsymbol{u}_{k-1}) \rangle \leq \|\nabla g(\boldsymbol{u}_{k-1})\|,$$

where the equality holds if and only if $\boldsymbol{u} = \frac{\nabla g(u_{k-1})}{\|\nabla g(u_{k-1})\|}$. But this is precisely the formula that gives $\boldsymbol{u}_k$ in Algorithm 2. Hence

$$\langle \boldsymbol{u}_k, \nabla g(\boldsymbol{u}_{k-1}) \rangle - \langle \boldsymbol{u}_{k-1}, \nabla g(\boldsymbol{u}_{k-1}) \rangle > 0,$$

which, in view of (4.4), implies that $g(\boldsymbol{u}_k)$ is increasing with k. The convergence follows from the fact that the restriction of $|g|$ to Σ is bounded from above, namely,

$$(4.6) \qquad \left| \frac{g(\boldsymbol{u})}{\|\boldsymbol{u}\|^N} \right| = \left| \frac{(\boldsymbol{u} \otimes \boldsymbol{u} \otimes \cdots \otimes \boldsymbol{u})^T \boldsymbol{T} (\boldsymbol{u} \otimes \boldsymbol{u} \otimes \cdots \otimes \boldsymbol{u})}{(\boldsymbol{u} \otimes \cdots \otimes \boldsymbol{u})^T (\boldsymbol{u} \otimes \cdots \otimes \boldsymbol{u})} \right| \leq |\lambda_1|,$$

with λ_1 denoting the eigenvalue of $\boldsymbol{T}$ with largest modulus.

The case of g being concave can be treated as above, by replacing g with $-g$. The only point that we need to comment on is that the quantity $\langle \boldsymbol{u}, \nabla g(\boldsymbol{u}_{k-1}) \rangle$ in (4.5) now has to take its minimum value, which occurs when $\boldsymbol{u} = -\frac{\nabla g(u_{k-1})}{\|\nabla g(u_{k-1})\|}$. However, the minus sign is not necessary as it does not affect the value of $g(\boldsymbol{u}_k)$ (recall that N is even). □

A distinction is made in [7] between the HOPM, as derived from the Lagrangian equations for the corresponding constrained minimization problem, and a gradient descent procedure. However, it can be shown, following similar arguments to those employed in [19] for the case of a norm function $g(\cdot)$, that the S-HOPM is in fact a gradient recursion using a strategic choice for the step-size parameter.

5. Properties of the matrix $\boldsymbol{T}$. In this section we will state and prove some properties that the supersymmetry of a tensor implies for its square matrix unfolding. The so-called *vec-permutation* matrix [13] plays a central role to the subsequent analysis.

DEFINITION 10 (vec-permutation matrix). *The vec-permutation matrix, $\boldsymbol{I}_{m,n}$, is defined as the $mn \times mn$ permutation matrix that satisfies the equality*

$$(5.1) \qquad \boldsymbol{I}_{m,n} \mathrm{vec}(\boldsymbol{A}) = \mathrm{vec}(\boldsymbol{A}^T)$$

for all $m \times n$ matrices $\boldsymbol{A}$.

An explicit way of defining $\boldsymbol{I}_{m,n}$ is given below as a theorem [13].

THEOREM 5 (explicit characterization of $\boldsymbol{I}_{m,n}$). *The (i,j) entry of $\boldsymbol{I}_{m,n}$ is equal to unity if $k = l'$ and $l = k'$ with*

$$i = (k-1)n + l, \quad 1 \leq k \leq m, \quad 1 \leq l \leq n,$$
$$j = (k'-1)m + l', \quad 1 \leq k' \leq n, \quad 1 \leq l' \leq m,$$

and is zero otherwise.

THEOREM 6 (properties of $\boldsymbol{T}$ (general even order)). *The square matrix unfolding, $\boldsymbol{T}$, of an $M \times M \times \cdots \times M$ supersymmetric tensor $\boldsymbol{\mathcal{T}}$ of order $N = 2L$ satisfies the following properties:*
   (i) $\boldsymbol{T}^T = \boldsymbol{T}$.
   (ii) $\boldsymbol{I}_{M^{L-1}, M} \boldsymbol{T} = \boldsymbol{T}$.

Proof. Recall the definition of $\boldsymbol{T}$ from (2.1)–(2.3). The symmetry of $\boldsymbol{T}$ follows easily from the fact that $\mathcal{T}_{i_1,i_2,\ldots,i_L,j_1,j_2,\ldots,j_L} = \mathcal{T}_{j_1,j_2,\ldots,j_L,i_1,i_2,\ldots,i_L}$.

Call $\boldsymbol{T}'$ the matrix $\boldsymbol{I}_{M^{L-1},M}\boldsymbol{T}$ and consider any of its entries, say $T'_{p,q}$. Write its indices as

$$p = M^{L-1}(k-1) + l, \quad 1 \le k \le M, \quad 1 \le l \le M^{L-1},$$
$$q = M^{L-1}(q_1 - 1) + \cdots + M(q_{L-1} - 1) + q_L, \quad 1 \le q_i \le M, \quad 1 \le i \le L.$$

Then, it is readily seen that Theorem 5 implies

$$T'_{p,q} = T_{M(l-1)+k,q}.$$

Writing further l in the form

$$l = M^{L-2}(l_1 - 1) + \cdots + M(l_{L-2} - 1) + l_{L-1}, \quad 1 \le l_i \le M,$$

and using the supersymmetry of $\mathcal{T}$, the above yields

$$\begin{aligned} T'_{p,q} &= T_{M^{L-1}(l_1-1)+\cdots+M(l_{L-1}-1)+k,q} \\ &= \mathcal{T}_{l_1,l_2,\ldots,l_{L-1},k,q_1,q_2,\ldots,q_L} \\ &= \mathcal{T}_{k,l_1,l_2,\ldots,l_{L-1},q_1,q_2,\ldots,q_L} \\ &= T_{M^{L-1}(k-1)+M^{L-2}(l_1-1)+\cdots+l_{L-1},q} \\ &= T_{M^{L-1}(k-1)+l,q} \\ &= T_{p,q}. \end{aligned}$$

This proves (ii). $\quad\square$

Being symmetric, $\boldsymbol{T}$ admits an eigenvalue decomposition with M^L real eigenvalues and orthonormal eigenvectors [28],

$$(5.2) \qquad\qquad \boldsymbol{T} = \sum_{i=1}^{M^L} \lambda_i \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T,$$

where

$$\boldsymbol{\xi}_i^T \boldsymbol{\xi}_j = \delta(i,j) \text{ for all } i,j$$

and the eigenvalues are numbered such that

$$|\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_{M^L}|.$$

Let us now confine our attention to fourth-order supersymmetric tensors ($L = 2$). We shall denote the corresponding permutation matrix $\boldsymbol{I}_{M,M}$ by $\boldsymbol{P}$. It is easy to see that $\boldsymbol{P}$ is symmetric. Let us also define the $M \times M$ matrices $\boldsymbol{\Xi}_i$ as the matrix versions of the corresponding eigenvectors of $\boldsymbol{T}$, i.e.,

$$\boldsymbol{\Xi}_i = \text{unvec}(\boldsymbol{\xi}_i).$$

For this special, yet important case, some more information on $\boldsymbol{T}$, involving its eigenstructure, can be revealed [23].

THEOREM 7 (properties of $\boldsymbol{T}$ ($N = 4$)). *The square matrix unfolding of any fourth-order supersymmetric tensor, $\mathcal{T}$, satisfies the following properties:*

(i) $PT = TP = PTP = T$.

(ii) *It admits $\frac{M(M+1)}{2}$ eigenvectors $\boldsymbol{\xi}$ with* positive *symmetry, i.e., $P\boldsymbol{\xi} = +\boldsymbol{\xi}$, and $\frac{M(M-1)}{2}$ eigenvectors with* negative *symmetry, i.e., $P\boldsymbol{\xi} = -\boldsymbol{\xi}$. The corresponding matrices $\boldsymbol{\Xi}$ are symmetric ($\boldsymbol{\Xi} = \boldsymbol{\Xi}^T$) and skew-symmetric ($\boldsymbol{\Xi}^T = -\boldsymbol{\Xi}$), respectively.*

(iii) *All eigenvectors of T having negative symmetry must correspond to a zero eigenvalue.*

(iv) $\mathrm{rank}(T) \leq \frac{M(M+1)}{2}$.

Proof.

(i) The proof of (i) follows easily from the symmetry of T, P and property (ii) of Theorem 6.

(ii) We will first prove that all eigenvectors $\boldsymbol{\xi}$ of T enjoy one of the above symmetries. Take the ith eigenpair

$$T\boldsymbol{\xi}_i = \lambda_i \boldsymbol{\xi}_i.$$

Premultiplying the above equation by P and taking into account the equality $PT = TP$ yields

(5.3) $$T \cdot P\boldsymbol{\xi}_i = \lambda_i \cdot P\boldsymbol{\xi}_i,$$

which shows that $P\boldsymbol{\xi}_i$ is also an eigenvector of T for the eigenvalue λ_i. If λ_i is simple, then the corresponding eigenvector is unique up to a sign factor, and hence $P\boldsymbol{\xi}_i = \pm\boldsymbol{\xi}_i$. In the case of a multiple eigenvalue, one can always choose an eigenvector from its invariant space that has the desired symmetry. For example, a possible choice could be the normalized version of $\boldsymbol{\xi}_i' = P\boldsymbol{\xi}_i \pm \boldsymbol{\xi}_i$. This eigenvector is seen to be orthogonal to the rest.

It follows from the definition of P (cf. (5.1)) that the symmetries $P\boldsymbol{\xi}_i = \pm\boldsymbol{\xi}_i$ satisfied by the eigenvectors of T are equivalent to $\boldsymbol{\Xi}_i^T = \pm\boldsymbol{\Xi}_i$, respectively.

Introduce now the following two subspaces of $\mathbb{R}^{M^2}$:

$$S^+ = \{x \in \mathbb{R}^{M^2} | Px = +x\},$$
$$S^- = \{x \in \mathbb{R}^{M^2} | Px = -x\}.$$

These subspaces are orthogonal to each other since if $x \in S^+$ and $y \in S^-$, then the orthogonality of P [13] implies

$$x^T y = -x^T P^T P y = -x^T y;$$

hence $x^T y = 0$. It also follows that

$$\dim S^+ = \frac{M(M+1)}{2},$$
$$\dim S^- = \frac{M(M-1)}{2},$$

since, as shown above, parameterizing S^+ (resp., S^-) is equivalent to parameterizing the set of symmetric (resp., skew-symmetric) $M \times M$ matrices. Since $\dim S^+ + \dim S^- = M^2$, we can write the orthogonal decomposition as

$$\mathbb{R}^{M^2} = S^+ \oplus S^-.$$

The result then follows from the fact that the eigenvectors of $\boldsymbol{T}$ belong to either S^+ or S^- and form an orthonormal basis of $\mathbb{R}^{M^2}$.[7]

(iii) Using the property $\boldsymbol{TP} = \boldsymbol{T}$ in (5.3) yields $\boldsymbol{T\xi}_i = \lambda_i \boldsymbol{P\xi}_i$. If $\boldsymbol{\xi}_i$ is such that $\boldsymbol{P\xi}_i = -\boldsymbol{\xi}_i$, then it follows that $\lambda_i \boldsymbol{\xi}_i = -\lambda_i \boldsymbol{\xi}_i$, and $\lambda_i = 0$.

(iv) The proof of (iv) follows directly from (ii) and (iii). $\quad\square$

COROLLARY 1 (dominant eigenvector of $\boldsymbol{T}$ ($N = 4$)). *The dominant eigenvector,* $\boldsymbol{\xi}_1$, *of the square matrix unfolding of a nonzero supersymmetric fourth-order tensor satisfies* $\boldsymbol{P\xi}_1 = +\boldsymbol{\xi}_1$. *Equivalently, its* $M \times M$ *matrix version,* $\boldsymbol{\Xi}_1$, *is symmetric.*

Proof. The proof follows from Theorem 7(iii) since $\lambda_1 \neq 0$. $\quad\square$

6. New initialization. We derive here an alternative initialization scheme for the S-HOPM for fourth-order tensors that is observed to be more effective than that based on the HOSVD in approaching the globally optimum point. The starting point is the inequality (4.6) that becomes an equality if and only if $\boldsymbol{u} \otimes \boldsymbol{u}$ coincides with a dominant eigenvector, $\pm\boldsymbol{\xi}_1$. That is, the global maximum of $h(\boldsymbol{u})$ (cf. (3.3)) would be attained if $\boldsymbol{\xi}_1$ could be written as a "Kronecker square," something which is in general not true.

Nonetheless, this remark suggests a way of computing an initial estimate for $\boldsymbol{u}$, namely, setting it equal to the best, in the LS sense, "Kronecker square root" of $\boldsymbol{\xi}_1$:

$$\boldsymbol{u} = \arg \min_{\varsigma \in \mathbb{R}, \|s\|=1} \|\boldsymbol{\xi}_1 - \varsigma \boldsymbol{s} \otimes \boldsymbol{s}\|.$$

Equivalently,

$$\boldsymbol{u} = \arg \min_{\varsigma \in \mathbb{R}, \|s\|=1} \|\boldsymbol{\Xi}_1 - \varsigma \boldsymbol{s}\boldsymbol{s}^T\|.$$

Since $\boldsymbol{\Xi}_1$ is symmetric (see Corollary 1), the latter problem is solved by setting $\boldsymbol{u}$ equal to the unit-norm eigenvector of $\boldsymbol{\Xi}_1$ that corresponds to its absolutely largest eigenvalue, say ς_1 [11]. Note that $\varsigma_1^2 \leq \|\boldsymbol{\Xi}_1\|^2 = \|\boldsymbol{\xi}_1\|^2 = 1$.

The proposed initialization method thus involves two symmetric matrix rank-1 approximation problems:

New Initialization

1. $\boldsymbol{\xi}_1 =$ dominant eigenvector of $\boldsymbol{T}$.
2. $\boldsymbol{u}_0 =$ dominant eigenvector of $\mathrm{unvec}(\boldsymbol{\xi}_1)$.

Using (5.2), $g(\boldsymbol{u}_0)$ can be written as

$$\begin{aligned} g(\boldsymbol{u}_0) &= \langle \boldsymbol{u}_0 \otimes \boldsymbol{u}_0, \boldsymbol{T}(\boldsymbol{u}_0 \otimes \boldsymbol{u}_0) \rangle \\ &= \sum_{i=1}^{M^2} \lambda_i ((\boldsymbol{u}_0 \otimes \boldsymbol{u}_0)^T \boldsymbol{\xi}_i)^2 \\ &= \sum_{i=1}^{M(M+1)/2} \lambda_i (\boldsymbol{u}_0^T \boldsymbol{\Xi}_i \boldsymbol{u}_0)^2, \end{aligned}$$

---

[7]This also proves that the vec-permutation matrix $P \stackrel{\triangle}{=} \boldsymbol{I}_{M,M}$ has eigenvalues ± 1 with multiplicities $\frac{M(M\pm 1)}{2}$, respectively, a property stated in [13].

where use was made of the well-known identity $\mathrm{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^T \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{B})$ [13] and the fact that $\lambda_i = 0$ for $i > \frac{M(M+1)}{2}$. Choosing $\boldsymbol{u}_0$ as above yields

$$g(\boldsymbol{u}_0) = \lambda_1 \varsigma_1^2 + \sum_{i=2}^{M(M+1)/2} \lambda_i (\boldsymbol{u}_0^T \boldsymbol{\Xi}_i \boldsymbol{u}_0)^2.$$

If $\boldsymbol{T}$ is sign (semi)definite (g is then convex/concave), the latter relation implies

(6.1) $$|g(\boldsymbol{u}_0)| \geq |\lambda_1|\varsigma^2.$$

This, in conjunction with (4.6), provides us with lower and upper bounds on the initial value of h, as follows.

THEOREM 8 (bounds on initial value). *The value assumed by h with the suggested initialization, when applied to a supersymmetric tensor $\boldsymbol{T}$ with sign (semi)definite square matrix unfolding, $\boldsymbol{T}$, is bounded as*

$$\lambda_1^2 \varsigma_1^4 \leq h(\boldsymbol{u}_0) \leq \lambda_1^2,$$

where λ_1 and ς_1 are the absolutely largest eigenvalues of the matrices $\boldsymbol{T}$ and $\boldsymbol{\Xi}_1$, respectively. This initial value approaches the global maximum as the vector $\boldsymbol{\xi}_1$ approaches Kronecker decomposability (i.e., as $|\varsigma_1|$ approaches one).

Example 2 *(continued).* Consider the $3 \times 3 \times 3 \times 3$ tensor given before. The corresponding matrix $\boldsymbol{T}$ is found to have the following singular values:

$$(|\lambda_i|)_{i=1}^9 = (0.2841, 0.2617, 0.2305, 0.0353, 0.0020, 0.0001, 0, 0, 0),$$

agreeing with Theorem 7(iii).[8] We ran the S-HOPM for this tensor, using both the HOSVD-based and the new initialization methods. The results are shown in Figure 6.1. In both cases the iterations converge to a global minimum. Nevertheless, the new initialization scheme is seen to lie much closer to the globally optimal solution than the HOSVD-based scheme. In fact, for this example successive iterations are nearly superfluous. Extensive simulations have shown this to be the typical case for tensors with an associated functional that is convex/concave.

The superior performance of the new method can also be seen in Figure 6.2 where the position of the two initial estimates in the parameter space, as well as the trajectories followed in each case, are shown.

The lower and upper bounds given in Theorem 8 are 0.0444 and 0.0807, respectively, and are seen to be satisfied by the initial value assumed by h in the new initialization scheme, namely, $h(\boldsymbol{u}_0) = 0.0758$. Note that the initial value suggested by the HOSVD-based scheme, namely, 0.0183, does not meet the lower bound.

We have found, however, some examples where one of the initialization methods leads to a local extremum. These cases are rare and, moreover, in all of them h assumes at the suboptimal point a value quite close to the optimal one.

Example 3. Let the tensor $\boldsymbol{T}$ be given by (4.2) with

$$\boldsymbol{H} = \begin{bmatrix} -0.1413 & -0.8318 & -0.0769 & -0.1434 & 0.4681 & 0.2054 & 0.0210 \\ 0.3194 & 0.0328 & 0.6555 & 0.1696 & 0.0224 & 0.6580 & 0.0716 \\ 0.4123 & -0.4371 & 0.1749 & -0.3828 & -0.6389 & -0.2315 & -0.0065 \end{bmatrix}$$

---

[8]Note that there are 6 nonzero eigenvalues, whereas there are 7 sources. The number of nonzero eigenvalues of $\boldsymbol{T}$ can reveal the number of (kurtic) sources only in the case that $K \leq \frac{M(M+1)}{2}$.
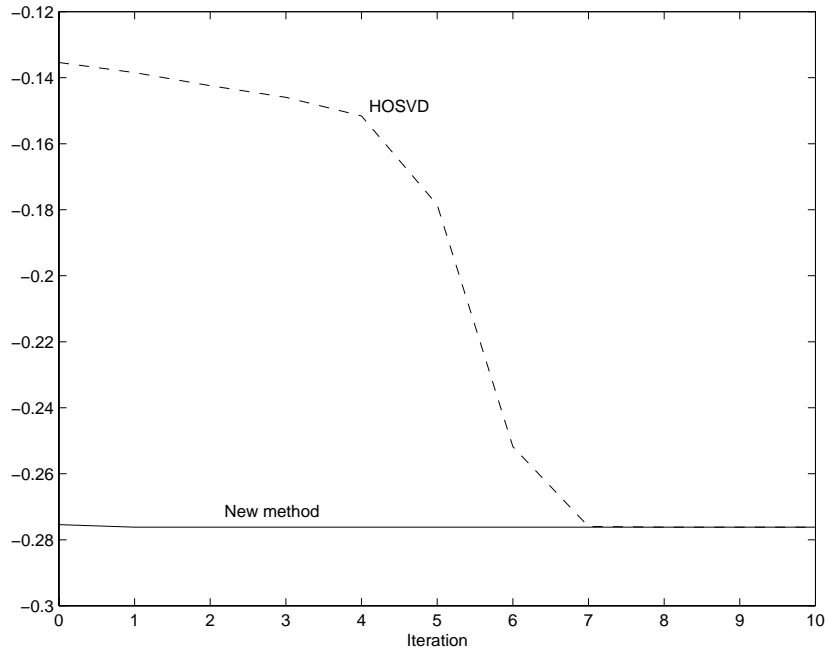
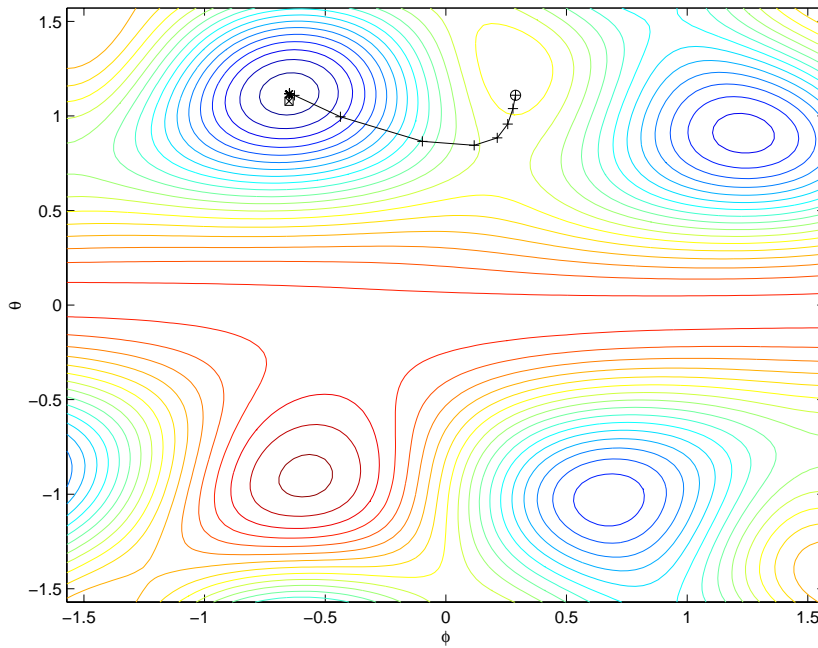FIG. 6.1. *Results of the S-HOPM for both the HOSVD-based and new initializations.*

FIG. 6.2. *Visualization of the S-HOPM algorithm for both initialization methods (Example 2). The HOSVD-based initial estimate is denoted by a small circle and the trajectory followed by +'s. The initial estimate provided by the proposed method is denoted by a small square and the subsequent estimates by ×'s.*
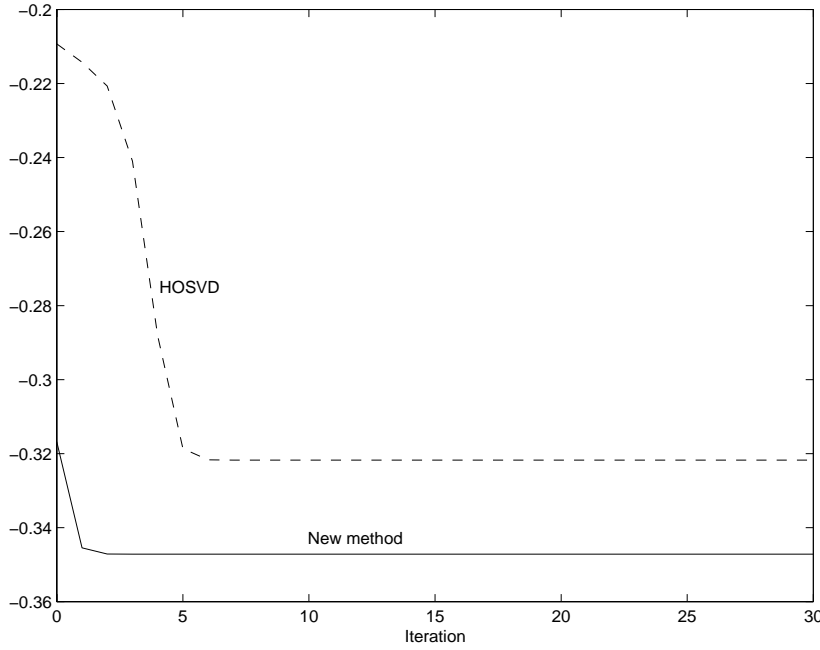
FIG. 6.3. *Results of S-HOPM for an initial estimate based on HOSVD and the new method. In the former case, the algorithm is trapped to a local minimum.*

and the source cumulants

$$(\mathrm{cum}_4(x_i))_{i=1}^7 = (-0.1204, -0.4336, -0.0961, -0.8479, -0.7684, -0.8408, -0.9204).$$

As shown in Figure 6.3, the S-HOPM, initialized with the aid of the HOSVD, is trapped to a local minimum.

The bounds of Theorem 8 are 0.0537 and 0.1272 and the new initialization method yields an initial value of 0.1004 for h. Again, the initial value provided by the HOSVD-based method, namely, 0.0438, does not satisfy the lower bound.

Example 4. The converse is seen to happen for the tensor built as in (4.2) with

$$\boldsymbol{H} = \begin{bmatrix} -0.5100 & 0.3056 & 0.2035 & 0.1959 & 0.4809 & 0.3216 & 0.4816 \\ 0.4881 & -0.4607 & 0.5045 & -0.2727 & 0.2863 & 0.2995 & 0.2211 \\ -0.0529 & -0.4287 & -0.2190 & 0.5228 & -0.3968 & 0.5673 & 0.1133 \end{bmatrix}$$

and

$$(\mathrm{cum}_4(x_i))_{i=1}^7 = (-0.4173, -0.3469, -0.2225, -0.2766, -0.5792, -0.4679, -0.7488).$$

Figure 6.4 shows that the new initialization method leads to a local minimum this time. Nevertheless, as we can see in Figure 6.5 there is no significant difference in the values of h at the local and global minima. Figure 6.6 shows the evolution of the algorithm on the parameter space for the two initializations. In this example, the bounds of Theorem 8, 0.0092 and 0.0387, are met by the initial values for h provided by both the new initialization method and the HOSVD-based one, namely, 0.0181 and 0.0174, respectively.
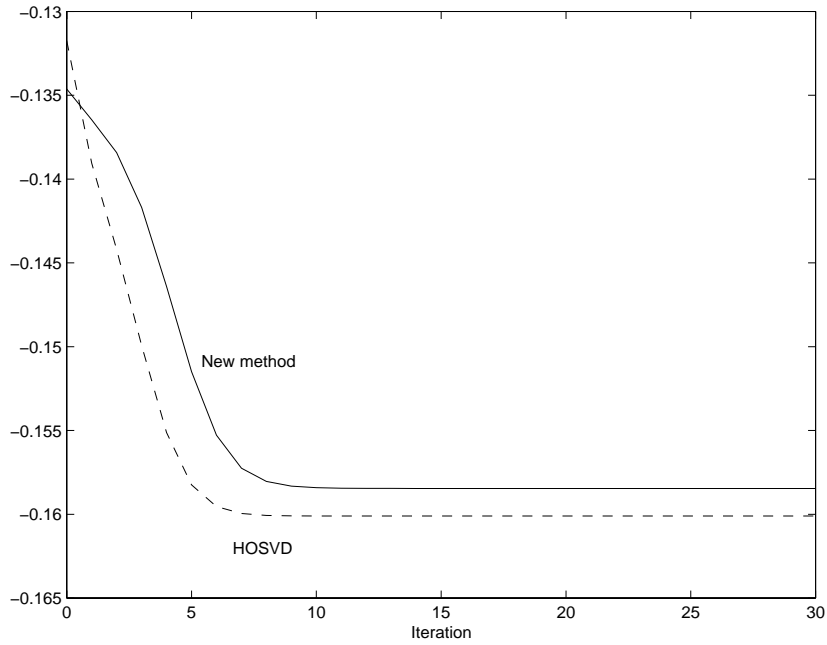
FIG. 6.4. *Results of S-HOPM for both the HOSVD-based and the new initialization methods. The latter leads to a local minimum.*
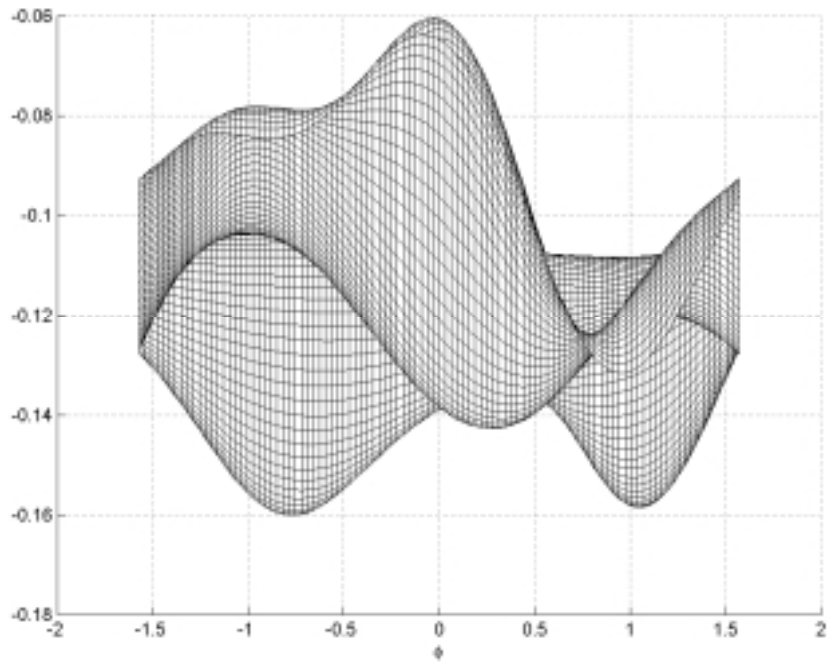

FIG. 6.5. *The function g for the tensor of Example* 4. *Notice the local minimum, corresponding to a value of the function quite close to its global minimum one.*
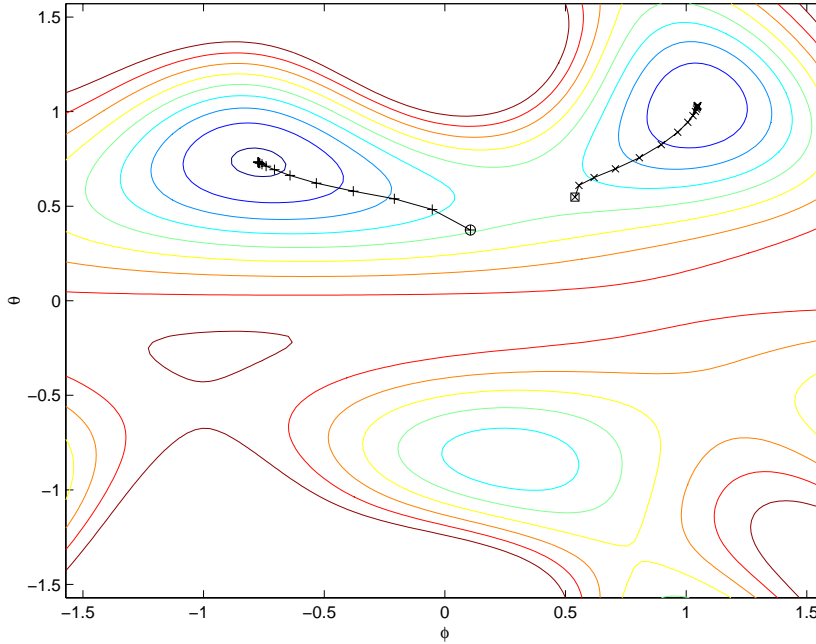
FIG. 6.6. *Visualization of the initial estimates and the trajectories followed by the S-HOPM in Example* 4, *for both initialization methods. Symbols are as in Figure* 6.2.

7. On the best rank-R approximation. It is known that by successively subtracting the LS rank-1 approximation from a given matrix R times results in its LS rank-R approximation [11]. One could wonder whether this fact still holds for higher-order tensors, as pointed out in [3]. Unfortunately, as the following (typical) example demonstrates, this is not the case.

Example 5. Take the tensor $\mathcal{T}$ described in Example 2, normalized to unit norm, and determine its best rank-1 approximant, $\hat{\mathcal{T}}$. Then do the same for the tensor $\mathcal{T} - \hat{\mathcal{T}}$, and so on. As shown in Figure 7.1, the norm of the residue indeed decreases to practically zero; however, this is done in about 200 iterations, while $\mathcal{T}$ has rank 7. It is also of interest to note that the rank of T remains equal to 6 all the way through. □

The same iterative process, but with every new rank-1 term being constrained to be orthogonal to the previous ones, was recently studied in [17]. Depending on the definition of the orthogonality adopted, this fails or is not certain to provide a valid rank-R approximation scheme. The above process would work in the case of a tensor with rank less than or equal to its dimension, M. This is the case in the BSS context when the mixing matrix is "tall," i.e., $M \geq K$. In fact, in such a case, all rank($\mathcal{T}$) rank-1 terms can be jointly determined by minimizing the norm $\|\mathcal{T} - \boldsymbol{H}^{\overset{S}{\star}N}\|$ subject to the constraint that $\boldsymbol{S}$ is diagonal and $\boldsymbol{H}$ has full column rank (usually assumed to have orthonormal columns). Works coping with this problem include [15, 21]. An algebraic approach via joint diagonalization of the matrices $\boldsymbol{\Xi}_i$ as defined above is proposed in [2]. Note that step 2 of our initialization method is part of the diagonalization of $\boldsymbol{\Xi}_1$. Nonetheless, it is the problem of the recovery of a single source (rank-1 approximation) that our method addresses. Moreover, it is applicable to BSS problems with "fat" mixing matrices ($M < K$) as well (see Example 2).
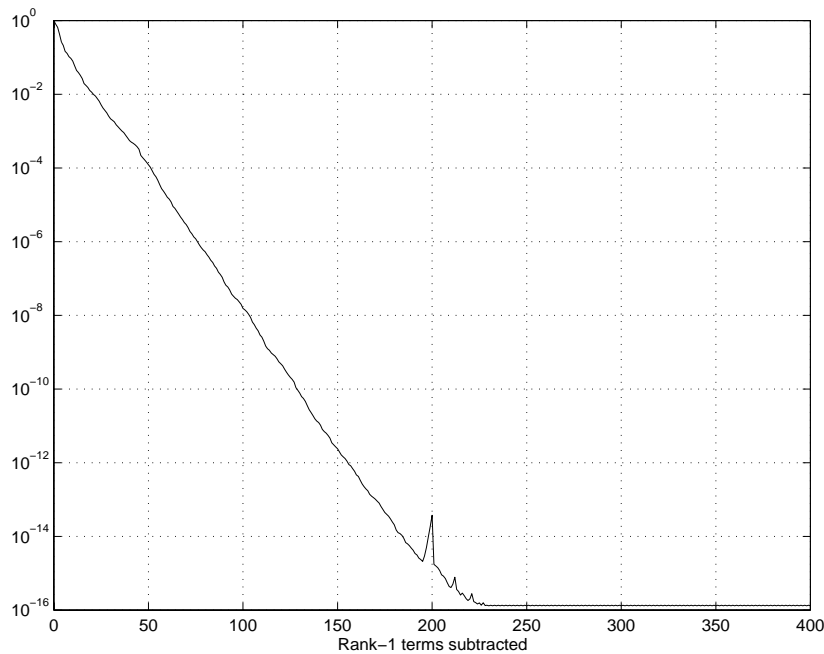
FIG. 7.1. *Norm of the residue remaining when successively subtracting rank-1 terms from the tensor of Example 5. The spike near iteration 200 may be due to numerical artifacts.*

For the latter more challenging problem, [1] develops an algebraic method for determining $\boldsymbol{H}$, based on the assumption of linear independence of the projectors on the spaces spanned by its columns. This can be seen to be equivalent to the matrix $\boldsymbol{H} \odot \boldsymbol{H}$, which occurs in (4.3), having full column rank.

Both algebraic- and optimization-based approaches for expanding a supersymmetric tensor in a sum of rank($\boldsymbol{\mathcal{T}}$) rank-1 terms have been developed in [6] based on its representation in terms of an homogeneous polynomial. The problem then becomes one of expressing the polynomial as a sum of powers of linear forms. Workable algorithms derived this way appear limited to small-sized tensors [3].

Finally, the HOPM can be viewed as a special case of the alternating least squares (ALS) iterative approach common in problems of multilinear model fitting (PARAFAC or CANDECOMP [18, 7, 27]) to multidimensional data. The ALS method can be used with no change to compute a rank-R approximation to a supersymmetric tensor as well, although this way the rich symmetry in the problem is not exploited.

8. Conclusions. The problem of computing the best, in the LS sense, rank-1 approximation to a given Nth-order supersymmetric tensor has been studied in this paper. A symmetric version of the higher-order power method, which was thought to be unreliable, has been shown to be convergent for tensors whose associated polynomial form is convex or concave. A new method for initializing the iterations has been developed for the fourth-order case and observed in extensive simulations to provide an estimate that lies closer to the globally optimal solution than that yielded by the HOSVD. Moreover, the proximity to the optimal solution is a priori quantifiable. It happens, though rarely, that the initial estimate provided by either the HOSVD-based

scheme or the one proposed here be in the basin of attraction of a locally optimal solution. However, this is not a serious problem since in all such cases encountered, the quality of approximation corresponding to the local optimum is observed to be quite close to the best attainable. As a byproduct of our study of the rank-1 approximation problem, some properties satisfied by the square matrix unfolding of *any* supersymmetric tensor have also been derived.

The applicability of the symmetric higher-order power method is accompanied by a significant reduction to the computational complexity of its general version. The convexity/concavity assumptions required to prove its convergence are plausible in many signal processing applications, such as in blind separation of multiuser communications channels, where all source signals have kurtoses of the same sign.

It is not an easy task to extend the results obtained here to the more general problem of computing the best rank-R approximation to a supersymmetric tensor, when $R > 1$. Simply imposing the symmetry constraint on the ALS method of fitting PARAFAC models to general tensors, in the same way that HOPM gave rise to the S-HOPM, does not always result in a convergent procedure.

REFERENCES

[1] J.-F. CARDOSO, *Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors*, in Proceedings of the IEEE International Conference on Acoust., Speech, and Signal Processing (ICASSP'91), Toronto, Canada, 1991.

[2] J.-F. CARDOSO AND A. SOULOUMIAC, *Blind beamforming for non-Gaussian signals*, Proc. IEE-F, 140 (1993), pp. 362–370.

[3] P. COMON, *Blind channel identification and extraction of more sources than sensors*, in Proc. SPIE Conference on Advanced Signal Processing VIII, San Diego, 1998.

[4] P. COMON, *Tensor decompositions—State of the art and applications*, keynote address in IMA Conf. Mathematics in Signal Processing, Warwick, UK, 2000.

[5] P. COMON AND P. CHEVALIER, *Blind source separation: Models, concepts, algorithms, and performance*, in Unsupervised Adaptive Filtering, Vol. I, S. Haykin, ed., John Wiley, New York, 2000.

[6] P. COMON AND B. MOURRAIN, *Decomposition of quantics in sums of powers of linear forms*, Signal Process., 53 (1996), pp. 96–107.

[7] L. DE LATHAUWER, *Signal Processing Based on Multilinear Algebra*, Ph.D. dissertation, Katholieke Universiteit, Leuven, Belgium, 1997.

[8] L. DE LATHAUWER, P. COMON, B. DE MOOR, AND J. VANDEWALLE, *Higher-order power method—Application in independent component analysis*, in Proceedings of the International Symposium on Nonlinear Theory and Its Applications (NOLTA'95), Las Vegas, NV, 1995, pp. 91–96.

[9] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.

[10] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.

[11] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 2nd ed., The Johns Hopkins University Press, Baltimore, MD, 1989.

[12] W. H. GREUB, *Multilinear Algebra*, Springer-Verlag, Berlin, 1967.

[13] H. V. HENDERSON AND S. R. SEARLE, *The vec-permutation matrix, the vec operator and Kronecker products: A review*, Linear and Multilinear Algebra, 9 (1981), pp. 271–288.

[14] A. HYVÄRINEN AND E. OJA, *A fast fixed-point algorithm for independent component analysis*, Neural Comput., 9 (1997), pp. 1483–1492.

[15] A. HYVÄRINEN, *Fast and robust fixed-point algorithms for independent component analysis*, IEEE Trans. Neural Networks, 10 (1999), pp. 626–634.

[16] E. KOFIDIS AND P. A. REGALIA, *Tensor approximation and signal processing applications*, in Structured Matrices in Mathematics, Computer Science and Engineering, Vol. I, Contemp. Math. 280, V. Olshevsky, ed., AMS, Providence, RI, 2001.

[17] T. G. KOLDA, *Orthogonal tensor decompositions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 243–255.

[18] P. M. KROONENBERG, *Singular value decompositions of interactions in three-way contingency tables*, in Multiway Data Analysis, R. Coppi and S. Bolasco eds., Elsevier Science Publishers, North Holland, 1989, pp. 169–184.

[19] M. MBOUP AND P. A. REGALIA, *A gradient search interpretation of the super-exponential algorithm*, IEEE Trans. Inform. Theory, 46 (2000), pp. 2731–2734.

[20] P. MCCULLAGH, *Tensor Methods in Statistics*, Chapman and Hall, New York, 1987.

[21] C. B. PAPADIAS, *Globally convergent blind source separation based on a multiuser kurtosis maximization criterion*, IEEE Trans. Signal Process., 48 (2000), pp. 3508–3519.

[22] P. A. REGALIA AND E. KOFIDIS, *The higher-order power method revisited: Convergence proofs and effective initialization*, in Proceedings of the IEEE International Conference on Acoust., Speech, and Signal Processing (ICASSP-2000), Istanbul, Turkey, 2000.

[23] P. A. REGALIA AND E. KOFIDIS, *A "unimodal" blind equalization criterion*, in Proceedings of the European Signal Processing Conference (EUSIPCO-2000), Tampere, Finland, 2000.

[24] P. A. REGALIA AND M. MBOUP, *Properties of some blind equalization criteria in noisy multiuser environments*, IEEE Trans. Signal Process., 49 (2001), pp. 3112–3122.

[25] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[26] O. SHALVI AND E. WEINSTEIN, *Super-exponential methods for blind deconvolution*, IEEE Trans. Inform. Theory, 39 (1993), pp. 504–519.

[27] N. SIDIROPOULOS AND R. BRO, *PARAFAC techniques for signal separation*, in Signal Processing Advances in Wireless & Mobile Communications—Trends in Single- and Multi-User Systems, G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, eds., Prentice-Hall, Englewood Cliffs, NJ, 2001.

[28] G. W. STEWART, *Introduction to Matrix Computations*, Academic Press, New York, 1973.