NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2007; **14**:503–519
Published online 1 June 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla.537

On the successive supersymmetric rank-1 decomposition of higher-order supersymmetric tensors

Yiju Wang[1,2,*,†] and Liqun Qi[3]

[1] *School of Operations Research and Management Sciences, Qufu Normal University, Rizhao Shandong 276800, China*
[2] *Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong*
[3] *Department of Mathematics, City University of Hong Kong, Kowloon Tong, Kowloon, Hong Kong*

SUMMARY

In this paper, a successive supersymmetric rank-1 decomposition of a real higher-order supersymmetric tensor is considered. To obtain such a decomposition, we design a greedy method based on iteratively computing the best supersymmetric rank-1 approximation of the residual tensors. We further show that a supersymmetric canonical decomposition could be obtained when the method is applied to an orthogonally diagonalizable supersymmetric tensor, and in particular, when the order is 2, this method generates the eigenvalue decomposition for symmetric matrices. Details of the algorithm designed and the numerical results are reported in this paper. Copyright © 2007 John Wiley & Sons, Ltd.

Received 12 September 2005; Revised 6 July 2006; Accepted 16 February 2007

KEY WORDS: higher-order tensors; rank-1 tensors; supersymmetry; decomposition

1. INTRODUCTION

A tensor of order m is an m-way array whose entries are addressed *via* m indices and it arises more and more often in signal and image processing, data analysis, higher-order statistics, as well as independent component analysis [1–5]. In particular, moments and cumulants of multivariate stochastic processes are higher-order tensors [6].

It is well known that higher-order tensors have some analogies with matrices and hence some concepts such as rank and lower-rank decomposition related to matrices have been extended to

*Correspondence to: Yiju Wang, School of Operations Research and Management Sciences, Qufu Normal University, Rizhao Shandong 276800, China.
†E-mail: wyiju@hotmail.com

Contract/grant sponsor: The Research Grant Council of Hong Kong; contract/grant numbers: CityU 501004, CityU 501606

higher-order tensors, e.g. [7–10]. More recently, Qi [11] developed the concepts of H-eigenvalue and symmetric hyperdeterminant of higher-order supersymmetric tensors and discussed their close links which extended some interesting property of symmetric matrix to higher-order case.

The topic of this paper is closely related to the rank of tensors and the following is a brief account on this issue. A tensor is said to be rank-1 if it can be expressed as an outer product of a number of vectors, that is, an mth order rank-1 tensor $\mathscr{T}$ assumes the following form:

$$\mathscr{T} = u^{(1)} \circ u^{(2)} \circ \cdots \circ u^{(m)} \triangleq \prod_{j=1}^{m} u^{(j)}$$

in the sense that

$$\mathscr{T}_{i_1,i_2,\ldots,i_m} = u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(m)}_{i_m}$$

where $u^{(j)} \in R^{n_j}$, $j = 1, 2, \ldots, m$. In particular, if these m vectors are all equal to vector u, then their out product is denoted by u^m and it is called a supersymmetric rank-1 tensor [12].

Rank-1 tensors have many interesting properties and play an important role in tensor analysis [13], and some researchers have payed more attention to finding the best rank-1 approximation to higher-order tensors, e.g. [14].

For a higher-order tensor, its rank is defined as the minimal number of rank-1 tensors that yield $\mathscr{T}$ in a linear combination [9, 10, 14, 15], and this decomposition is called canonical decomposition [16] and is also called PARAFAC decomposition [17], which preserves the uniqueness under mild conditions [9, 10, 18]. For an overview on the recent development on this kind of decomposition, see [15] and papers therein.

In analogy to symmetric matrices, a real higher-order tensor is called supersymmetric if its entries are invariant under any permutation of their indices [19]. It is well known that supersymmetric tensors and homogeneous polynomials are bijectively associated [3, 20]. In the sequel of this paper, we denote the set of mth order n-dimensional supersymmetric tensor by $\mathscr{S}_{m,n}$. It can be verified that the set $\mathscr{S}_{m,n}$ constitutes a linear space of dimension $\binom{n+m-1}{m}$ [3]. Kofidis and Regalia [12], discussed the best supersymmetric rank-1 approximation to higher-order supersymmetric tensors in the least-squares sense. Besides the promising numerical experiments, they also stressed the efficiency of the proposed method in theory.

Similarly, the symmetric rank of a tensor $\mathscr{T} \in \mathscr{S}_{m,n}$ is defined as the minimum number of supersymmetric rank-1 tensors that yield $\mathscr{T}$ in a linear combination [4]. The rank of a tensor $\mathscr{T} \in \mathscr{S}_{m,n}$ may be larger than the dimension n but it is upbounded by a polynomial of n and m [3, 4, 21]. That is, any tensor $\mathscr{T} \in \mathscr{S}_{m,n}$ can be expressed as a combination of supersymmetric rank-1 tensors with finitely many terms, and if the number of terms is minimal, then we call this decomposition symmetric canonical decomposition or symmetric PARAFAC factors decomposition.

Low-rank decomposition of higher-order (not necessarily supersymmetric) tensor finds applications in the analysis of multiway data, as well as high-order statistics and independent component analysis [2, 22]. Up to date, there is no really efficient way to establish a symmetric canonical decomposition for a tensor in $\mathscr{S}_{m,n}$ except for cubic and binary cases [3]. So, in this paper, we consider the decomposition of supersymmetric tensors into the sums of supersymmetric rank-1 tensors. Based on the bijective relation of supersymmetric tensors to homogeneous polynomials, we know that this kind of decomposition is equivalent to expressing an m-degree homogeneous polynomial as the sum of mth powers of linear forms.

DOI: 10.1002/nla

In the next section, we will propose a greedy method to decompose a higher-order supersymmetric tensor into a successive supersymmetric rank-1 tensors. At each step of this method, we need to compute the best supersymmetric rank-1 approximation of the residual tensor, and to obtain such an approximation we will employ the projected gradient method for solving constrained minimization problem [23]. The main contributions of this paper are as follows.

- We show that the greedy method really generates a successive supersymmetric rank-1 tensors for any real supersymmetric tensor.
- We give an example to show that the proposed method could not generate a supersymmetric canonical decomposition, in general, when it is applied to a tensor in $\mathscr{S}_{m,n}$. However, we show that this kind of decomposition could be obtained when the method is applied to an orthogonally diagonalizable supersymmetric tensor.
- For the sequence $\{|\lambda_k|\}$ generated by the method, we borrow an example from [24] to show that this generated sequence may be non-monotone, and we further discuss the conditions under which this generated sequence is monotone.
- We give some preliminary experiments to show the efficiency of our proposed method.

We end this section with some notations used in this paper. Throughout this paper, all variables take values in the real field. Vectors will be denoted by lowercase letters (e.g. u), while matrices will be denoted by uppercase letters (e.g. A), higher-order tensors will be denoted by bold, calligraphic, uppercase letters (e.g. $\mathscr{T}$). Superscripts with brackets of vectors or tensors are used to denote different vectors or tensors. The element of tensor $\mathscr{T}$ with index $(i_1, i_2, \ldots, i_m)$ is denoted by $\mathscr{T}_{i_1,i_2,\ldots,i_m}$. The vector e_i denotes the unit vector in R^n such that the ith element is 1 and others are zero. We use $\|\cdot\|$ to denote the 2-norm of vectors and call a vector to be unit if its 2-norm is 1. For two different tensors, $\mathscr{T}^{(1)}$ and $\mathscr{T}^{(2)}$ of the same dimension and same order, their inner product is defined as follows:

$$\mathscr{T}^{(1)} \cdot \mathscr{T}^{(2)} = \sum_{i_1,i_2,\ldots,i_m} \mathscr{T}^{(1)}_{i_1,i_2,\ldots,i_m} \mathscr{T}^{(2)}_{i_1,i_2,\ldots,i_m}$$

and the Frobenius norm of the tensor $\mathscr{T}$ is defined as follows:

$$\|\mathscr{T}\| = (\mathscr{T} \cdot \mathscr{T})^{1/2} = \left(\sum_{i_1,i_2,\ldots,i_m} \mathscr{T}^2_{i_1,i_2,\ldots,i_m} \right)^{1/2}$$

For simplicity, we omit the dot when we express the inner product of two tensors.

2. DECOMPOSITION METHOD AND ITS EFFICIENCY

First, we formally state the problem considered in [12]: given a tensor $\mathscr{T} \in \mathscr{S}_{m,n}$, determine a scalar λ and a unit vector $u \in R^n$ such that the supersymmetric rank-1 tensor $\bar{\mathscr{T}} = \lambda u^m$ minimizes the function

$$f(\bar{\mathscr{T}}) = \|\mathscr{T} - \bar{\mathscr{T}}\|^2$$

over the manifold of supersymmetric rank-1 tensors. If $\bar{\mathscr{T}} = \lambda u^m$ is the global minimizer of the function, then it is said to be the best supersymmetric rank-1 approximation of the tensor $\mathscr{T}$.

Using the optimality condition, we have the followings (also see Theorem 2 in [12]).

DOI: 10.1002/nla

Lemma 2.1
For unit-norm vector $u \in R^n$ and $\lambda \in R$, the tensor λu^m is the best supersymmetric rank-1 approximation to the tensor $\mathscr{T} \in \mathscr{S}_{m,n}$ if and only if the vector u globally maximizes the function

$$g(v) = |\mathscr{T} v^m| = \left| \sum_{i_1, i_2, \ldots, i_m} \mathscr{T}_{i_1, i_2, \ldots, i_m} v_{i_1} v_{i_2} \ldots v_{i_m} \right|$$

over the unit sphere $\{v \in R^n : \|v\| = 1\}$ or the unit ball $\{v \in R^n : \|v\| \leqslant 1\}$, and

$$\lambda = \mathscr{T} u^m, \quad \mathscr{T} u^{m-1} = \lambda u$$

Remark 2.1
Since tensor $\mathscr{T} \in \mathscr{S}_{m,n}$ corresponds to a symmetric matrix $T \in R^{n \times n}$ when $m = 2$, from Lemma 2.1, we know that if $(\hat{\lambda}, \hat{u})$ is an eigenpair of T such that

$$\hat{\lambda} = \text{argmax } \{|\lambda| : \lambda \text{ is an eigenvalue of } T\}$$

then $\hat{\lambda}(\hat{u})^2$ is the best supersymmetric rank-1 approximation to $\mathscr{T}$ and *vice versa*.

For $m \geqslant 2$, suppose λu^m is the best supersymmetric rank-1 approximation of the tensor $\mathscr{T} \in \mathscr{S}_{m,n}$, from Lemma 2.1, we have

$$\|\mathscr{T} - \lambda u^m\|^2 = \|\mathscr{T}\|^2 - 2\lambda \langle \mathscr{T}, u^m \rangle + \lambda^2 = \|\mathscr{T}\|^2 - \lambda^2 \leqslant \|\mathscr{T}\|^2$$

Obviously, $\mathscr{T}' := \mathscr{T} - \lambda u^m$ is also a supersymmetric tensor of the same order, and its Frobenius norm is strictly less than that of $\mathscr{T}$ if and only if $\lambda \neq 0$. Inspired by this fact, we can design the following algorithm to express a supersymmetric tensor $\mathscr{T}$ as the sum of supersymmetric rank-1 tensors.

Algorithm 2.1

Initial step: Input a supersymmetric tensor $\mathscr{T}$ and $\varepsilon \geqslant 0$, let $k = 1$ and $\mathscr{T}^{(k)} = \mathscr{T}$.
Iterative step: If $\|\mathscr{T}^{(k)}\| \leqslant \varepsilon$, stop. Otherwise, compute $u^{(k)}$ such that

$$u^{(k)} = \arg \max_{\|u\|=1} |\mathscr{T} u^m|$$

Let $\lambda_k := \mathscr{T} \cdot (u^{(k)})^m$, $\mathscr{T}^{(k+1)} := \mathscr{T}^{(k)} - \lambda_k (u^{(k)})^m$ and $k := k + 1$.

For Algorithm 2.1, if $\varepsilon = 0$ and the algorithm terminates within finite number of steps, then we can decompose tensor $\mathscr{T} \in \mathscr{S}_{m,n}$ into a sum of finite supersymmetric rank-1 tensors. Certainly, if $\|\mathscr{T}^{(k)}\| > \varepsilon$ and $\lambda_k = 0$ at a certain step, then the algorithm could not be further executed. However, the following lemma implies that this case never happens. (The following result may be not new, however, we include a compact stand-alone proof below, both for the sake of completeness and because the proof is used in proving our next result.)

Lemma 2.2
For a non-zero tensor $\mathscr{T} \in \mathscr{S}_{m,n}$, it holds that

$$\max\{|\mathscr{T} u^m| : u \in R^n, \|u\| = 1\} \neq 0$$

DOI: 10.1002/nla

Proof
It suffices to show that there exists a unit vector $u \in R^n$ such that $\mathscr{T} u^m \neq 0$ for any given non-zero tensor $\mathscr{T} \in \mathscr{S}_{m,n}$.

Among all the non-zero entries of tensor $\mathscr{T}$, we collect all the ones with the maximum index repetition in the following sense:

(1) the index has t distinct numbers;
(2) any entry of $\mathscr{T}$ whose index has strictly less than t distinct numbers is zero.

Since $\mathscr{T} \neq 0$, this can always be done. Pick up one from these entries. Without loss of generality, we suppose its index has the following t distinct numbers: $1, 2, \ldots, t$. Define the index set

$$S^* := \{i_1 i_2 \ldots i_m : i_j \text{ is taken from } \{1, 2, \ldots, t\} \text{ for } j = 1, 2, \ldots, m\}$$

For the homogeneous multivariate polynomial $\mathscr{T} u^m$ with respect to u, we arrange the non-zero monomials of $\mathscr{T} u^m$ whose index is taken from S^* in the lexicographical order with respect to the exponential of u

$$c_1 u_1^{p_1^1} u_2^{p_2^1} \ldots u_t^{P_t^1}, \quad c_2 u_1^{p_1^2} u_2^{p_2^2} \ldots u_t^{P_t^2}, \ldots, \quad c_s u_1^{p_1^s} u_2^{p_2^s} \ldots u_t^{P_t^s}$$

where $(p_1^j, p_2^j, \ldots, p_t^j)$, $j = 1, 2, \ldots, s$, are the exponentials of u, c_j is the sum of coefficients of the monomials which u has the same exponential, and s denotes the number of these monomials whose coefficient c_j is non-zero.

Note that $(p_1^i, p_2^i, \ldots, p_t^i) \succ (p_1^j, p_2^j, \ldots, p_t^j)$ for $1 \leqslant i < j \leqslant s$ in the sense of lexicographical order, and $(p_1^i, p_2^i, \ldots, p_t^i) \neq (p_1^j, p_2^j, \ldots, p_t^j)$ for $1 \leqslant i \neq j \leqslant s$. Furthermore, $p_1^j + p_2^j + \cdots + p_t^j = m$ for $j = 1, 2, \ldots, s$.

Denote $\kappa = \max\{\max_{1 \leqslant i \leqslant s-1} |c_{i+1}|/|c_i|, 1\}$ and choose $\hat{u} \in R^n$ as follows: $\hat{u}_1 = 1$ and

$$\hat{u}_i = \begin{cases} \dfrac{\hat{u}_{i-1}}{2\kappa} & \text{if } i \in \{2, \ldots, t\} \\ 0 & \text{otherwise} \end{cases}$$

From the choice of $\hat{u}$, we know that

$$|c_{j+1} \hat{u}_1^{p_1^{j+1}} \hat{u}_2^{p_2^{j+1}} \ldots \hat{u}_t^{P_t^{j+1}}| \leqslant \tfrac{1}{2} |c_j \hat{u}_1^{p_1^j} \hat{u}_2^{p_2^j} \ldots \hat{u}_t^{P_t^j}| \quad \text{for } j = 1, 2, \ldots, s-1$$

Thus,

$$\begin{aligned} |\mathscr{T} \hat{u}^m| &= \left| \sum_{j=1}^{s} c_j \hat{u}_1^{p_1^j} \hat{u}_2^{p_2^j} \ldots \hat{u}_t^{P_t^j} \right| \\ &\geqslant |c_1 \hat{u}_1^{p_1^1} \hat{u}_2^{p_2^1} \ldots \hat{u}_t^{P_t^1}| - \sum_{j=2}^{s} |c_j \hat{u}_1^{p_1^j} \hat{u}_2^{p_2^j} \ldots \hat{u}_t^{P_t^j}| \\ &\geqslant \frac{1}{2^{s-1}} |c_1 \hat{u}_1^{p_1^1} \hat{u}_2^{p_2^1} \ldots \hat{u}_t^{P_t^1}| \\ &\geqslant \frac{1}{2^{N-1}} |c_1 \hat{u}_1^{p_1^1} \hat{u}_2^{p_2^1} \ldots \hat{u}_t^{P_t^1}| \\ &> 0 \end{aligned}$$

where $N = n^m$.

DOI: 10.1002/nla

It is ready to verify that $\|\hat{u}\| \leqslant 2$. Define $\bar{u} = \hat{u}/\|\hat{u}\|$, then

$$|\mathscr{T}\bar{u}^m| = \frac{1}{\|\hat{u}\|^m}|\mathscr{T}\hat{u}^m| \geqslant \frac{1}{2^m}|\mathscr{T}\hat{u}^m| > 0$$

The desired result follows. □

This lemma, in conjunction with Lemma 2.1, implies that for any non-zero tensor $\mathscr{T} \in \mathscr{S}_{m,n}$, there exists a unit vector $u \in R^n$ and $\lambda \neq 0$ such that λu^m is the best rank-1 approximation to $\mathscr{T}$ and

$$\|\mathscr{T} - \lambda u^m\|^2 = \|\mathscr{T}\|^2 - \lambda^2 < \|\mathscr{T}\|^2$$

Furthermore, the following conclusion tells us that $|\lambda|$ would not be sufficiently small unless $\|\mathscr{T}\|$ is.

Lemma 2.3
For a sequence $\{\mathscr{T}^{(k)}\} \subset \mathscr{S}_{m,n}$, suppose that $\lambda_k (u^{(k)})^m$ is the best supersymmetric rank-1 approximation of $\mathscr{T}^{(k)}$. Then $|\lambda_k| \to 0$ only if $\|\mathscr{T}^{(k)}\| \to 0$.

Proof
Suppose the assertion does not holds, then there exists a sequence of tensors $\{\mathscr{T}^{(k)}\} \subset \mathscr{S}_{m,n}$ such that $\lim_{k\to\infty} \|\mathscr{T}^{(k)}\| > 0$ and $\lim_{k\to\infty} |\lambda_k| = 0$, where $\lambda_k (u^{(k)})^m$ is the best supersymmetric rank-1 approximation to tensor $\mathscr{T}^{(k)}$. Without loss of generality, we assume that the sequence $\{\mathscr{T}^{(k)}\}$ converges to $\hat{\mathscr{T}}$.

The following arguments are based on the proof of Lemma 2.2.

Denote

$$\mathscr{I} = \{(i_1, i_2, \ldots, i_m) : \hat{\mathscr{T}}_{i_1,i_2,\ldots,i_m} \neq 0\}$$
$$\tau = \min\{|\hat{\mathscr{T}}_{i_1,i_2,\ldots,i_m}| : (i_1, i_2, \ldots, i_m) \in \mathscr{I}\}$$

Since $\hat{\mathscr{T}} \neq 0$, so $\tau > 0$. By hypothesis $\lambda_k \to 0$ as $k \to \infty$, we know that there exists sufficiently large $K > 0$ such that for any $k \geqslant K$, it holds that

$$\|\mathscr{T}^{(k)}\| \leqslant 2\|\hat{\mathscr{T}}\|, \quad |\lambda_k| < \frac{\tau}{2^{N+m+1}} \left(\frac{\tau}{4N\|\hat{\mathscr{T}}\|}\right)^{mn}$$

and

$$|\mathscr{T}^{(k)}_{i_1,i_2,\ldots,i_m}| = \begin{cases} < \dfrac{\tau}{M} & \text{if } (i_1, i_2, \ldots, i_m) \notin \mathscr{I} \\ > \dfrac{\tau}{2} & \text{otherwise} \end{cases}$$

where $M = N2^{N+1}((4N\|\hat{\mathscr{T}}\|)/\tau)^{mn}$ and $N = n^m$.

Let $\hat{\mathscr{T}}^{(k)} \in \mathscr{S}_{m,n}$ be such that

$$\hat{\mathscr{T}}^{(k)}_{i_1,i_2,\ldots,i_m} = \begin{cases} 0 & \text{if } (i_1, i_2, \ldots, i_m) \notin \mathscr{I} \\ \mathscr{T}^{(k)}_{i_1,i_2,\ldots,i_m} & \text{otherwise} \end{cases}$$

DOI: 10.1002/nla

We take $\hat{\mathscr{T}}^{(k)}$ as the tensor considered in Lemma 2.2, and adopt the convention used in the proof of Lemma 2.2, then

$$\kappa = \max\left\{\max_{1\leqslant i\leqslant s-1}\frac{|c_{i+1}|}{|c_i|}, 1\right\}\leqslant\frac{N\|\hat{\mathscr{T}}\|}{\tau/2}=\frac{2N\|\hat{\mathscr{T}}\|}{\tau}$$

It is easy to verify that $1\geqslant\hat{u}_i\geqslant(1/2\kappa)^n$ for $i=1,2,\ldots,t$. Thus,

$$|\hat{\mathscr{T}}^{(k)}\hat{u}^m|\geqslant\frac{1}{2^{s-1}}|c_1\hat{u}_1^{p_1^1}\hat{u}_2^{p_2^1}\ldots\hat{u}_t^{P_t^1}|\geqslant\frac{\tau}{2^N}\left(\frac{1}{2\kappa}\right)^{mn}$$

and hence

$$\begin{aligned}
|\mathscr{T}^{(k)}\hat{u}^m| &= \left|\hat{\mathscr{T}}^{(k)}\hat{u}^m+\sum_{(i_i,i_2,\ldots,i_m)\notin\mathscr{I}}\mathscr{T}^{(k)}_{i_i,i_2,\ldots,i_m}\hat{u}_1^{p_1^1}\hat{u}_2^{p_2^1}\ldots\hat{u}_t^{P_t^1}\right|\\
&\geqslant|\hat{\mathscr{T}}^{(k)}\hat{u}^m|-\sum_{(i_i,i_2,\ldots,i_m)\notin\mathscr{I}}|\mathscr{T}^{(k)}_{i_i,i_2,\ldots,i_m}\hat{u}_1^{p_1^1}\hat{u}_2^{p_2^1}\ldots\hat{u}_t^{P_t^1}|\\
&\geqslant\frac{1}{2^{s-1}}|c_1\hat{u}_1^{p_1^1}\hat{u}_2^{p_2^1}\ldots\hat{u}_t^{P_t^1}|-\sum_{(i_i,i_2,\ldots,i_m)\notin\mathscr{I}}|\mathscr{T}^{(k)}_{i_i,i_2,\ldots,i_m}|\\
&\geqslant\frac{\tau}{2^N}\left(\frac{1}{2\kappa}\right)^{mn}-N\frac{\tau}{M}\\
&\geqslant\frac{\tau}{2^N}\left(\frac{\tau}{4N\|\hat{\mathscr{T}}\|}\right)^{mn}-N\frac{\tau}{M}\\
&\geqslant\frac{\tau}{2^{N+1}}\left(\frac{\tau}{4N\|\hat{\mathscr{T}}\|}\right)^{mn}
\end{aligned}$$

where the last inequality follows from the definition of M.

Since $\|\hat{u}\|\leqslant 2$, if we let $\bar{u}=\hat{u}/\|\hat{u}\|$, then for $k\geqslant K$

$$\max_{\substack{u\in R^n\\ \|u\|=1}}|\mathscr{T}^{(k)}u^m|\geqslant|\mathscr{T}^{(k)}\bar{u}^m|\geqslant\frac{\tau}{2^{N+m+1}}\left(\frac{\tau}{4N\|\hat{\mathscr{T}}\|}\right)^{mn}$$

Recalling that

$$|\lambda_k|<\frac{\tau}{2^{N+m+1}}\left(\frac{\tau}{4N\|\hat{\mathscr{T}}\|}\right)^{mn}$$

for sufficiently large k, from Lemma 2.1, we obtain a contradiction and we are done. □

Based on the previous lemmas, we can establish our first main result in this section.

DOI: 10.1002/nla

Theorem 2.1
Suppose $\varepsilon=0$. When Algorithm 2.1 is applied to a tensor $\mathscr{T}\in\mathscr{S}_{m,n}$, if it generates an infinite sequence of supersymmetric rank-1 tensors $\{\lambda_k(u^{(k)})^m\}$, then

$$\mathscr{T}=\sum_{k=1}^{\infty}\lambda_k(u^{(k)})^m,\quad \|\mathscr{T}\|^2=\sum_{k=1}^{\infty}\lambda_k^2$$

Proof
From Lemma 2.2, we know that if $\|\mathscr{T}^{(k)}\|\neq 0$, Algorithm 2.1 generates a unit vector $u^{(k)}\in R^n$ and a non-zero scalar λ_k such that $\lambda_k(u^{(k)})^m$ is the best rank-1 approximation of $\mathscr{T}^{(k)}$. Using Lemma 2.1, one has

$$\begin{aligned}\left\|\mathscr{T}-\sum_{i=1}^{k}\lambda_i(u^{(i)})^m\right\|^2&=\left\|\mathscr{T}-\sum_{i=1}^{k-1}\lambda_i(u^{(i)})^m\right\|^2-\lambda_k^2\\&=\|\mathscr{T}\|^2-\sum_{i=1}^{k}\lambda_i^2\end{aligned}$$

Letting $k\to\infty$, we get

$$\left\|\mathscr{T}-\sum_{i=1}^{\infty}\lambda_i(u^{(i)})^m\right\|^2=\|\mathscr{T}\|^2-\sum_{i=1}^{\infty}\lambda_i^2$$

Using Lemma 2.3, we know that

$$\mathscr{T}=\sum_{i=1}^{\infty}\lambda_i(u^{(i)})^m \qquad\square$$

This conclusion shows that we can get a successive supersymmetric rank-1 decomposition of tensors in $\mathscr{S}_{m,n}$ by Algorithm 2.1. Certainly, it would be very excellent if the algorithm could generate the supersymmetric canonical decomposition when it is applied to a tensor in $\mathscr{S}_{m,n}$. However, this is not the case in general as seen from the following example.

Let $u^{(1)},\ldots,u^{(s)}$ be s ($s\geqslant 2$) linearly independent but non-orthogonal vectors in R^n, and $\mathscr{T}$ be a third-order supersymmetric tensor

$$\mathscr{T}=\sum_{j=1}^{s}(u^{(j)})^3$$

which is known to be the unique supersymmetric canonical decomposition of tensor $\mathscr{T}$ [9]. Let $u^{(1)},u^{(2)}$ be normal vectors in R^2 such that $\langle u^{(1)},u^{(2)}\rangle=\frac{1}{2}$. In this case, $\mathscr{T}=(u^{(1)})^3+(u^{(2)})^3$.

For any unit vector $v\in R^2$, there exist numbers α_1 and α_2 such that $v=\alpha_1u^{(1)}+\alpha_2u^{(2)}$. Now, we consider the tensor product of $\mathscr{T}$ and vector v^3

$$\begin{aligned}\mathscr{T}v^3&=((u^{(1)})^3+(u^{(2)})^3)(\alpha_1u^{(1)}+\alpha_2u^{(2)})^3\\&=(\alpha_1+\tfrac{1}{2}\alpha_2)^3+(\alpha_2+\tfrac{1}{2}\alpha_1)^3\\&=\alpha_1^3+\tfrac{3}{2}\alpha_1^2\alpha_2+\tfrac{3}{4}\alpha_1\alpha_2^2+\tfrac{1}{8}\alpha_2^3+\alpha_2^3+\tfrac{3}{2}\alpha_2^2\alpha_1+\tfrac{3}{4}\alpha_2\alpha_1^2+\tfrac{1}{8}\alpha_1^3\end{aligned}$$

DOI: 10.1002/nla

$$= \tfrac{9}{8}(\alpha_1^3 + \alpha_2^3) + \tfrac{9}{4}\alpha_1\alpha_2(\alpha_1 + \alpha_2)$$
$$= \tfrac{9}{8}(\alpha_1 + \alpha_2)(\alpha_1^2 + \alpha_2^2 - \alpha_1\alpha_2) + \tfrac{9}{4}\alpha_1\alpha_2(\alpha_1 + \alpha_2)$$
$$= \tfrac{9}{8}(\alpha_1 + \alpha_2)$$

where the last equality uses the following equality on α_1 and α_2, deduced from the fact that v is a unit vector:

$$\alpha_1^2 + \alpha_2^2 + 2\alpha_1\alpha_2\langle u^{(1)}, u^{(2)}\rangle = 1$$

i.e.

$$\alpha_1^2 + \alpha_2^2 + \alpha_1\alpha_2 = 1$$

Obviously, $\alpha_1 + \alpha_2 > 1$ for any positive numbers α_1 and α_2 satisfying the equality above, and consequently,

$$\max_{v \in R^2, \|v\|=1} |\mathscr{T} v^3| = \sum_{j=1}^{2} |(u^{(j)})^3 v^3| > \sum_{j=1}^{2} (u^{(j)})^3 (u^{(i)})^3, \quad i = 1, 2$$

Recalling Lemma 2.1, we deduce that neither of $(u^{(i)})^3$, $i = 1, 2$, is the best supersymmetric rank-1 approximation to tensor $\mathscr{T}$.

This says that Algorithm 2.1 may not terminate within finite steps or generate the supersymmetric canonical decomposition for tensors in $\mathscr{S}_{m,n}$ in general. In the following, we consider a special case in which Algorithm 2.1 generates the supersymmetric canonical decomposition. To this end, the following definition is needed.

For a tensor $\mathscr{T} \in \mathscr{S}_{m,n}$, the elements $\mathscr{T}_{i,i,\ldots,i}$, for $i = 1, 2, \ldots, n$, are said to be the diagonal elements of $\mathscr{T}$, and the other elements in $\mathscr{T}$ are said to be off-diagonal elements. A supersymmetric higher-order tensor $\mathscr{T}$ is said to be diagonal if all the off-diagonal elements in $\mathscr{T}$ are zero, and a supersymmetric higher-order tensor $\mathscr{T}$ is said to be orthogonally diagonalizable if there exists an orthogonal matrix $A \in R^{n\times n}$ such that $A^m\mathscr{T}$ is diagonal, where $A^m\mathscr{T}$ is defined as

$$(A^m\mathscr{T})_{i_1,i_2,\ldots,i_m} = \sum_{j_1,j_2,\ldots,j_m} \mathscr{T}_{j_1,j_2,\ldots,j_m} A_{i_1,j_1} A_{i_2,j_2} \ldots A_{i_m,j_m} \quad \text{for } i_1, i_2, \ldots, i_m = 1, 2, \ldots, n$$

This kind of transformation can be taken as an orthogonal transformation to tensors as seen from the following result.

Proposition 2.1
For $\mathscr{T} \in \mathscr{S}_{m,n}$, $v \in R^n$ and orthogonal matrix $A \in R^{n\times n}$, the followings hold.

(1) $\|A^m\mathscr{T}\| = \|\mathscr{T}\|$,
(2) $(A^\top)^m(A^m\mathscr{T}) = \mathscr{T}$,
(3) $A^m v^m = (Av)^m$,
(4) $(A^m\mathscr{T})v^m = \mathscr{T}(A^\top v)^m$.

Proof
Here, we only prove the first statement, as the others can similarly be proved.

Since

$$
\begin{aligned}
(A^m \mathscr{T}) \cdot (A^m \mathscr{T}) &= \sum_{i_1,\ldots,i_m} \left(\sum_{j_1,\ldots,j_m} \mathscr{T}_{j_1,\ldots,j_m} A_{i_1,j_1} \ldots A_{i_m,j_m} \sum_{j'_1,\ldots,j'_m} \mathscr{T}_{j'_1,\ldots,j'_m} A_{i_1,j'_1} \ldots A_{i_m,j'_m} \right) \\
&= \sum_{j_1,\ldots,j_m} \sum_{j'_1,\ldots,j'_m} \mathscr{T}_{j_1,\ldots,j_m} \mathscr{T}_{j'_1,\ldots,j'_m} \left(\sum_{i_1,\ldots,i_m} A_{i_1,j_1} \ldots A_{i_m,j_m} A_{i_1,j'_1} \ldots A_{i_m,j'_m} \right) \\
&= \sum_{j_1,\ldots,j_m} \sum_{j'_1,\ldots,j'_m} \mathscr{T}_{j_1,\ldots,j_m} \mathscr{T}_{j'_1,\ldots,j'_m} \left(\sum_{i_1} A_{i_1,j_1} A_{i_1,j'_1} \right) \cdots \left(\sum_{i_m} A_{i_m,j_m} A_{i_m,j'_m} \right) \\
&= \sum_{j_1,\ldots,j_m} \mathscr{T}_{j_1,\ldots,j_m} \mathscr{T}_{j_1,\ldots,j_m} = \|\mathscr{T}\|^2
\end{aligned}
$$

where the next to the last equality follows from the orthogonality of A, we conclude that (1) holds. □

To discuss the finite step termination of Algorithm 2.1 when it is applied to an orthogonally diagonalizable tensor in $\mathscr{S}_{m,n}$, we need the following conclusion.

Lemma 2.4
For $\mathscr{T} \in \mathscr{S}_{m,n}$ and orthogonal matrix $A \in R^{n\times n}$, if λu^m is the best supersymmetric rank-1 approximation to the tensor $\mathscr{T}$, then $\lambda(Au)^m$ is the best supersymmetric rank-1 approximation to the tensor $A^m \mathscr{T}$.

Proof
First, we note that any global maximizer $u \in R^n$ of the optimization problem

$$\max\{|\mathscr{T} v^m| \mid \|v\| = 1, v \in R^n\}$$

corresponds to the the best supersymmetric rank-1 approximation to the tensor $\mathscr{T}$ with $\lambda = \mathscr{T} u^m$ from Lemma 2.1. On the other hand, from (4) in Proposition 2.1, we know that any global maximizer $u \in R^n$ of the optimization problem stated above corresponds to the global maximizer Au of the optimization problem

$$\max\{|(A^m \mathscr{T}) v^m| \mid \|v\| = 1, v \in R^n\}$$

Hence, $\lambda(Au)^m$ is the best supersymmetric rank-1 approximation to the tensor $A^m \mathscr{T}$. □

Theorem 2.2
Suppose $\varepsilon = 0$. If Algorithm 2.1 is applied to an orthogonally diagonalizable tensor $\mathscr{T} \in \mathscr{S}_{m,n}$, then it terminates at most n steps.

Proof
For orthogonally diagonalizable tensor $\mathscr{T} \in \mathscr{S}_{m,n}$, there exists orthogonal matrix A such that $A^m \mathscr{T}$ is diagonal tensor in $\mathscr{S}_{m,n}$ with non-zero diagonal element $\mu_1, \mu_2, \ldots, \mu_r$, and $r \leqslant n$. Certainly,

$$A^m \mathscr{T} = \sum_{i=r}^{r} \mu_i e_i^m$$

From Proposition 2.1, one has

$$\mathscr{T}=\sum_{i=1}^{r}\mu_i(A^\top e_i)^m$$

Without loss of generality, we assume that $\{|\mu_i|\}$ is non-increasing. From Lemma 2.1, it is easy to see that $\mu_{i+1}e_{i+1}^m$ is the best supersymmetric rank-1 approximation to the tensor $A^m\mathscr{T}-\sum_{j=1}^{i}\mu_j e_j^m$, for $i=0,1,2,\ldots,r-1$. From Lemma 2.4, we know that $\mu_{i+1}(A^\top e_{i+1})^m$ is the best supersymmetric rank-1 approximation to the tensor $\mathscr{T}-\sum_{j=1}^{i}\mu_j(A^\top e_j)^m$. The desired result follows from the fact that Algorithm 2.1 generates the best supersymmetric rank-1 approximation to the tensor $\mathscr{T}-\sum_{i=1}^{r}\mu_i(A^\top e_i)^m$ at each step. □

From the proof above, we know that for any orthogonally diagonalizable tensor $\mathscr{T}\in\mathscr{S}_{m,n}$, Algorithm 2.1 generates a non-increasing finite sequence $\{|\lambda_k|\}$. Obviously, any second-order supersymmetric tensor, i.e. symmetric matrix is always orthogonally diagonalizable, so if Algorithm 2.1 is applied to a second-order supersymmetric tensor, then it generates the eigenvalue decomposition and the generated sequence $\{|\lambda_k|\}$ is non-increasing in this case. Now, an interesting question is whether the monotonicity of the generated sequence $\{|\lambda_k|\}$ can be kept in general. In the rest of this section, we will focus on this issue.

First, consider three-order three-dimensional supersymmetric tensor

$$\mathscr{T}=x\circ y\circ z+y\circ z\circ x+z\circ x\circ y$$

where x,y,z are orthonormal vectors in R^3. As shown in [24], the best supersymmetric rank-1 approximation of this tensor is λu^3 with $u=\alpha_1x+\alpha_2y+\alpha_3z$ such that $\alpha_1^2=\alpha_2^2=\alpha_3^2=\frac{1}{3}$ and $|\lambda|=3|\alpha_1\alpha_2\alpha_3|=\sqrt{3}/3$.

Obviously, there are many choices of α_i's, and without loss of generality, we let $\alpha_1=\alpha_2=\alpha_3=\sqrt{3}/3$, then $\lambda=\sqrt{3}/3$ and λu^3 is the best supersymmetric rank-1 approximation to $\mathscr{T}$.

Now, consider the best supersymmetric rank-1 approximation to the residue tensor $\bar{\mathscr{T}}=\mathscr{T}-\lambda u^3$. To this end, we let $v=\beta_1x+\beta_2y+\beta_3z$ with $\beta_1^2+\beta_2^2+\beta_3^2=1$. Then

$$\begin{aligned}\bar{\mathscr{T}}v^3&=(x\circ y\circ z+y\circ z\circ x+z\circ x\circ y-\tfrac{1}{9}(x+y+z)^3)v^3\\&=3\beta_1\beta_2\beta_3-\frac{1}{9}(\beta_1+\beta_2+\beta_3)^3\end{aligned}$$

and if we let $\beta_1=-\beta_2=-\beta_3=\sqrt{3}/3$, then

$$\bar{\mathscr{T}}v^3=\frac{\sqrt{3}}{3}+\frac{1}{9}\left(\frac{\sqrt{3}}{3}\right)^3$$

Combining this with Lemma 2.1, we conclude that $|\lambda_1|<|\lambda_2|$ when we apply Algorithm 2.1 to tensor $\mathscr{T}$.

The following result shows that the sequence $\{|\lambda_k|\}$ is monotone under certain conditions.

Proposition 2.2
For the sequence $\{\lambda_k\}$ generated by applying Algorithm 2.1 to a non-zero tensor $\mathscr{T} \in \mathscr{S}_{m,n}$, it holds that

$$|\lambda_{k+1}| < 2|\lambda_k| \quad \text{for all } k \text{ such that } \lambda_k \neq 0$$

Furthermore, $|\lambda_k| \geqslant |\lambda_{k+1}|$ in one of the following cases:

(1) if m is an even number, λ_k and λ_{k+1} have the same sign;
(2) the vectors corresponding to λ_k and λ_{k+1} for the best rank-1 approximation are orthogonal.

To prove this proposition, we need the following lemma.

Lemma 2.5
Suppose u and v are two unit vectors in R^n. Then $|u^m \cdot v^m| \leqslant 1$ and equality holds iff $u = \pm v$.

Proof
Suppose that $u = (u_1, u_2, \ldots, u_n)^\top$ and $v = (v_1, v_2, \ldots, v_n)^\top$. Using Cauchy–Schwarz inequality, we obtain that

$$\begin{aligned}
|u^m \cdot v^m| &= \left| \sum_{i_1, i_2, \ldots, i_m}^{n} u_{i_1} v_{i_1} u_{i_2} v_{i_2}, \ldots, u_{i_m} v_{i_m} \right| \\
&= \left| \sum_{i_1=1}^{n} u_{i_1} v_{i_1} \sum_{i_2=1}^{n} u_{i_2} v_{i_2}, \ldots, \sum_{i_m=1}^{n} u_{i_m} v_{i_m} \right| \\
&= |(u^\top v)^m| \\
&\leqslant \|u\|^m \cdot \|v\|^m \\
&= 1
\end{aligned}$$

Certainly, the equality holds if and only if $u = \pm v$. □

Proof of Proposition 2.2
Suppose $\lambda_1 (u^{(1)})^m$ and $\lambda_2 (u^{(2)})^m$ are the best supersymmetric rank-1 approximation to $\mathscr{T}$ and $\mathscr{T} - \lambda_1 (u^{(1)})^m$, respectively. To prove the first statement of the assertion, we only need consider the following two cases for vectors $u^{(1)}$ and $u^{(2)}$:

(a) If $u^{(1)}$ and $u^{(2)}$ are linearly dependent, i.e. $u^{(1)} = \pm u^{(2)}$, then

$$\begin{aligned}
\|\mathscr{T}\|^2 - \lambda_1^2 - \lambda_2^2 &= \|\mathscr{T} - \lambda_1 (u^{(1)})^m - \lambda_2 (u^{(2)})^m\|^2 \\
&= \begin{cases} \|\mathscr{T} - (\lambda_1 + \lambda_2)(u^{(1)})^m\|^2 & \text{if } u^{(1)} = u^{(2)} \text{ or } m \text{ is even} \\ \|\mathscr{T} - (\lambda_1 - \lambda_2)(u^{(1)})^m\|^2 & \text{if } u^{(1)} = -u^{(2)} \text{ and } m \text{ is odd} \end{cases} \\
&= \|\mathscr{T}\|^2 - \lambda_1^2 + \lambda_2^2
\end{aligned}$$

which implies that $\lambda_2 = 0$, the desired result follows from the fact that $\lambda_1 \neq 0$.

DOI: 10.1002/nla

(b) If $u^{(1)}$ and $u^{(2)}$ are linearly independent. From Lemma 2.1, we know that

$$|\lambda_1| = |\mathscr{T}(u^{(1)})^m| \geqslant |\mathscr{T}(u^{(2)})^m|, \quad \lambda_2 = (\mathscr{T} - \lambda_1(u^{(1)})^m)(u^{(2)})^m$$

Using Lemma 2.5, one has

$$|\lambda_2| < 2|\lambda_1|$$

Combining these two cases and using the induction method, we obtain the first statement of the conclusion.

For the second statement, since

$$\|\mathscr{T} - \lambda_2(u^{(2)})^m\|^2 \geqslant \|\mathscr{T} - \lambda_1(u^{(1)})^m\|^2$$

it follows that

$$\begin{aligned}
\|\mathscr{T}\|^2 - \lambda_1^2 - \lambda_2^2 &= \|\mathscr{T} - \lambda_1(u^{(1)})^m - \lambda_2(u^{(2)})^m\|^2 \\
&= \|\mathscr{T} - \lambda_2(u^{(2)})^m\|^2 - 2\langle \mathscr{T} - \lambda_2(u^{(2)})^m, \lambda_1(u^{(1)})^m\rangle + \lambda_1^2 \\
&\geqslant \|\mathscr{T} - \lambda_1(u^{(1)})^m\|^2 - \lambda_1^2 + 2\lambda_1\lambda_2(u^{(1)})^m(u^{(2)})^m \\
&= \|\mathscr{T}\|^2 - 2\lambda_1^2 + 2\lambda_1\lambda_2(u^{(1)})^m(u^{(2)})^m
\end{aligned}$$

i.e.

$$\lambda_1^2 - \lambda_2^2 \geqslant 2\lambda_1\lambda_2(u^{(1)})^m(u^{(2)})^m$$

The desired result follows from the facts that $\lambda_1\lambda_2(u^{(1)})^m(u^{(2)})^m \geqslant 0$ for cases (1) and (2). □

3. NUMERICAL EXPERIMENTS

When we apply Algorithm 2.1 to tensors in $\mathscr{S}_{m,n}$ to get its successive supersymmetric rank-1 decomposition, the best supersymmetric rank-1 approximation to a supersymmetric tensor is needed at each step, in other words, we should find the global minimizer of the following constrained optimization problem:

$$\begin{aligned}
\min \quad & f(u) = -|\mathscr{T}u^m| \\
\text{s.t.} \quad & \|u\|^2 \leqslant 1
\end{aligned}$$

Kofidis and Regalia [25] proposed a supersymmetric higher-order power method (S-HOPM) to accomplish this job. In the following, we will employ the projected gradient method to do this job due to that the projection from R^n onto the unit ball can easily be obtained [23].

To shed light on the projected gradient method for solving the constrained optimization problem, we give the definition of projection operator below.

Denote $\Omega = \{v \in R^n : \|v\| \leqslant 1\}$. The projection from R^n onto the set Ω is defined as

$$P(u) = \arg\min\{\|u - v\| \mid v \in \Omega\} \quad \forall u \in R^n$$

Obviously, for $u \in R^n$,

$$P(u) = \begin{cases} u & \text{if } \|u\| \leqslant 1 \\ \dfrac{u}{\|u\|} & \text{otherwise} \end{cases}$$

Hence $u^{(k)}$ in Algorithm 2.1 can be obtained by applying the following algorithm to the tensor:

$$\mathscr{T}^{(k)} = \mathscr{T} - \sum_{i=1}^{k-1} \lambda_i (u^{(i)})^m$$

Algorithm 3.1

Step 1: Choose $\beta > 0$, $\sigma, \gamma \in (0, 1)$, $\varepsilon \geqslant 0$, and $v^{(0)} \in R^n$ such that $\|v^{(0)}\| = 1$ and $\mathscr{T}^{(k)}(v^{(0)})^m \neq 0$. If $\mathscr{T}^{(k)}(v^{(0)})^m > 0$, then let $f(v) = -\mathscr{T}^{(k)} v^m$, otherwise, $f(v) = \mathscr{T}^{(k)} v^m$. Let $t = 0$.

Step 2: Compute vector $v^{(t)}(1) := P(v^{(t)} - \nabla f(v^{(t)}))$. If $\|v^{(t)} - v^{(t)}(1)\| \leqslant \varepsilon$, stop and output $u^{(k)} = v^{(t)}$.

Step 3: Compute the stepsize $\alpha_t = \beta \gamma^{m_t}$, where m_t is the smallest non-negative integer m such that

$$f(P(v^{(t)} - \beta \gamma^m \nabla f(v^{(t)}))) \leqslant f(v^{(t)}) + \sigma \langle \nabla f(v^{(t)}), P(v^{(t)} - \beta \gamma^m \nabla f(v^{(t)})) - v^{(t)} \rangle$$

Step 4: Let $v^{(t+1)} = P(v^{(t)} - \alpha_t \nabla f(v^{(t)}))$, $t := t + 1$ and go to Step 2.

Remark 3.1
The algorithm consists of an initialization stage (Step 1) in which we set the algorithm parameters, of verification stage (Step 2), and of a recursive stage which consists of two steps: Steps 3 and 4. Step 2 is used to verify if the current iterative point is an approximate stationary point of the problem which sometimes can be taken as a solution point [23], Steps 3 and 4 generate the next iterate *via* the Armijo stepsize rule which guarantees that the generated sequence of objective value is monotonically decreasing.

Remark 3.2
At each step of Algorithm 2.1, to find the best supersymmetric rank-1 approximation to a tensor $\mathscr{T}^{(k)} \in \mathscr{S}_{m,n}$, i.e. to find the global minimizer of the constrained optimization problem

$$\begin{aligned} \min \quad & f(u) = -|\mathscr{T}^{(k)} u^m| \\ \text{s.t.} \quad & \|u\|^2 \leqslant 1 \end{aligned}$$

we would combine Algorithm 3.1 with the uniform grid method of higher-order accuracy.

In our following numerical experiments, we will alternatively use Algorithms 2.1 and 3.1 to test the efficiency of our method. The following first two examples are taken from [12]. Obviously, computing the best supersymmetric rank-1 approximation to a tensor $\mathscr{T} \in \mathscr{S}_{m,n}$, i.e. computing the global minimizer of the constrained optimization problem, is much more expensive, and may be impossible sometimes. In fact, a local minimizer of the problem suffices. Suppose $u \in R^n$ is a local minimizer of the problem. Certainly, $\|u\| = 1$. If we take $\lambda = \mathscr{T} u^m$, then

$$\|\mathscr{T} - \lambda u^m\|^2 = \|\mathscr{T}\|^2 - \lambda^2$$

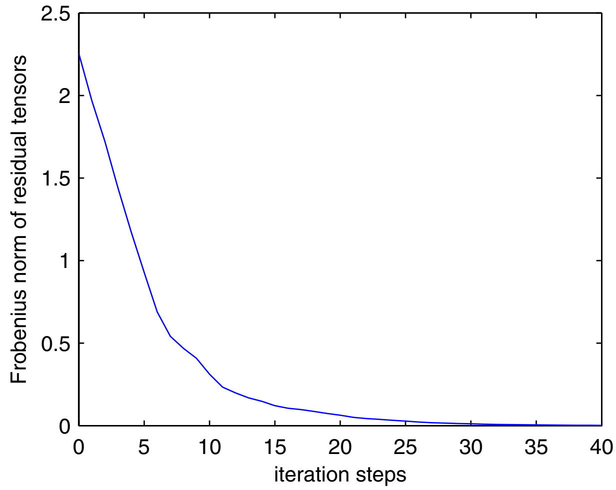

Figure 1. Numerical result of Example 3.1.

Thus, if $\mathscr{T}u^m \neq 0$, then $\|\mathscr{T} - \lambda u^m\| < \|\mathscr{T}\|$, which shows that the sequence $\{\|\mathscr{T}^{(k)}\|\}$ strictly decreases even if $u^{(k)} \in R^n$ is a local and non-global minimizer. Based on this fact, in the following numerical experiments, we would use the uniform grid method with lower-order accuracy to choose the initial point $v^{(0)}$ in Algorithm 3.1. The numerical results of the following examples can be seen from Figures 1–3.

Example 3.1
Consider the tensor $\mathscr{T} \in \mathscr{S}_{4,3}$ with entries:

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

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

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

Example 3.2
Consider the normalized supersymmetric tensor $\mathscr{T} \in \mathscr{T}_{4,3}$ in Example 2 in [12].

Example 3.3
Consider the supersymmetric tensor $\mathscr{T} \in \mathscr{T}_{4,3}$ whose entries are generated from $[-1, 1]$ randomly:

$\mathscr{T}_{1111} = 0.2137$, $\mathscr{T}_{1112} = -0.0280$, $\mathscr{T}_{1113} = 0.7826$, $\mathscr{T}_{1122} = 0.5242$, $\mathscr{T}_{1123} = -0.0871$

$\mathscr{T}_{1133} = -0.9630$, $\mathscr{T}_{1222} = 0.6428$, $\mathscr{T}_{1223} = -0.1106$, $\mathscr{T}_{1233} = 0.2309$, $\mathscr{T}_{1333} = 0.8436$

$\mathscr{T}_{2222} = 0.4764$, $\mathscr{T}_{2223} = -0.6475$, $\mathscr{T}_{2233} = -0.1886$, $\mathscr{T}_{2333} = 0.8709$, $\mathscr{T}_{3333} = 0.8338$

DOI: 10.1002/nla

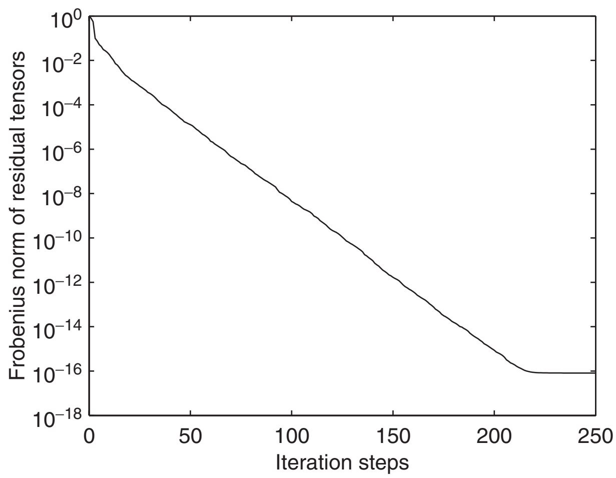

Figure 2. Numerical result of Example 3.2.

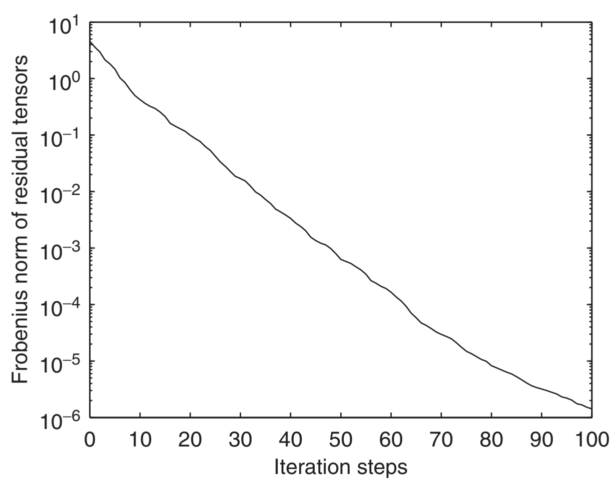

Figure 3. Numerical result of Example 3.3.

In the end of this section, we give a theoretical analysis on the S-HOPM method and Algorithm 3.1 for computing the best rank-1 approximation of a tensor in $\mathscr{S}_{m,n}$. It is well known that these two methods both converges (at most) linearly, and they all need lower capacity during the computing. The reason why we adopt the projected gradient method here is that this method always converges to a stationary point of the concerned problem and it can be easily combined with the uniform grid method to compute an approximate global optimal solution. As for the computation time of our method, it mainly depends on the accuracy order we choose when we use the uniform grid method to select the initial point $v^{(0)}$ in Algorithm 3.1.

ACKNOWLEDGEMENTS

The authors wish to give their sincere thanks to the anonymous referees for their valuable suggestions and helpful comments which improved the presentation of the paper.

REFERENCES

1. Cardoso JF. High-order contrasts for independent component analysis. *Neural Computation* 1999; **11**:157–192.
2. Comon P. Independent component analysis, a new concept? *Signal Processing* 1994; **36**:287–314.
3. Comon P, Mourrain B. Decomposition of quantics in sums of powers of linear forms. *Signal Processing* 1996; **53**:96–107.
4. Comon P. Tensor decompositions In *Mathematics in Signal Processing V*, McWhirter JG, Proudler IK (eds). Oxford University Press: Oxford, U.K., 2001.
5. Nikias CL, Mendel JM. Signal processing with higher-order spectra. *IEEE Signal Processing Magazine* 1993; **10**:10–37.
6. Mccullagh P. *Tensor Methods in Statistics*. Monographs on Statistics and Applied Probability. Chapman & Hall: London, 1987.
7. De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**:1253–1278.
8. De Lathauwer L, De Moor B. From matrix to tensor: multilinear algebra and signal processing In *Mathematics in Signal Processing IV*, McWhirter J (ed.). Oxford University Press: Oxford, U.K., 1998; 1–15.
9. Kruskal JB. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications* 1977; **18**:95–138.
10. Sidiropoulos ND, Bro R. On the uniqueness of multilinear decomposition of N-way arrays. *Journal of Chemometrics* 2000; **14**:229–239.
11. Qi LQ. Eigenvalues of an even-order real supersymmetric tensor. *Journal of Symbolic Computation* 2005; **40**:1302–1324.
12. Kofidis E, Regalia PA. On the best rank-1 approximation of higher-order supersymmetric tensors. *SIAM Journal on Matrix Analysis and Applications* 2002; **23**:863–884.
13. Zhang T, Golub GH. Rank-1 approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications* 2001; **23**:534–550.
14. De Lathauwer L, De Moor B, Vandewalle J. On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensor. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**:1324–1342.
15. De Lathauwer L, De Moor B, Vandewalle J. Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition. *SIAM Journal on Matrix Analysis and Applications* 2004; **26**: 295–327.
16. Carrol JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckardt-Young' decomposition. *Psychometrika* 1970; **35**:283–319.
17. Harshman RA. Foundations of PARAFAC procedure: models and conditions for an 'explanatory' multi-mode factor analysis. *UCLA Working Papers in Phonetics* 1970; **16**:1–84.
18. Harshman RA. Determination and proof of minimum uniqueness conditions for parafac. *UCLA Working Papers in Phonetics* 1972; **22**:111–117.
19. Brini A, Huang RQ, Teolis AGB. The umbral symbolic method for supersymmetric tensors. *Advances in Mathematics* 1992; **96**:123–193.
20. Ehrenborg R, Rota GC. Apolarity and canonical forms for homogeneous polynomials. *European Journal of Combinatorics* 1993; **14**:157–181.
21. Reznick B. Sums of even powers of real linear forms. *Memoirs of the American Mathematical Society* 1992; **96**(463).
22. De Lathauwer L, De Moor B, Vandewalle J. An introduction to independent component analysis. *Journal of Chemometrics* 2000; **14**:123–149.
23. Calamai PH, Moré JJ. Projected gradient methods for linearly constrained problems. *Mathematical Programming* 1987; **39**:93–116.
24. De Lathauwer L. First-order perturbation analysis of the best rank-(R_1, R_2, R_3) approximation in multilinear algebra. *Journal of Chemometrics* 2004; **18**(1):2–11.
25. Kofidis E, Regalia PA. Tensor approximation and signal processing applications. In *Structured Matrices in Mathematics*, *Computer Science and Engineering I*, vol. 280, Olshevsky V (ed.). AMS: Providence, RI, 2001; 103–133.

DOI: 10.1002/nla