

Z-eigenvalue methods for a global polynomial optimization problem

Liqun Qi · Fei Wang · Yiju Wang

Received: 6 June 2006 / Accepted: 24 August 2007 / Published online: 29 September 2007
© Springer-Verlag 2007

Abstract As a global polynomial optimization problem, the best rank-one approximation to higher order tensors has extensive engineering and statistical applications. Different from traditional optimization solution methods, in this paper, we propose some Z-eigenvalue methods for solving this problem. We first propose a direct Z-eigenvalue method for this problem when the dimension is two. In multidimensional case, by a conventional descent optimization method, we may find a local minimizer of this problem. Then, by using orthogonal transformations, we convert the underlying supersymmetric tensor to a pseudo-canonical form, which has the same E-eigenvalues and some zero entries. Based upon these, we propose a direct orthogonal transformation Z-eigenvalue method for this problem in the case of order three and dimension three. In the case of order three and higher dimension, we propose a heuristic orthogonal transformation Z-eigenvalue method by improving the local minimum with the lower-dimensional Z-eigenvalue methods, and a heuristic cross-hill Z-eigenvalue method by using the two-dimensional Z-eigenvalue method to find more local minimizers. Numerical experiments show that our methods are efficient and promising.

This work is supported by the Research Grant Council of Hong Kong and the Natural Science Foundation of China (Grant No. 10771120).

L. Qi (✉)
Department of Applied Mathematics, The Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong
e-mail: maqilq@polyu.edu.hk

F. Wang
Department of Mathematics, Hunan City University, Yiyang, Hunan, China
e-mail: mafwang@163.com

Y. Wang
School of Operations Research and Management Sciences, Qufu Normal University,
Rizhao Shandong 276800, China
e-mail: wyiju@hotmail.com

Keywords Polynomial optimization · Supersymmetric tensor · Orthogonal transformation · Z-eigenvalue

Mathematics Subject Classification (2000) 90C30

1 Introduction

In this paper, we consider the following global polynomial optimization problem

$$\begin{aligned} \min f(x) = \mathcal{A}x^m &= \sum_{i_1, i_2, \dots, i_m=1}^n a_{i_1, i_2, \dots, i_m} x_{i_1} x_{i_2} \cdots x_{i_m} \\ \text{s.t. } x^\top x &= 1 \end{aligned} \quad (1)$$

where $x \in \mathfrak{R}^n$, $m, n \geq 2$ are integers, f is a homogeneous polynomial of degree m with n variables, and \mathcal{A} is an m -th order n -dimensional real supersymmetric tensor. The tensor \mathcal{A} is an m -way array whose entries are addressed via m indices, and it is said to be supersymmetric if its entries $a_{i_1 \dots i_m}$ are invariant under any permutation of their indices $\{i_1, \dots, i_m\}$ [8, 15, 20]. In this paper, unless stated otherwise, all tensors involved are supersymmetric.

Let $\mathcal{A}x^{m-1}$ be a vector in \mathfrak{R}^n with its i th component as

$$(\mathcal{A}x^{m-1})_i = \sum_{i_2, \dots, i_m=1}^n a_{i, i_2, \dots, i_m} x_{i_2} \cdots x_{i_m}.$$

Obviously, the critical points of (1) satisfy the following equations for some $\lambda \in R$:

$$\begin{cases} \mathcal{A}x^{m-1} = \lambda x, \\ x^\top x = 1. \end{cases} \quad (2)$$

A real number λ satisfying (2) with a real vector x is called a **Z-eigenvalue** of \mathcal{A} in [20]. We call the real vector x with which a Z-eigenvalue λ solves (2) a **Z-eigenvector** of \mathcal{A} associated with the Z-eigenvalue λ . Then problem (1) is equivalent to finding the smallest Z-eigenvalue λ_{\min} and the corresponding Z-eigenvector. The properties of Z-eigenvalues were studied in [17, 20–22]. In [18, 20, 22], the number of Z-eigenvalues of an m -th order n -dimensional real supersymmetric tensor was discussed. For other types of eigenvalues/eigenvectors of tensors, see [17, 20].

An important application of problem (1) is the best rank-one approximation to a higher order supersymmetric tensor. A rank-one tensor means that it can be expressed as an outer product of a number of the identical vectors [8, 15]. Given a higher order supersymmetric tensor \mathcal{A} , if there exist a scalar λ and a unit-norm vector u such that the rank-one tensor $\bar{\mathcal{A}} \triangleq \lambda u^m$ minimizes the least-squares cost function

$$\tau(\bar{\mathcal{A}}) = \|\mathcal{A} - \bar{\mathcal{A}}\|_F^2$$

over the manifold of rank-one tensors, where $\|\cdot\|_F$ is the Frobenius norm, then λu^m is called the best rank-one approximation to tensor \mathcal{A} . The best rank-one approximation to a supersymmetric tensor has applications in signal processing, wireless communication systems, signal and image processing, data analysis, higher-order statistics, as well as independent component analysis [6–9, 11, 15, 19, 25]. As shown in [20], the Z-eigenvalue λ of \mathcal{A} with the largest absolute value and its Z-eigenvector x form the best rank-one approximation λx^m to \mathcal{A} . The same result is available in [8, 15] and possibly others as well. When m is odd, if λ is a Z-eigenvalue of \mathcal{A} with a Z-eigenvector x , then $-\lambda$ is a Z-eigenvalue of \mathcal{A} with a Z-eigenvector $-x$. Hence, in this case, the smallest Z-eigenvalue λ_{\min} is the Z-eigenvalue of the largest absolute value. On the other hand, when m is even, the Z-eigenvalue of the largest absolute value can be found by comparing the absolute values of λ_{\min} and the largest Z-eigenvalue λ_{\max} , which is the smallest Z-eigenvalue of $-\mathcal{A}$.

Another application of problem (1) is the positive definiteness identification problem for the multivariate form $f(x)$, defined in (1). The multivariate form $f(x)$ is called positive definite if $f(x) > 0$ for all $x \in \mathfrak{R}^n$, $x \neq 0$. This only applies to the case that m is even. Clearly, $f(x)$ is positive definite if and only if the smallest eigenvalue λ_{\min} of \mathcal{A} is positive. The positive definiteness of such a homogeneous polynomial form $f(x)$ plays an important role in the stability analysis of nonlinear autonomous systems via Liapunov's direct method in automatic control [1–4, 10, 12, 13, 16, 24].

When $m = 2$, \mathcal{A} is a symmetric matrix, and we may use some direct or iterative transformation methods to convert it to a canonical form, i.e., its diagonal form, to find its eigenvalues. Can this idea be extended to the higher dimensional case? If so, what is the canonical form for a supersymmetric tensor when $m \geq 3$? A notable property of Z-eigenvalues is that all of the Z-eigenvalues of a supersymmetric tensor are invariant under orthogonal transformation [20]. Can we use orthogonal transformation to find a “canonical form” of a supersymmetric tensor \mathcal{A} when $m \geq 3$? In this paper, we will define some “canonical forms” for a supersymmetric tensor, discuss their properties when $m \geq 2$, and propose some orthogonal transformation Z-eigenvalue methods for finding such “canonical forms” and hence solve problem (1) when $m = 3$ and $n \geq 2$. In the literature, there are discussions on “orthogonal tensor decomposition” [14], which is not related to “orthogonal transformation” discussed here.

When $n = 2$, all the Z-eigenvalues can be found easily. In the next section, we describe a direct Z-eigenvalue method for this case.

In Sect. 3, we will define the pseudo-canonical form of a supersymmetric tensor \mathcal{A} when $m \geq 2$. Such a pseudo-canonical form can be found by a conventional descent optimization method and orthogonal transformations. In Sect. 4, we show that we may find all the Z-eigenvalues and their Z-eigenvectors of a supersymmetric tensor \mathcal{A} when $m = n = 3$, based upon such a pseudo-canonical form. We call such a method a direct orthogonal transformation Z-eigenvalue method.

We may use a conventional method to find a critical point of (1), hence a Z-eigenvector of \mathcal{A} . Is such a Z-eigenvector a local minimizer of (1)? In Sect. 5, we give a sufficient condition and a necessary condition for this.

When $n \geq 4$, it is not efficient to find all the Z-eigenvalues directly. In Sect. 6, we first define the canonical form of a supersymmetric tensor \mathcal{A} . The canonical form always exists. If we have the canonical form, then we know the smallest Z-eigenvalue

λ_{\min} . However, such a canonical form is unknown and cannot be used as the target of an algorithm directly. Thus we define the r -th order pseudo-canonical form which is a good approximation to the canonical form. In Sect. 7, we propose a heuristic orthogonal transformation Z-eigenvalue method to find λ_{\min} in the case that $m = 3$ and $n \geq 4$. The target of this method is to find the third order pseudo-canonical form of \mathcal{A} . The direct methods studied in Sects. 2 and 4 for the cases that $m = 3, n = 2$ and 3 are subroutines of this method. We use them to improve the local minimum obtained by a conventional descent optimization method.

In Sect. 8, we propose another heuristic Z-eigenvalue method for the case that $m = 3$ and $n \geq 3$. After finding one local minimizer of the problem (1), we use the two-dimensional Z-eigenvalue method to find more local minimizers. We call this method a heuristic cross-hill Z-eigenvalue method.

Numerical results are reported in Sect. 9. They show that our algorithms are efficient and promising.

We use the entry a_{i_1, \dots, i_m} with the natural order for its indices, i.e., $i_1 \leq i_2 \leq \dots \leq i_m$ to represent other entries which are equal to it because of the supersymmetry. For example, when $m = 3$ and $n = 3$, $a_{123} = a_{132} = a_{213} = a_{231} = a_{312} = a_{321}$. We use a_{123} to represent them. When we assume that $a_{123} = 0$, it means that all of these six entries are zero.

By [20], if we know a Z-eigenvector x of \mathcal{A} , we may calculate the corresponding Z-eigenvalue by

$$\lambda = \mathcal{A}x^m. \tag{3}$$

For $j = 1, \dots, n$, we denote by $e^{(j)}$ the unit vector in \mathfrak{R}^n , whose j -th element is 1 and other elements are zero. Similarly, we denote by $e^{(1,j)}$ the unit vector in \mathfrak{R}^j , whose first element is 1 and other elements are zero.

2 A direct Z-eigenvalue method For $n = 2$

We assume that $n = 2$ in this section. Denote

$$\alpha_j = a_{i_1, \dots, i_m},$$

where $i_1 = \dots = i_{m-j} = 1, i_{m-j+1} = \dots = i_m = 2$ and $0 \leq j \leq m$.

Theorem 1 *Suppose that $n = 2$.*

If $\alpha_1 = a_{1, \dots, 1, 2} = 0$, then $\lambda = \alpha_0 = a_{1, \dots, 1}$ is a Z-eigenvalue of \mathcal{A} , with a Z-eigenvector $x = (1, 0)^\top$. If furthermore m is odd, then $\lambda = -a_{1, \dots, 1}$ is also a Z-eigenvalue of \mathcal{A} , with a Z-eigenvector $x = (-1, 0)^\top$.

The other Z-eigenvalues and corresponding Z-eigenvectors of \mathcal{A} can be found by finding real roots of the following one dimensional polynomial equation of t :

$$\sum_{j=0}^{m-1} \binom{m-1}{j} [\alpha_j t^{m-j} - \alpha_{j+1} t^{m-j+1}] = 0, \tag{4}$$

and substituting such real values of t to

$$x_1 = \pm \frac{t}{\sqrt{1+t^2}}, \quad x_2 = \pm \frac{1}{\sqrt{1+t^2}}, \tag{5}$$

and

$$\lambda = \sum_{j=0}^m \binom{m}{j} \alpha_j x_1^{m-j} x_2^j. \tag{6}$$

Proof When $n = 2$, (2) has the form

$$\begin{cases} \sum_{j=0}^{m-1} \binom{m-1}{j} \alpha_j x_1^{m-j} x_2^j = \lambda x_1 \\ \sum_{j=0}^{m-1} \binom{m-1}{j} \alpha_{j+1} x_1^{m-j} x_2^j = \lambda x_2 \\ x_1^2 + x_2^2 = 1 \end{cases} \tag{7}$$

and (3) has the form (6). By the second equation of (7), (7) has a solution with $x_2 = 0$ if and only if $\alpha_1 = 0$. In all the other cases, $x_2 \neq 0$. Let $t = \frac{x_1}{x_2}$. From the first two equations of (7), we have (4). By the third equation of (7), we have (5). The conclusions follow now. □

Obviously, Eq. (4) has at most $m + 1$ real roots. After finding all the Z-eigenvalues of \mathcal{A} , and the Z-eigenvectors associated with them, we may easily solve (1). Clearly, the case that $n = 2$ is trivial. But as stated in the introduction, we will use it as a subroutine for the higher dimensional case.

3 Orthogonal transformation and the pseudo-canonical form

Let \mathcal{A} be an m -th order n -dimensional supersymmetric tensor, $P = (p_{ij})$ be an $n \times n$ real matrix. Define $\mathcal{B} = P^m \mathcal{A}$ as another m -th order n -dimensional tensor with entries

$$b_{i_1, i_2, \dots, i_m} = \sum_{j_1, j_2, \dots, j_m=1}^n p_{i_1 j_1} p_{i_2 j_2} \cdots p_{i_m j_m} a_{j_1, j_2, \dots, j_m}.$$

Then \mathcal{B} is also a supersymmetric tensor. If P is an orthogonal matrix, then we say that \mathcal{A} and \mathcal{B} are **orthogonally similar**.

By [20], we have the following theorem.

Theorem 2 (Qi 2005) *Suppose that \mathcal{A} is an m -th order n -dimensional supersymmetric tensor, $\mathcal{B} = P^m \mathcal{A}$, P is an $n \times n$ orthogonal matrix. Then \mathcal{A} and \mathcal{B} have the same Z-eigenvalues. If λ is a Z-eigenvalue of \mathcal{A} with a Z-eigenvector x , then λ is a Z-eigenvalue of \mathcal{B} with a Z-eigenvector $y = Px$.*

Suppose that λ is a Z-eigenvalue of \mathcal{A} with a Z-eigenvector x . Let P be an orthogonal matrix with x^\top as its first row. Let $\mathcal{B} = P^m \mathcal{A}$. Then we see that $y = Px = e^{(1)}$. By

(2), we see that

$$b_{1,\dots,1} = \lambda$$

and

$$b_{1,\dots,1,i} = 0, \tag{8}$$

for $i = 2, \dots, n$. Based upon this, we now define the pseudo-canonical form of \mathcal{A} .

Definition 1 An m -th order n -dimensional supersymmetric tensor \mathcal{B} is said to be a pseudo-canonical form of another m -th order n -dimensional supersymmetric tensor \mathcal{A} if \mathcal{A} and \mathcal{B} are orthogonally similar and

$$b_{i,\dots,i,j} = 0 \tag{9}$$

for all $1 \leq i < j \leq n$.

If $m = 2$, then a pseudo-canonical form is a diagonal matrix.

Let \mathcal{A} be an m -th order n -dimensional supersymmetric tensor and $1 \leq j \leq n$. Denote $\mathcal{A}^{(j)}$ as an m -th order $(n - j + 1)$ -dimensional supersymmetric tensor obtained by the entries a_{i_1,\dots,i_m} , with $i_1, \dots, i_m = j, \dots, n$. By (2), we have the following observation:

Proposition 1 An m -th order n -dimensional supersymmetric tensor \mathcal{A} is a pseudo-canonical form if and only if the diagonal elements $a_{j,\dots,j}$ are Z -eigenvalues of $\mathcal{A}^{(j)}$ with $e^{(1,n-j+1)}$ as their Z -eigenvectors respectively for $j = 1, \dots, n$.

Suppose now we have found a local minimizer of (1) via a conventional descent optimization method. Then this local minimizer and its objective function value are a Z -eigenvector and the corresponding Z -eigenvalue. Using the method described after Theorem 2, we may convert \mathcal{A} to another tensor \mathcal{B} which is orthogonally similar to \mathcal{A} and satisfies (8) for $i = 2, \dots, n$. Now, by applying the conventional descent optimization method to $\mathcal{B}^{(2)}$, we may convert \mathcal{B} to a new tensor \mathcal{C} which is orthogonally similar to \mathcal{B} and hence \mathcal{A} , and satisfies (9) (with the letter b replaced by c in (9)) for $i = 1, 2$ and $i < j \leq n$. Continuing this process at most $n - 1$ times, we may find a pseudo-canonical form of \mathcal{A} .

4 A direct orthogonal transformation Z -eigenvalue method for $m = n = 3$

Suppose that $m = n = 3$ and \mathcal{A} is a pseudo-canonical form. We now describe a method to find all the Z -eigenvalues and corresponding Z -eigenvectors of \mathcal{A} .

Theorem 3 Suppose that $m = n = 3$ and \mathcal{A} is a pseudo-canonical form.

- (a) Then $\lambda = \pm a_{111}$ are Z -eigenvalues of \mathcal{A} with Z -eigenvectors $x = \pm(1, 0, 0)^\top$ respectively.
- (b) If $a_{122} = 0$, then $\lambda = \pm a_{222}$ are Z -eigenvalues of \mathcal{A} with Z -eigenvectors $x = \pm(0, 1, 0)^\top$ respectively.

(c) If $a_{123} = 0$, then

$$\lambda = \pm \frac{2a_{122t} + a_{222}}{\sqrt{t^2 + 1}} \tag{10}$$

are Z-eigenvalues of \mathcal{A} with Z-eigenvectors

$$x = \pm \frac{1}{\sqrt{t^2 + 1}}(t, 1, 0)^T \tag{11}$$

where t is a real root of the following quadratic equation of t :

$$(a_{111} - 2a_{122})t^2 - a_{222}t + a_{122} = 0. \tag{12}$$

(d) The other Z-eigenvalues and corresponding Z-eigenvectors of \mathcal{A} can be found by finding real solutions of the following polynomial equations of u and v :

$$\begin{cases} 2a_{123}u^2v + (2a_{133} - a_{111})u^2 + 2a_{233}uv + a_{333}u - a_{122}v^2 - 2a_{123}v - a_{133} = 0, \\ 2a_{123}uv^2 + 2(a_{133} - a_{122})uv - 2a_{123}u + (2a_{233} - a_{222})v^2 + a_{333}v - a_{233} = 0, \end{cases} \tag{13}$$

and substituting such real values of u and v to

$$x = \pm \frac{1}{\sqrt{u^2 + v^2 + 1}}(u, v, 1)^T \tag{14}$$

and

$$\lambda = \pm \frac{2a_{123}uv + 2a_{133}u + 2a_{233}v + a_{333}}{\sqrt{u^2 + v^2 + 1}}. \tag{15}$$

Proof Since \mathcal{A} is a pseudo-canonical form, we have $a_{112} = a_{113} = a_{223} = 0$. Then (2) has the form

$$\begin{cases} a_{111}x_1^2 + a_{122}x_2^2 + a_{133}x_3^2 + 2a_{123}x_2x_3 = \lambda x_1, \\ 2a_{122}x_1x_2 + 2a_{123}x_1x_3 + a_{222}x_2^2 + a_{233}x_3^2 = \lambda x_2, \\ 2a_{123}x_1x_2 + 2a_{133}x_1x_3 + 2a_{233}x_2x_3 + a_{333}x_3^2 = \lambda x_3, \\ x_1^2 + x_2^2 + x_3^2 = 1. \end{cases} \tag{16}$$

The conclusions (a) and (b) follow from (16) directly.

By the third equation of (16), if $a_{123} = 0$ and $x_3 = 0$, (16) becomes

$$\begin{cases} a_{111}x_1^2 + a_{122}x_2^2 = \lambda x_1, \\ 2a_{122}x_1 + a_{222}x_2 = \lambda, \\ x_1^2 + x_2^2 = 1. \end{cases} \tag{17}$$

Let $t = \frac{x_1}{x_2}$. From the first two equations of (17), we have (12). By the third equation of (17), we have (11). By the second equation of (17), we have (10). This proves (c).

It is easy to verify that the conclusions (a), (b) and (c) cover all the case that $x_3 = 0$. Assume that $x_3 \neq 0$. Let $u = \frac{x_1}{x_3}$ and $v = \frac{x_2}{x_3}$. From the first three equations of (16),

we have (13). By the fourth equation of (16), we have (14). By (3), (15) holds. The conclusion (d) follows now. \square

Remark The second equation of (13) is linear in variable u . By eliminating u in (13), we have

$$h(v) \equiv h_0v^6 + h_1v^5 + \dots + h_5v + h_6 = 0,$$

where

$$\begin{aligned} h_0 &= -4a_{122}a_{123}^2, \\ h_1 &= 2a_{123}a_{222}^2 - 8a_{122}a_{123}a_{133} + 8a_{122}^2a_{123} - 8a_{123}^3 - 4a_{123}a_{222}a_{233}, \\ h_2 &= -2a_{123}a_{222}a_{333} + 2a_{133}a_{222}^2 - 4a_{133}a_{222}a_{233} - 4a_{111}a_{233}^2 - a_{111}a_{222}^2 \\ &\quad + 4a_{111}a_{222}a_{233} + 8a_{122}a_{233}^2 - 4a_{122}a_{222}a_{233} - 4a_{122}a_{133}^2 - 4a_{122}^3 \\ &\quad + 8a_{122}^2a_{133} + 24a_{122}a_{123}^2 - 20a_{123}^2a_{133}, \\ h_3 &= -2a_{133}a_{222}a_{333} - 4a_{111}a_{233}a_{333} + 2a_{111}a_{222}a_{333} + 8a_{122}a_{233}a_{333} + 4a_{123}a_{233}^2 \\ &\quad - 2a_{122}a_{222}a_{333} + 32a_{122}a_{123}a_{133} - 16a_{122}^2a_{123} - 16a_{123}a_{133}^2 + 16a_{123}^3, \\ h_4 &= -4a_{133}a_{233}^2 + 4a_{133}a_{222}a_{233} - a_{111}a_{333}^2 + 4a_{111}a_{233}^2 - 2a_{111}a_{222}a_{233} \\ &\quad - 4a_{122}a_{233}^2 + 2a_{122}a_{333}^2 + 6a_{123}a_{233}a_{333} - 2a_{123}a_{222}a_{333} - 20a_{122}a_{123}^2 \\ &\quad + 24a_{123}^2a_{133} - 4a_{133}^3 + 8a_{122}a_{133}^2 - 4a_{122}^2a_{133}, \\ h_5 &= -2a_{123}a_{233}^2 - 2a_{133}a_{233}a_{333} + 2a_{111}a_{233}a_{333} - 2a_{122}a_{233}a_{333} + 2a_{123}a_{333}^2 \\ &\quad + 8a_{123}a_{133}^2 - 8a_{122}a_{123}a_{133} - 8a_{123}^3, \\ h_6 &= 2a_{133}a_{233}^2 - a_{111}a_{233}^2 - 2a_{123}a_{233}a_{333} - 4a_{123}^2a_{133}. \end{aligned}$$

Note that $h_0 = h_1 = 0$ when $a_{123} = 0$. For each real root of $h(v) = 0$, if it satisfies

$$2a_{123}v^2 + 2(a_{133} - a_{122})v \neq 2a_{123},$$

then we have

$$u = -\frac{(2a_{233} - a_{222})v^2 + a_{333}v - a_{233}}{2a_{123}v^2 + 2(a_{133} - a_{122})v - 2a_{123}};$$

if it satisfies

$$2a_{123}v^2 + 2(a_{133} - a_{122})v = 2a_{123}$$

and

$$(2a_{233} - a_{222})v^2 + a_{333}v = a_{233},$$

then we may substitute it to the first equation of (13) to find all possible real solution of u . In this way, we may find all real solutions (u, v) of (13).

5 Conditions for local minimizers

We now give a necessary condition and a sufficient condition for a Z-eigenvector to be a local minimizer of (1).

Theorem 4 *Assume that $m \geq 2$. Suppose that x is a Z-eigenvector of \mathcal{A} , associated with a Z-eigenvalue λ . Then λ is an eigenvalue of the symmetric matrix $A \triangleq \mathcal{A}x^{m-2}$, and x is an eigenvector of A , associated with the eigenvalue λ . Suppose that $\lambda_i, i = 2, \dots, n$, are the other eigenvalues of A . Then the vector x is a local minimizer of (1) only if*

$$\lambda_i \geq \frac{1}{m-1}\lambda \tag{18}$$

for $i = 2, \dots, n$; and the vector x is a local minimizer of (1) if

$$\lambda_i > \frac{1}{m-1}\lambda, \tag{19}$$

for $i = 2, \dots, n$.

Proof Since x is a Z-eigenvector of \mathcal{A} , associated with a Z-eigenvalue λ , by (2), we have

$$Ax = \mathcal{A}x^{m-1} = \lambda x.$$

This shows that λ is an eigenvalue of the symmetric matrix A , and x is an eigenvector of A , associated with the eigenvalue λ . The Lagrangian function of (1) can be written as

$$\frac{1}{m}\mathcal{A}x^m - \lambda(x^\top x - 1).$$

Certainly, any local minimizer of problem (1) with the optimal Lagrange multiplier constitute a Z-eigenpair of tensor \mathcal{A} . Now, we consider the Hessian of the Lagrangian function at x , i.e. $(m-1)A - \lambda I$, where I is the $n \times n$ unit matrix. Then by the optimization theory, x is a local minimizer of (1) only if for any unit vector u in the critical cone of (1) at x ,

$$u^\top [(m-1)A - \lambda]u = (m-1)u^\top Au - \lambda \geq 0. \tag{20}$$

Since the feasible region of (1) is the unit sphere in \mathfrak{R}^n , the critical cone of (1) is the set of all vectors perpendicular to x . Since A is a real symmetric matrix and λ is an eigenvalue of A , with x as an eigenvector of A , associated with λ , by the properties of eigenvalues and eigenvectors of real symmetric matrices, we see that the condition (20) is equivalent to the condition (18). Similarly, we may show that (19) is a sufficient condition for x to be a local minimizer of (1). This completes the proof. \square

As a referee pointed out, this theorem has a similar flavor to Theorem 2 of [23]. It is possible to reconcile these two results under appropriate conditions. This may be studied further in future.

6 The canonical form and the r th order pseudo-canonical form

We now define the canonical form.

Definition 2 An m -th order n -dimensional supersymmetric tensor \mathcal{B} is called a canonical form of another m -th order n -dimensional supersymmetric tensor \mathcal{A} if it is a pseudo-canonical form of \mathcal{A} and $b_{i,\dots,i}$ is the smallest Z -eigenvalue of $\mathcal{B}^{(i)}$ for $i = 1, \dots, n$.

The canonical form always exists. When $n \geq 4$, it is not realistic to use the canonical form as a direct target for an algorithm. A pseudo-canonical form is an approximation to the canonical form. We now define some better approximations to the canonical form. In the next section, we will use such approximations as targets to guide a conventional descent optimization method to go further from a local minimizer of (1).

Definition 3 For two m -th order n -dimensional supersymmetric tensors \mathcal{A} and \mathcal{B} , tensor \mathcal{B} is called a first order pseudo-canonical form of tensor \mathcal{A} if it is a pseudo-canonical form of \mathcal{A} and $b_{i,\dots,i} \leq b_{j,\dots,j}$ for $1 \leq i < j \leq n$.

Let $1 \leq j < k \leq n$. We use $\mathcal{B}(j, k)$ to denote the m -th order two dimensional supersymmetric tensor whose entries are b_{i_1, i_2, \dots, i_m} , where $i_1, i_2, \dots, i_m = j, k$, and use $[\mathcal{B}(j, k)]_{\min}$ to denote the smallest Z -eigenvalue of $\mathcal{B}(j, k)$.

Definition 4 An m -th order n -dimensional supersymmetric tensor \mathcal{B} is called a second order pseudo-canonical form of another m -th order n -dimensional supersymmetric tensor \mathcal{A} if it is a first order pseudo-canonical form of \mathcal{A} and

$$b_{1,\dots,1} = \min_{1 \leq j < k \leq n} [\mathcal{B}(j, k)]_{\min}. \quad (21)$$

Let $1 \leq j < k < l \leq n$. We use $\mathcal{B}(j, k, l)$ to denote the m -th order three dimensional supersymmetric tensor whose entries are b_{i_1, i_2, \dots, i_m} , where $i_1, i_2, \dots, i_m = j, k, l$, and use $[\mathcal{B}(j, k, l)]_{\min}$ to denote the smallest Z -eigenvalue of $\mathcal{B}(j, k, l)$.

Definition 5 For two m -th order n -dimensional supersymmetric tensors \mathcal{A} and \mathcal{B} , tensor \mathcal{B} is called a third order pseudo-canonical form of tensor \mathcal{A} if \mathcal{B} is a first order pseudo-canonical form of \mathcal{A} and

$$b_{1,\dots,1} = \min_{1 \leq j < k < l \leq n} [\mathcal{B}(j, k, l)]_{\min}. \quad (22)$$

It is not difficult to see that a third order pseudo-canonical form is also a second order pseudo-canonical form. A third order pseudo-canonical form is a better approximation to the canonical form as (22) involves all entries of \mathcal{B} . We may define higher order pseudo-canonical forms similarly. We do not go to such details.

7 A heuristic orthogonal transformation Z-eigenvalue method For $m = 3$ and $n \geq 4$

In this section, we assume that $m = 3$ and $n \geq 4$. Suppose that we have a conventional descent optimization algorithm, called Algorithm M. Given a unit vector $x \in \mathfrak{R}^n$, we may use Algorithm M to find a local minimizer y of (1) such that $f(y) \leq f(x)$. In the algorithm given below, there are three optional stopping places, at Steps 8, 9 and 10, respectively. Then we get a first order, or a second order, or a third order pseudo-canonical form of tensor \mathcal{A} , respectively. The higher order implies more works and better global optimality.

Algorithm 1 *Step 0* Let P be the unit $n \times n$ matrix I .
Step 1 Let

$$|a_{iii}| = \max_{1 \leq j \leq n} |a_{jjj}|.$$

Let $x = e^{(i)}$ if $a_{iii} \leq 0$. Otherwise, let $x = -e^{(i)}$.

Step 2 Use Algorithm M to find a local minimizer y of (1) such that $f(y) \leq f(x)$. Replace x by y .

Step 3 Use vector x^\top as the first row to construct an orthogonal matrix Q such that $b_{111} = f(x)$, where $\mathcal{B} = Q^3 \mathcal{A}$. Replace P by QP .

Step 4 Let $i = 1$.
Step 5 If

$$|b_{iii}| = \max_{i \leq j \leq n} |b_{jjj}|$$

and

$$b_{ijj} = 0$$

for all j satisfying $i < j \leq n$, go to Step 6. Otherwise, go to Step 7.

Step 6 If $i = n$, go to Step 8. Otherwise, replace i by $i + 1$ and go to Step 5.
Step 7 Assume that

$$|b_{kkk}| = \max_{i \leq j \leq n} |b_{jjj}|.$$

Add additional constraints $x_l = 0$ for $1 \leq l < i$ to problem (1). Denote this problem as $(1)_i$ and its objective function as $f^{(i)}(x)$. Take $w = e^{(k)} \in \mathfrak{R}^n$ as an initial point to find a local minimizer z of $(1)_i$ via Algorithm M such that $f^{(i)}(z) \leq f^{(i)}(w) = b_{kkk}$. Use $(e^{(1)})^\top, \dots, (e^{(i-1)})^\top, z^\top$ as the first i rows to construct an orthogonal matrix Q such that

$$c_{jjj} = b_{jjj}$$

for $1 \leq j < i$, and

$$c_{iii} = b_{kkk},$$

where $C = Q^m B$. Replace B by C and P by QP . Go to Step 4.

Step 8 For the first order heuristic orthogonal transformation Z-eigenvalue method, we stop here and the output is $x = P^T e^{(1)}$. Otherwise go to the next step.

Step 9 Search all possible j and k such that

$$[B(j, k)]_{\min} < -|b_{111}|.$$

As long as we find such a pair (j, k) , let y be a vector in \mathfrak{N}^n such that its j -th and k -th components form a Z-eigenvector of $B(j, k)$ and other components are zero. Let $x = P^T y$ and $P = I$. Go to Step 2. If such a pair (j, k) does not exist, stop here with the output $x = P^T e^{(1)}$ if we conduct the second order heuristic orthogonal transformation Z-eigenvalue method; otherwise go to the next step.

Step 10 Search all possible j, k and l such that

$$[B(j, k, l)]_{\min} < -|b_{111}|.$$

As long as we find such a triple (j, k, l) , let y be a vector in \mathfrak{N}^n such that its j -th, k -th and l -th components form a Z-eigenvector of $B(j, k, l)$ and other components are zero. Let $x = P^T y$ and $P = I$. Go to Step 2. If such a triple (j, k, l) does not exist, stop with the output $x = P^T e^{(1)}$.

Note that we have used the lower dimensional methods discussed in Sects. 2 and 4 as subroutines in Steps 8 and 9 of Algorithm 1. We may use a simple orthogonal transformation to convert all the diagonal elements of B to be non-positive. Now it is not difficult to prove the following theorem.

Theorem 5 *Algorithm 1 terminates in a finitely many iterations. It will find a first, or second, or third order pseudo-canonical form of \mathcal{A} , depending upon the termination criterion set at Steps 8, 9 or 10.*

8 A heuristic cross-hill Z-eigenvalue method For $m = 3$ and $n \geq 3$

In this section, we would propose another heuristic Z-eigenvalue method to solve (1) for $m = 3$ and $n \geq 3$.

Certainly, the global minimum of (1) is negative unless \mathcal{A} is a zero tensor. Suppose that we have found a local minimizer $y = y^{(1)}$ of (1) with a negative objective function value. We may find $n - 1$ unit vectors $y^{(i)}$ of \mathfrak{N}^n for $i = 2, \dots, n$ such that $\{y^{(i)} : i = 1, \dots, n\}$ constitute an orthogonal basis of \mathfrak{N}^n . Now, for $i = 2, \dots, n$, restrict problem (1) on the plane spanned by y and $y^{(i)}$. Certainly, this is a two-dimensional form of problem (1) and the current local minimizer y of (1) is also a local minimizer of the restricted problem. By Theorem 1, this two dimensional problem has at most two local minimizers with negative objective function values. Then we

may use the method discussed in Sect. 2 to find $z^{(i)}$, the other local minimizer of the two-dimensional problem with a negative objective function value for $i = 2, \dots, n$. Since $z^{(i)}$ may not be a local minimizer of (1), and y and $z^{(i)}$ are separated by a “hill” of the objective function value in two dimensional case, if we use a conventional descent optimization method with $z^{(i)}$ as the starting point, we will find a local minimizer $w^{(i)}$ of (1), which has a negative objective function value and is different from y . We may continue this process until no new local minimizers can be found. Comparing the objective function values of these local minimizers of (1), we may have a better solution. Assume that $m = 3$ and $n \geq 3$. Our algorithm is as follows.

Algorithm 2 *Step 1* Let

$$|a_{iii}| = \max_{1 \leq j \leq n} |a_{jjj}|.$$

Let $x = e^{(i)}$ if $a_{iii} \leq 0$. Otherwise, let $x = -e^{(i)}$.

Step 2 Use Algorithm M to find a local minimizer $y = y^{(1)}$ of (1) such that $f(y) \leq f(x)$. Let $P = \{y^{(1)}\}$.

Step 3 Find $n - 1$ unit vectors $y^{(i)}$ of \mathfrak{R}^n for $i = 2, \dots, n$ such that $\{y^{(i)} : i = 1, \dots, n\}$ is an orthogonal basis of \mathfrak{R}^n .

Step 4 For $i = 2, \dots, n$, consider the restriction problem of (1) on the plane spanned by y and $y^{(i)}$. Use the method discussed in Sect. 2 to find another local minimizer $z^{(i)}$ with a negative objective function value, of the restricted two-dimensional problem, if such a point exists. Then use Algorithm M to find a local minimizer $w^{(i)}$ of (1), such that $f(w^{(i)}) \leq f(z^{(i)})$.

Step 5 For each local minimizer $\{w^{(i)}\}$ found in Step 4, if it is not in P , add it to P , and put it as $y^{(1)}$ and repeat Steps 3–4, until no new local minimizers can be found.

Step 6 Compare the function value of f for all local minimizers found in P . The point with smallest value of f is the solution found by this algorithm.

Clearly, this algorithm also terminates in a finitely many iterations as long as the number of local minimizers is finite. This covers the nonsingular cases.

9 Numerical results

In this section, we present preliminary numerical experiments for some third order n dimensional supersymmetric tensors with Algorithms 1 and 2. The computation was done on a personal computer (Pentium IV, 2.8 GHz) by running Matlab 7.0. As for the descent solution method for solving the constrained optimization problem, i.e., Algorithm M used in Algorithms 1 and 2, we adopt the projected gradient method proposed in [5] as the projection from a point in R^n to the unit sphere can easily be computed.

To construct the testing problems in our numerical experiments, we take two sets of third order supersymmetric tensors of different dimensions. The elements of tensors in one set are generated by a uniform distribution in the interval $(-1,1)$; while the elements of tensors in another set are generated by a standard normal distribution.

Table 1 Results of TPI

| Dim | Num | Alg1 | | Alg2 | | |
|-----|-------|--------|--------|--------|--------|-----|
| | | RS (%) | AT (s) | RS (%) | AT (s) | AN |
| 3 | 1,000 | 100.0 | 0.013 | 99.4 | 0.038 | 2.9 |
| 4 | 1,000 | 96.6 | 0.031 | 99.5 | 0.103 | 3.6 |
| 6 | 1,000 | 91.6 | 0.089 | 99.7 | 0.335 | 4.6 |
| 8 | 100 | 88 | 0.196 | 99.0 | 0.728 | 5.6 |
| 10 | 100 | 89 | 0.379 | 99.0 | 1.359 | 6.4 |

Table 2 Results of TPII

| Dim | Num | Alg1 | | Alg2 | | |
|-----|-------|--------|--------|--------|--------|-----|
| | | RS (%) | AT (s) | RS (%) | AT (s) | AN |
| 3 | 1,000 | 100.0 | 0.014 | 99.2 | 0.042 | 2.9 |
| 4 | 1,000 | 96.5 | 0.031 | 99.3 | 0.097 | 3.6 |
| 6 | 1,000 | 92.2 | 0.085 | 99.8 | 0.265 | 4.5 |
| 8 | 100 | 88 | 0.178 | 99.0 | 0.521 | 5.3 |
| 10 | 100 | 87 | 0.331 | 99.0 | 0.933 | 6.2 |

In the following, these two sets of testing problems are labeled as TPI and TPII, respectively.

To analyze the numerical results of our algorithms, we compute the global minimizers of the testing problems by the uniform grid method along with the descent solution method.

Tables 1 and 2 show the performance of Algorithms 1 and 2 for problems TPI and TPII, where **Dim** denotes the dimension of the tensor \mathcal{A} , **Num** denotes the number of tests for each dimension, **Alg1** and **Alg2** denote Algorithms 1 and 2, respectively, **RS** denotes the success ratio of finding the global minimizers, **AT** denotes the average time for each sample, and **AN** denotes the average number of local minimizers found by Algorithm 2 for each sample.

Numerical results show that Algorithms 1 and 2 are efficient and promising. They show that Algorithm 2 has higher success ratios, while Algorithm 1 uses less computational time.

The following example shows the procedures of Algorithms 1 and 2 in detail.

Example 9.1 Consider a supersymmetric $6 \times 6 \times 6$ tensor whose entries are $a_{iii} = i$ for $i = 1, 2, \dots, 6$, and $a_{ii,i+1} = 10$ for $i = 1, 2, \dots, 5$, and zero otherwise. The global minimizer of (1) for this example is

$$x^* = -(0.0, 0.0, 0.0, 0.6577, 0.6802, 0.3237)^\top$$

and the global minimum value is $f(x^*) = -16.235$. By the projection gradient descent method, the initial estimate is $x^{(0)} = -(0, 0, 0, 0, 0, 1)^\top$ and $f(x^{(0)}) = -6$.

By Algorithm 1, the output of the first order termination criterion is $x^{(1)} = x^{(0)}$ and $f(x^{(1)}) = -6$, the output of the second order termination criterion is

$x^{(2)} = -(0.0, 0.0, 0.0, 0.0, 0.8249, 0.5653)^\top$ and $f(x^{(2)}) = -15.430$, and the output of the third order termination criterion is $x^{(3)} = x^*$ and $f(x^{(3)}) = -16.235$.

By Algorithm 2, based on $x^{(0)}$, the local minimizers found by one round of the heuristic cross-hill Z-eigenvalue method are

$$\bar{x}^{(1)} = -(0.0, 0.0, 0.0, 0.6577, 0.6802, 0.3237)^\top$$

and

$$\bar{x}^{(2)} = -(0.0, 0.0, 0.0, 0.0, 0.8249, 0.5653)^\top.$$

We have $f(\bar{x}^{(1)}) = -16.235$ and $f(\bar{x}^{(2)}) = -15.430$. Based on $\bar{x}^{(1)}$, the local minimizers found by a new round of the heuristic cross-hill Z-eigenvalue method are

$$\bar{x}^{(3)} = -(0.0, 0.0, 0.0, 0.2061, 0.8139, 0.5432)^\top$$

and

$$\bar{x}^{(4)} = (0.0, 0.0, 0.0, 0.0, 0.6923, -0.7216)^\top.$$

We have $f(\bar{x}^{(3)}) = -15.455$ and $f(\bar{x}^{(4)}) = -10.971$. However, based on $\bar{x}^{(2)}$ or $\bar{x}^{(3)}$ or $\bar{x}^{(4)}$, there are no new local minimizers which can be found by a round of the heuristic cross-hill Z-eigenvalue method. Therefore, Algorithm 2 terminates in five iterations and the global minimizer found by this Algorithm is, i.e., $x^* = \bar{x}^{(1)}$.

Acknowledgments The authors are thankful to the associate editor and two referees for their comments.

References

1. Anderson, B.D., Bose, N.K., Jury, E.I.: Output feedback stabilization and related problems-solutions via decision methods. *IEEE Trans. Autom. Control* **AC20**, 55–66 (1975)
2. Bose, N.K., Kamat, P.S.: Algorithm for stability test of multidimensional filters. *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-22**, 307–314 (1974)
3. Bose, N.K., Modarressi, A.R.: General procedure for multivariable polynomial positivity with control applications. *IEEE Trans. Autom. Control* **AC-21**, 596–601 (1976)
4. Bose, N.K., Newcomb, R.W.: Tellegon's theorem and multivariable realizability theory. *Int. J. Electron.* **36**, 417–425 (1974)
5. Calamai, P.H., Moré, J.J.: Projected gradient methods for linearly constrained problems. *Math. Program.* **39**, 93–116 (1987)
6. Cardoso, J.F.: High-order contrasts for independent component analysis. *Neural Comput.* **11**, 157–192 (1999)
7. Comon, P.: Independent component analysis, a new concept?. *Signal Process.* **36**, 287–314 (1994)
8. De Lathauwer, L., De Moor, B., Vandewalle, J.: On the best rank-1 and rank- (R_1, R_2, \dots, R_N) approximation of higher-order tensor. *SIAM J. Matrix Anal. Appl.* **21**, 1324–1342 (2000)
9. De Lathauwer, L., Comon, P., De Moor, B., Vandewalle, J.: 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
10. Fu, M.: Comments on 'A procedure for the positive definiteness of forms of even-order'. *IEEE Trans. Autom. Control* **43**, 1430 (1998)

11. Grigorasu, V.S., Regalia, P.A.: Tensor displacement structures and polyspectral matching. In: Kailath, T., Sayed, A.H., (eds.) *Fast Reliable Algorithms for Structured Matrices*, Chap. 9. SIAM Publications, Philadelphia (1999)
12. Hasan, M.A., Hasan, A.A.: A procedure for the positive definiteness of forms of even-order. *IEEE Trans. Autom. Control* **AC-41**, 615–617 (1996)
13. Jury, E.I., Mansour, M.: Positivity and nonnegativity conditions of a quartic equation and related problems. *IEEE Trans. Autom. Control* **AC26**, 444–451 (1981)
14. Kolda, T.G.: Orthogonal tensor decomposition. *SIAM J. Matrix Anal. Appl.* **23**, 243–255 (2001)
15. Kofidis, E., Regalia, P.A.: On the best rank-1 approximation of higher-order supersymmetric tensors. *SIAM J. Matrix Anal. Appl.* **23**, 863–884 (2002)
16. Ku, W.H.: Explicit criterion for the positive definiteness of a general quartic form. *IEEE Trans. Autom. Control* **10**, 372–373 (1965)
17. Lim, L.-H.: Singular values and eigenvalues of tensors: A variational approach. In: *Proceedings of the First IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, December 13–15, 2005, pp. 129–132
18. Ni, G., Qi, L., Wang, F., Wang, Y.: The degree of the E-characteristic polynomial of an even order tensor. *J. Math. Anal. Appl.* **329**, 1218–1229 (2007)
19. Nikias, C.L., Petropulu, A.P.: *Higher-Order Spectra Analysis, A Nonlinear Signal Processing Framework*. Prentice-Hall, Englewood Cliffs (1993)
20. Qi, L.: Eigenvalues of a real supersymmetric tensor. *J. Symb. Comput.* **40**, 1302–1324 (2005)
21. Qi, L.: Rank and eigenvalues of a supersymmetric tensor, a multivariate homogeneous polynomial and an algebraic surface defined by them. *J. Symb. Comput.* **41**, 1309–1327 (2006)
22. Qi, L.: Eigenvalues and invariants of tensors. *J. Math. Anal. Appl.* **325**, 1363–1377 (2007)
23. Regalia, P.A., Mboup, M.: Properties of some blind equalization criteria in noisy multi-user environments. *IEEE Trans. Signal Process.* **49**, 3112–3122 (2001)
24. Wang, F., Qi, L.: Comments on ‘Explicit criterion for the positive definiteness of a general quartic form’. *IEEE Trans. Autom. Control* **50**, 416–418 (2005)
25. Zhang, T., Golub, G.H.: Rank-1 approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* **23**, 534–550 (2001)