THE POINT-WISE CONVERGENCE OF SHIFTED SYMMETRIC HIGHER ORDER POWER METHOD

GANG LUO
School of Mathematical Sciences and LPMC, Nankai University
Tianjin 300071, China

QINGZHI YANG*
School of Mathematical Sciences and LPMC, Nankai University
Tianjin 300071, China
and School of Mathematics and Statistics, Kashi University
Kashi 844006, China

(Communicated by Jie Sun)

Abstract. Shifted symmetric higher-order power method (SS-HOPM) is an effective method of computing tensor eigenpairs. However the point-wise convergence of SS-HOPM has not been proven yet. In this paper, we provide a solid proof of the point-wise convergence of SS-HOPM via Lojasiewicz inequality. In particular, we establish a mapping from the sequence generated by the algorithm to a specially defined sequence. Using Lojasiewicz inequality, we prove the convergence of the new sequence, then the original sequence is convergent based on the relation of two sequences.

1. Introduction. Tensor eigenvalues and eigenvectors have received much attention in recent years due to its wide applications in many fields [7, 8, 10, 11, 9, 14]. There are several definitions of tensor eigenvalues and eigenvectors. A tensor eigenvalue and an associated eigenvector form an eigenpair. In this paper, we refer to eigenpair as Z-eigenpair [8] or \( l^2 \) eigenpair [7] which is defined below.

Definition 1.1. Assume that \( A \) is a symmetric \( m \)-th order \( n \)-dimensional real-valued tensor. For any \( n \)-dimensional vector \( x \), define

\[
(Ax^{m-1})_i = \sum_{i_2,\ldots,i_m=1}^{n} a_{i_1i_2\cdots i_m} x_{i_1} \cdots x_{i_m},
\]

for \( i_1 = 1,\ldots,n \). If there exist vector \( x \in \mathbb{R}^n \) and real number \( \lambda \in \mathbb{R} \) such that

\[
Ax^{m-1} = \lambda x \quad \text{and} \quad x^\top x = 1,
\]

then \( \lambda \) is called an eigenvalue of \( A \), \( x \) is called the associated eigenvector, and \( (\lambda, x) \) is called an eigenpair of \( A \).

2010 Mathematics Subject Classification. Primary: 15A18, 90C26; Secondary: 15A69.

Key words and phrases. Z-eigenvalue, shifted symmetric higher-order power method, point-wise convergence, Lojasiewicz inequality.

The second author is supported by Natural Science Foundation of Xinjiang (Grant No. 2017D01A14).

* Corresponding author: Qingzhi Yang.
A tensor is called symmetric if its elements are invariant under any permutation of its indices. It is shown in [7] that any eigenpair for a symmetric tensor $A$ corresponds to a Karush-Kuhn-Tucker (KKT) point of the constrained optimization problems

$$\begin{align*}
\max \quad & Ax^m \\
\text{s.t.} \quad & x^\top x = 1,
\end{align*}$$

or

$$\begin{align*}
\min \quad & Ax^m \\
\text{s.t.} \quad & x^\top x = 1,
\end{align*}$$

where $Ax^m = \sum_{i_1,\ldots,i_m=1}^n a_{i_1\cdots i_m} x_{i_1} \cdots x_{i_m}$. Finding the best symmetric rank-one approximation to a given symmetric tensor $A$ is equivalent to finding its largest eigenvalue in the absolute value sense, and it is equivalent to solving problem (3) or (4).

Problem (3) or (4) is nonlinear and nonconvex in general. In [4], Kofidis and Regalia presented the symmetric higher-order power method (S-HOPM) to solve the above problems, which is a direct generalization of classic power method[5]. The S-HOPM is not guaranteed to converge generally. Kofidis and Regalia[4] showed that for even order case if $Ax^m$ was convex, the generated sequence of the object function values would monotonically increase. In the context of independent component analysis (ICA), in which a fourth-order (i.e., even order) tensor is involved, shifted variants of power method have been developed [2, 13]. In [17], Kolda and Mayo extended the shifted power method to arbitrary order case, and proposed a shifted symmetric higher-order power method (SS-HOPM).

The SS-HOPM can be seen as a special variant of S-HOPM. The generated object function value sequence guarantees a monotonic convergence with sufficiently large or small shift term added. The factor sequence generated converges to an eigenvector of the related tensor. For generic tensors, the number of the eigenvectors is finite, and under this condition the factor sequence of SS-HOPM is guaranteed to converge globally whenever the sequence has one isolated accumulation point. Each accumulation point of factor sequence generated by SS-HOPM is an eigenvector of the related tensor. For generic tensors, the number of the eigenvectors is finite, and under this condition the factor sequence of SS-HOPM is guaranteed to converge globally. In [3], Dustin and Sturmfels showed that the number of the eigenvalues for a symmetric tensor is finite, while no similar conclusion exists for eigenvectors of a symmetric tensor. On page 1107 of [17], Kolda and Mayo conjectured that the factor sequence converges without the one isolated accumulation point condition, i.e., the factor sequence converges for all symmetric tensors. In this paper, we will show that this conjecture is true. To make a distinguish, we refer to the point-wise convergence of SS-HOPM as the global convergence of the generated factor sequence.

The square summability of the increment of the factor sequence has been shown in [17], while the global convergence cannot be inferred from it. Lojasiewicz inequality [16] is a powerful feature of real analytic functions. In [12], it is shown that the validity of Lojasiewicz inequality at a cluster point of gradient-related descent iteration together with the sufficient decrease condition enforce absolute summability of increments, which implies the global convergence of the iteration. The power of Lojasiewicz inequality in proving the convergence of algorithms on non-convex optimization problems can be seen in many literatures, for example [15, 19] and references therein. In [1], Uschmajew proved the point-wise convergence of Gauss-Seidel higher-order power method [6] by connecting the original constrained optimization to an unconstrained optimization, and proved the convergence of the
unconstrained optimization using the results in [12]. Motivated by this, we try to use this approach to prove the point-wise convergence of SS-HOPM, which is the main contribution of this paper. Specifically we define a new sequence and a related analytic function based on the sequence generated by SS-HOPM. The new sequence and function form a new gradient-related descent iteration, and the sufficient decrease condition holds. Using the well established results in [12], we show the global convergence of the new sequence, which in return guarantees the point-wise convergence of SS-HOPM. The point-wise convergence of many tensor related algorithms has not been proven yet, for example, [18, 2, 13], while the numerical experiments have witnessed their global convergence, thus the approach in our paper may give a novel framework to prove such point-wise convergence.

The rest of this paper is organized as follows. In Section 2, we recall the shifted symmetric higher-order power method and present some useful properties. In Section 3, first we introduce the global convergence theorem from [12], and then define a special sequence and function on the sequence generated by SS-HOPM. In section 4, we prove the global convergence of the new sequence, from which the point-wise convergence of SS-HOPM follows. A brief conclusion is made in Section 5.

Notation: Vectors are denoted by lowercase letters, e.g., $x$, matrices by capital letters, e.g., $X$, and tensors by calligraphic letters, e.g., $A$. $\| \cdot \|$ denotes Euclidean norm for vectors. The matrix $I$ is the identity matrix. Given a matrix $A$, $\rho(A)$ denotes its spectral norm. For a sequence $\{a_n\}$, $a_n \to a^*$ denotes that the sequence $\{a_n\}$ converges to $a^*$. $\Omega$ denotes the set $\{x \in \mathbb{R}^n \mid \|x\| = 1\}$ and $\mathcal{R}^{[m,n]}$ denotes the set of all $m$th-order $n$-dimensional real-valued tensors.

2. Some properties of SS-HOPM. In this section, we recall the shifted symmetric higher-order power method [17] and present some useful properties.

To solve (3) or (4), symmetric higher-order power method (Algorithm 1) was presented in [4].

Algorithm 1 Symmetric higher-order power method (S-HOPM)

Given a symmetric tensor $A \in \mathbb{R}^{[m,n]}$.

Require: $x_0 \in \mathbb{R}^n$ with $\|x_0\| = 1$. Let $\lambda_0 = Ax_0^m$.

1: for $k = 0, 1, \ldots$ do
2: $\hat{x}_{k+1} \leftarrow Ax_k^{m-1}$
3: $x_{k+1} \leftarrow \hat{x}_{k+1}/\|\hat{x}_{k+1}\|$
4: $\lambda_{k+1} \leftarrow Ax_k^m$
5: end for

In each iteration, S-HOPM computes the gradient of $Ax^m$ and projects it to unit sphere $\Omega$. Unfortunately S-HOPM may not be stable for general cases, later [17] presented a shift symmetric higher-order power method (SS-HOPM), which is shown in the following. In SS-HOPM, a “constant” term $\alpha(x^\top x)^{m/2}$ is added to $Ax^m$. Denote $g(x) \equiv Ax^m + \alpha(x^\top x)^{m/2}$. In each iteration of SS-HOPM, we compute the gradient of $g(x)$, and project it to $\Omega$. When $A$ is symmetric, one can verify that for any $x \neq 0$,

$$\nabla g(x) = mAx^{m-1} + m\alpha(x^\top x)^{m/2-1}x,$$
$$\nabla^2 g(x) = m(m-1)Ax^{m-2} + m\alpha(x^\top x)^{m/2-1}I + m(m-2)\alpha(x^\top x)^{m/2-2}xx^\top,$$
Require: $x_0 \in \mathbb{R}^n$ with $\|x_0\| = 1$, $\alpha \in \mathbb{R}$. Let $\lambda_0 = A x_0^m$.

1: for $k = 0, 1, \ldots$ do
2:   if $\alpha \geq 0$ then
3:     $\hat{x}_{k+1} \leftarrow A x_{k}^{m-1} + \alpha x_k$  \hspace{1cm} (Assumed convex)
4:   else
5:     $\hat{x}_{k+1} \leftarrow -(A x_{k}^{m-1} + \alpha x_k)$ \hspace{1cm} (Assumed concave)
6:   end if
7: $x_{k+1} \leftarrow \hat{x}_{k+1}/\|\hat{x}_{k+1}\|$
8: $\lambda_{k+1} \leftarrow A x_{k+1}^m$
9: end for

Algorithm 2 Shifted symmetric higher-order power method (SS-HOPM)

Given a symmetric tensor $A \in \mathbb{R}^{[m,n]}$.

Theorem 2.1. \cite{17} Let $A \in \mathbb{R}^{[m,n]}$ be symmetric. For $\alpha > \beta(A)$, the iterates $\{\lambda_k, x_k\}$ produced by Algorithm 2 satisfy the following properties. (a) The sequence $\{\lambda_k\}$ is nondecreasing, and there exists $\lambda_*$ such that $\lambda_k \to \lambda_*$. (b) The sequence $\{x_k\}$ has an accumulation point. (c) For every such accumulation point $x_*$, the pair $(\lambda_*, x_*)$ is an eigenpair of $A$. (d) If $A$ has finitely many real eigenvectors, then there exists $x_*$ such that $x_k \to x_*$.

Theorem 2.1 reveals the convergence of sequence $\{\lambda_k\}$ in the convex case. The sequence $\{x_k\}$ converges globally whenever $A$ has finitely many real eigenvectors. Here we give a simple example of tensor with infinite number of eigenvectors: $A$ is a symmetric 3rd-order 3-dimensional tensor. Let $a_{111} = 1$, and the rest elements of $A$ be 0, then it is clear that any $(0, y, z)$ with $y^2 + z^2 = 1$ is an eigenvector corresponding to 0 eigenvalue, thus $A$ has infinite number of eigenvectors. The convergence property of sequence $\{x_k\}$ remains unsettled, and in this paper we mainly address this problem. For concise, we start upon the results in Theorem
2.1, readers can refer to [17] for more details. Nextly we present further properties of the sequence \( \{\lambda_k, x_k\} \) generated by Algorithm 2, which will be used later. The convex case in Algorithm 2 is focused on, the properties and convergence results for concave case can be easily derived in a similar way.

**Proposition 1.** Let \( \mathcal{A} \in \mathbb{R}^{[m,n]} \) be symmetric and \( \alpha > \beta(\mathcal{A}) \), the sequence \( \{\lambda_k, x_k\} \) generated by Algorithm 2 satisfies the following properties.

1. \( \lambda_k + \alpha > 0 \).
2. \( \|x_{k+1} - x_k\| \to 0 \).
3. \( \|\hat{x}_k\| \to \lambda_* + \alpha \), where \( \lambda_* \) is the limit point of \( \{\lambda_k\} \).

**Proof.** (1). From Algorithm 2, \( \lambda_k = \mathcal{A}x_k^m \). From (8), \( \mathcal{A}x_k^m \geq -\frac{\beta(\mathcal{A})}{m-1} \). Then

\[
\lambda_k + \alpha \geq -\frac{\beta(\mathcal{A})}{m-1} + \alpha > -\beta(\mathcal{A}) + \alpha > 0.
\]

(2). Property \( \|x_{k+1} - x_k\| \to 0 \) comes from the proof of Theorem 2.1 in [17].

(3). From the algorithm we have

\[
\hat{x}_k^\top x_k = \mathcal{A}x_k^m + \alpha\|x_k\|^2 = \lambda_k + \alpha = \|\hat{x}_{k+1}\|^2 k x_k.
\]

For \( \|x_{k+1} - x_k\| \to 0 \), \( \|x_{k+1}\| = \|x_k\| = 1 \), then \( x_{k+1}^\top x_k \to 1 \). There exists a \( \lambda_* \) such that \( \lambda_k \to \lambda_* \) by Theorem 2.1. So \( \|\hat{x}_k\| \to \lambda_* + \alpha \). \( \square \)

**Proposition 2.** Let \( \alpha > \beta(\mathcal{A}) \), the sequence \( \{\lambda_k, x_k\} \) generated by Algorithm 2 satisfies

\[
\lambda_{k+1} - \lambda_k \geq \frac{r}{2}\|x_{k+1} - x_k\|^2,
\]

for sufficiently large \( k \) and some \( r > 0 \).

**Proof.** For \( \|x_{k+1} - x_k\| \to 0 \) (from Proposition 1), there exists a \( K \) such that for all \( k > K \), \( \|x_{k+1} - x_k\| \leq \sqrt{3} \), i.e., \( x_{k+1}^\top x_k \geq -\frac{3}{2} \). Then for any \( \theta \in (0,1) \), it holds that

\[
\|\theta x_k + (1-\theta)x_{k+1}\| = \sqrt{\|\theta x_k + (1-\theta)x_{k+1}\|^2}
\]

\[
= \sqrt{\theta^2 + (1-\theta)^2 + 2\theta(1-\theta)x_{k+1}^\top x_k}
\]

\[
\geq \sqrt{\theta^2 + (1-\theta)^2 - \theta(1-\theta)}
\]

\[
\geq \sqrt{3(\theta - \frac{1}{2})^2 + \frac{1}{4}}
\]

\[
\geq \frac{1}{2}
\]

By Taylor expansion of \( g(x) \) and equality (7), one has

\[
\lambda_{k+1} - \lambda_k = g(x_{k+1}) - g(x_k)
\]

\[
= \langle \nabla g(x_k), x_{k+1} - x_k \rangle
\]

\[
+ \frac{1}{2}(x_{k+1} - x_k)^\top \nabla^2 g(\theta x_k + (1-\theta)x_{k+1})(x_{k+1} - x_k)
\]

\[
\geq \langle \hat{x}_{k+1}^\top x_k, x_{k+1} - x_k \rangle + \frac{1}{2}(m-1)m(\alpha - \beta(\mathcal{A}))\|x_{k+1} - x_k\|^2
\]

\[
\geq \left( \frac{1}{2} \right)^{m-1}m(\alpha - \beta(\mathcal{A}))\|x_{k+1} - x_k\|^2,
\]

for some \( \theta \in (0,1) \) and sufficiently large \( k \). Denote \( r \equiv \left( \frac{1}{2} \right)^{m-1}m(\alpha - \beta(\mathcal{A})) \), then \( r > 0 \). The proof is complete. \( \square \)
Combine the above proposition and the boundedness of \( \{\lambda_k\} \), we can conclude that the sequence \( \{\|x_{k+1} - x_k\|^2\} \) is summable, however the global convergence of \( \{x_k\} \) cannot be inferred from it.

3. The point-wise convergence via Lojasiewicz inequality. In this section we introduce a convergence theorem for iterations with Lojasiewicz inequality and sufficient decrease condition hold, then we build a map from the original constrained optimization to an implicit unconstrained optimization by defining a new sequence and function on the sequence generated by SS-HOPM.

**Theorem 3.1.** [[12], Theorem 3.2] Let \( f : V \to \mathbb{R} \) be a real-analytic function on a finite-dimensional real vector space \( V \), and let \( \{y_k\} \subset \mathbb{R}^n \) be a sequence satisfying
\[
f(y_k) - f(y_{k+1}) \geq \sigma \|\nabla f(y_k)\| \|y_{k+1} - y_k\| \tag{12}
\]
for all large enough \( k \) and some \( \sigma > 0 \). Assume further that the implication
\[
[f(y_{k+1}) = f(y_k)] \implies y_{k+1} = y_k \tag{13}
\]
holds. Then a cluster point \( y_* \) of the sequence \( \{y_k\} \) must be its limit. In particular, if the sequence is bounded, it is convergent.

Inequality (12) is the called the sufficient decrease condition. The Lojasiewicz gradient inequality:
\[
|f(y) - f(y_*)|^{1-\theta} \leq \Lambda \|\nabla f(y)\|, \text{ for some } \Lambda > 0 \text{ and } \theta \in (0, \frac{1}{2}].
\]
which holds for real-analytic functions, plays an important role in proving Theorem 3.1. Further with Theorem 3.2, we may prove the limit point of sequence to be a critical point for certain function.

**Theorem 3.2.** [[1], Theorem 3.2] Under the conditions of Theorem 3.1, assume that there exists \( \eta > 0 \) such that
\[
\|y_{k+1} - y_k\| \geq \eta \|\nabla f(y_k)\| \tag{14}
\]
for sufficiently large \( k \), then \( \nabla f(y_*) = 0 \), and the convergence rate can be estimated as follows:
\[
\|y_* - y_k\| \leq \begin{cases} 
q^k & \text{if } \theta = \frac{1}{2} \text{ (for some } 0 < q < 1), \\
k^{-\frac{\theta}{1-\theta}} & \text{if } 0 < \theta < \frac{1}{2}. 
\end{cases} \tag{15}
\]

We define a sequence \( \{y_k\} \) related to \( \{\lambda_k, x_k\} \) generated by SS-HOPM. If \( \alpha > \beta(A) \), define
\[
y_k \equiv (\lambda_k + \alpha)\frac{1}{m} x_k. \tag{16}
\]
Clearly, \( y_k \) is well defined from Proposition 1. If \( \alpha < -\beta(A) \), we can prove \( \lambda_k + \alpha < 0 \) and in this case, we define \( y_k \) as
\[
y_k \equiv (-\lambda_k - \alpha)\frac{1}{m} x_k. \tag{17}
\]
If we show that \( \{y_k\} \) is convergent, for \( \lambda_k + \alpha \) converges to \( \lambda_* + \alpha \), \( \lambda_* + \alpha \neq 0 \), then \( \{x_k\} \) must converge. Two functions over \( y \) are defined below.
\[
f_1 \equiv -2(\mathbf{Ay}^m + \alpha(y^\top y)^{\frac{m}{2}}) + (y^\top y)^m \tag{18}
\]
for (16), and
\[
f_2 \equiv 2(\mathbf{Ay}^m + \alpha(y^\top y)^{\frac{m}{2}}) + (y^\top y)^m \tag{19}
\]
for (17).
Here we give some interpretation of defining such functions. \( \| \bar{A} - \lambda x \circ x \circ \cdots \circ x \| \)
is a measure for symmetric rank-1 approximation of \( \bar{A} \) with constraint \( \|x\| = 1 \), where \( x \circ x \circ \cdots \circ x \) denotes the \( m \)-th order \( n \)-dimensional tensor whose \((i_1, \ldots, i_m)\) element is \( x_{i_1} \cdots x_{i_m} \), for \( i_1, \ldots, i_m = 1, \ldots, n \). \( \| \bar{A} \pm y \circ y \circ \cdots \circ y \|_F^2 \) can be regarded as a measure for unconstrained-type symmetric rank-1 approximation of \( \bar{A} \) in some sense. \( \| \bar{A} - y \circ y \circ \cdots \circ y \|_F^2 \) corresponds to the situation \( \lambda \geq 0 \) and \( \lambda \) is absorbed into the factor \( y \). \( \| \bar{A} + y \circ y \circ \cdots \circ y \|_F^2 \) corresponds to the situation \( \lambda < 0 \), and \( -\lambda \) is absorbed into the factor \( y \).

\[
\| \bar{A} \pm y \circ y \circ \cdots \circ y \|_F^2 = \| \bar{A} \|_F^2 + 2\bar{A}y^m + (y^\top y)^m. \tag{20}
\]

Replacing \( \bar{A}y^m \) in (20) by \( A(y^m + \alpha (y^\top y)^{\frac{m}{2}}) \), we get the functions defined in (18) and (19). Substitute \( y_k \) into function \( f_1 \) and \( f_2 \), we have

\[
f_1(y_k) = -2((\lambda_k + \alpha)Ax_k^m + \alpha(\lambda_k + \alpha)) + (\lambda_k + \alpha)^2 = -2(\lambda_k^2 + 2\alpha\lambda_k + \alpha^2) + (\lambda_k + \alpha)^2 \tag{21}
\]

\[
f_2(y_k) = 2((-\lambda_k - \alpha)Ax_k^m + \alpha(-\lambda_k - \alpha)) + (-\lambda_k - \alpha)^2 = -2(\lambda_k^2 + 2\alpha\lambda_k + \alpha^2) + (\lambda_k + \alpha)^2 \tag{22}
\]

We see that the sequence \( \{y_k\} \) correspondent products a monotonic decrease in \( \mp 2(Ay^m + \alpha (y^\top y)^{\frac{m}{2}}) + (y^\top y)^m \), which is required for (12). Thus the sequence \( \{y_k\} \) and function \( f \) form a monotonic decreasing iteration, under which an unconstrained optimization hides. The transform strategy from constrained optimization to unconstrained optimization is motivated by [1], and the unconstrained optimization here is implicit.

In the last part of this section, we show the implication condition in Theorem 3.1 is satisfied for \( \{y_k\} \) and the function defined.

**Proposition 3.** Assume \( y_k \) is defined as (16) and \( f \) as (18) or \( y_k \) as (17) and \( f \) as (19), then the implication

\[
[f(y_k) = f(y_{k+1})] \implies y_k = y_{k+1}
\]

holds.

**Proof.** Assume \( \alpha > \beta(A) \), from Proposition 1, \( \lambda_k + a > 0 \), then \( [f(y_k) = f(y_{k+1})] \) implies \( \lambda_k = \lambda_{k+1} \). Then from Proposition 2, we have \( x_k = x_{k+1} \), i.e., \( y_k = y_{k+1} \). For the case of \( \alpha < -\beta(A) \), it can be proved with the same way. \( \square \)

4. **Point-wise convergence of SS-HOPM.** In this section, we will first explore some properties on \( \{y_k\} \), and then apply Theorem 3.1 to establish the convergence of \( \{y_k\} \), which shows the point-wise convergence of SS-HOPM.

**Proposition 4.** Let \( \alpha > \beta(A), \{\lambda_k, x^k\} \) be the sequence generated by Algorithm 2, \( \{y_k\} \) be defined by (16), then \( \{y_k\} \) is bounded above.
Proof. Since \( \lambda_k = Ax_k^n \) and \( \|x_k\| = 1 \), then there exists an \( M > 0 \) such that \( |\lambda_k| < M \). For \( y_k = (\lambda_k + \alpha)^\frac{1}{m} x_k \), we have \( \|y_k\| \leq (M + \alpha)^\frac{1}{m} \), so \( \{y_k\} \) is bounded above. \( \square \)

The \( \{y_k\} \) defined by (17) is also bounded above with a similar proof.

**Proposition 5.** Let \( \alpha > \beta(A) \), \( \{\lambda_k, x_k\} \) be the sequence generated by Algorithm 2, \( \{y_k\} \) be defined by (16), then there exists a \( \tau > 0 \) such that

\[
\|y_{k+1} - y_k\| \geq \tau \|x_{k+1} - x_k\|
\]  

for sufficiently large \( k \).

**Proof.**

\[
\|y_{k+1} - y_k\| = \|(\lambda_{k+1} + \alpha)^\frac{1}{m} x_{k+1} - (\lambda_k + \alpha)^\frac{1}{m} x_k\| \\
= (\lambda_{k+1} + \alpha)^\frac{1}{m} \|x_{k+1} - (\frac{\lambda_k + \alpha}{\lambda_{k+1} + \alpha})^\frac{1}{m} x_k\|. 
\]

From Proposition 1, \( \|x_{k+1} - x_k\| \to 0 \), then there exists a \( K \) such that for all \( k \geq K \), \( \|x_{k+1} - x_k\| \leq 1 \). From simple triangle view, we have

\[
\|x_{k+1} - (\frac{\lambda_k + \alpha}{\lambda_{k+1} + \alpha})^\frac{1}{m} x_k\| \geq \frac{\sqrt{3}}{2} \|x_{k+1} - x_k\|. 
\]

So

\[
\|y_{k+1} - y_k\| \geq \frac{\sqrt{3}}{2} (\lambda_{k+1} + \alpha)^\frac{1}{m} \|x_{k+1} - x_k\| \\
\geq \frac{\sqrt{3}}{4} (\lambda_\ast + \alpha)^\frac{1}{m} \|x_{k+1} - x_k\|
\]

for sufficiently large \( k \), where \( \lambda_\ast \) is the limit of \( \{\lambda_k\} \). The second inequality in (26) follows from the monotonic increase of \( \{\lambda_k\} \). Denote \( \tau = \frac{\sqrt{3}}{4} (\lambda_\ast + \alpha)^\frac{1}{m} \), then \( \tau > 0 \).

The proof is complete. \( \square \)

**Proposition 6.** Let \( \alpha > \beta(A) \), \( \{\lambda_k, x_k\} \) be the sequence generated by Algorithm 2, \( \{y_k\} \), \( f \) be defined by (16) and (18) respectively, then

\[
f(y_k) - f(y_{k+1}) \geq \sigma \|y_{k+1} - y_k\|^2
\]  

for sufficiently large \( k \) and some \( \sigma > 0 \).

**Proof.** From the definitions of \( y_k \) and \( f \), we have

\[
f(y_k) - f(y_{k+1}) = (\lambda_{k+1} + \alpha)^2 - (\lambda_k + \alpha)^2 = (\lambda_{k+1} + \lambda_k + 2\alpha)(\lambda_{k+1} - \lambda_k) \\
\geq (\lambda_\ast + \alpha)(\lambda_{k+1} - \lambda_k)
\]

for all sufficiently large \( k \), where \( \lambda_\ast \) is the limit of \( \{\lambda_k\} \). From Proposition 2, there exists an \( r > 0 \) such that

\[
\lambda_{k+1} - \lambda_k \geq \frac{r}{2} \|x_{k+1} - x_k\|^2 
\]

for all sufficiently large \( k \).

\[
\|y_{k+1} - y_k\| = \|(\lambda_{k+1} + \alpha)^\frac{1}{m} x_{k+1} - (\lambda_k + \alpha)^\frac{1}{m} x_k\| \\
= (\lambda_{k+1} + \alpha)^\frac{1}{m} \|x_{k+1} - (\frac{\lambda_k + \alpha}{\lambda_{k+1} + \alpha})^\frac{1}{m} x_k\| \\
\leq (\lambda_\ast + \alpha)^\frac{1}{m} \|x_{k+1} - (\frac{\lambda_k + \alpha}{\lambda_{k+1} + \alpha})^\frac{1}{m} x_k\|. 
\]
From (7), there exists an $R > 0$ such that $\rho(\nabla^2 g(x)) \leq R$ for $x \in \{ x \mid \frac{1}{2} \leq \| x \| \leq 1 \}$. Since $\| x_{k+1} - x_k \| \to 0$, the line segment generated by $x_{k+1}$, $x_k$ is guaranteed to lie in $\{ x \mid \frac{1}{2} \leq \| x \| \leq 1 \}$ for sufficiently large $k$. Then

$$
\begin{align*}
\lambda_{k+1} - \lambda_k &\leq \langle \nabla g(x_k), x_{k+1} - x_k \rangle + \frac{R}{2} \| x_{k+1} - x_k \|^2 \\
&\leq (\| \nabla g(x_k) \| + \frac{R}{2} \| x_{k+1} - x_k \|) \| x_{k+1} - x_k \| \\
&\leq M \| x_{k+1} - x_k \|
\end{align*}
$$

(31)

for a sufficiently large $M$, where the last inequality comes from the boundedness of $\{ x_k \}$ and the boundedness of $\nabla g(x)$ over $\{ x \mid \frac{1}{2} \leq \| x \| \leq 1 \}$. It is easy to verify that $0 \leq 1 - (\frac{\lambda_k + \alpha}{\lambda_{k+1} + \alpha})^\frac{1}{\lambda} \leq 1 - \frac{\lambda_k + \alpha}{\lambda_{k+1} + \alpha}$, thus we have

$$
\begin{align*}
\| x_{k+1} - \left( \frac{\lambda_k + \alpha}{\lambda_{k+1} + \alpha} \right)^\frac{1}{\lambda} x_k \| &\leq \| x_{k+1} - x_k \| + \| x_k (1 - (\frac{\lambda_k + \alpha}{\lambda_{k+1} + \alpha})^\frac{1}{\lambda})) \| \\
&\leq \| x_{k+1} - x_k \| + 1 - \frac{\lambda_k + \alpha}{\lambda_{k+1} + \alpha} \\
&\leq (1 + \frac{M}{\lambda_{k+1} + \alpha}) \| x_{k+1} - x_k \| \\
&\leq (1 + \frac{2M}{\lambda_k + \alpha}) \| x_{k+1} - x_k \|
\end{align*}
$$

(32)

for all sufficiently large $k$. So

$$
\| y_{k+1} - y_k \| \leq (\lambda_k + \alpha)^\frac{1}{\lambda} (1 + \frac{2M}{\lambda_k + \alpha}) \| x_{k+1} - x_k \|. 
$$

(33)

Combine inequalities (28), (29), and (33), we have

$$
\begin{align*}
f(y_k) - f(y_{k+1}) &\geq (\lambda_k + \alpha) (\lambda_{k+1} - \lambda_k) \\
&\geq \frac{(\lambda_k + \alpha)^r}{2} \| x_{k+1} - x_k \|^2 \\
&\geq \frac{(\lambda_k + \alpha)^r}{2((\lambda_k + \alpha)^\frac{1}{\lambda} (1 + \frac{2M}{\lambda_k + \alpha}))^2} \| y_{k+1} - y_k \|^2.
\end{align*}
$$

(34)

Denote $\sigma \equiv \frac{(\lambda_k + \alpha)^r}{2((\lambda_k + \alpha)^\frac{1}{\lambda} (1 + \frac{2M}{\lambda_k + \alpha}))^2}$, then $\sigma > 0$. The proof is complete. 

\begin{proposition}
Let $\alpha > \beta(A)$, $\{ \lambda_k, x_k \}$ be the sequence generated by Algorithm 2, $\{ y_k \}$, $f$ be defined by (16) and (18) respectively, then there exists a $\eta > 0$ such that $\| y_{k+1} - y_k \| \geq \eta \| \nabla f(y_k) \|$ for all sufficiently large $k$.
\end{proposition}

\begin{proof}
\begin{align*}
\nabla f(y_k) &= -2m(\mathcal{A} y_k^{m-1} + \alpha (y_k^T y_k)^\frac{1}{2} - y_k + 2m(y_k^T y_k)^{m-1} y_k \\
&= -2m((\lambda_k + \alpha)^\frac{m-1}{m} \mathcal{A} y_k^{m-1} + \alpha (\lambda_k + \alpha)^\frac{m-1}{m} x_k) + 2m(\lambda_k + \alpha)^\frac{2m-1}{m} x_k \\
&= -2m(\lambda_k + \alpha)^\frac{m-1}{m} (\mathcal{A} y_k^{m-1} - \lambda_k x_k).
\end{align*}

(35)

\end{proof}
For
\[ \alpha + \lambda_k = \| \hat{x}_{k+1} x_{k+1} x_k, \]
one has
\[ (x_{k+1} - \alpha + \lambda_k) x_k = x_{k+1} x_k - \alpha + \lambda_k = 0, \]
thus
\[ \| x_{k+1} - \alpha + \lambda_k x_k \| \leq \| x_{k+1} - x_k \|. \]

From Proposition 5, there exists a \( \tau > 0 \) such that
\[ \| y_{k+1} - y_k \| \geq \tau \| x_{k+1} - x_k \| \]
for all sufficiently large \( k \). Combine inequalities (35), (39), and (40), we have
\[ \| y_{k+1} - y_k \| \geq \tau \| x_{k+1} - x_k \| \]
\[ \geq \tau \| x_{k+1} - \alpha + \lambda_k x_k \| \]
\[ \geq \frac{\tau}{2m} \| \hat{x}_{k+1} \| (\lambda_k + \alpha) \frac{m-1}{m} \| \nabla f(y_k) \| \]
\[ \geq \frac{\tau}{4m(\lambda_k + \alpha)} \frac{2m-1}{m} \| \nabla f(y_k) \| \]
for all sufficiently large \( k \). The last inequality in (41) follows from Proposition 1 and the monotonic increase of \( \lambda_k + \alpha \). Denote \( \eta = \frac{\tau}{4m(\lambda_k + \alpha)^{2m-1}} \), then \( \eta > 0 \). The proof is complete.

**Theorem 4.1.** Let \( \alpha > \beta(A) \), \( \{ \lambda_k, x_k \} \) be the sequence generated by Algorithm 2, \( \lambda_* \) be the limit of \( \{ \lambda_k \} \), \( \{ y_k \} \), \( f \) be defined by (16) and (18) respectively, then \( \{ y_k \} \) converges to a point \( y_* \) with \( \nabla f(y_*) = 0 \). The convergence rate estimation is in (17) in term of the exponent in the Lojasiewicz gradient inequality at \( y_* \). The sequence \( \{ x_k \} \) converges to \( \frac{y_*}{(\lambda_k + \alpha)^m} \).

**Proof.** From Proposition 6 and 7, there exist \( \sigma > 0, \eta > 0 \) such that
\[ f(y_k) - f(y_{k+1}) \geq \sigma \| \nabla f(y_k) \| \| y_{k+1} - y_k \| \]
for all sufficiently large \( k \). The implication
\[ [ f(y_{k+1} - f(y_k)) \implies y_{k+1} = y_k ] \]
is proved in Proposition 3. So the conditions of Theorem 3.1 hold. By the boundedness of \( \{ y_k \} \), we conclude that \( \{ y_k \} \) converges to some point \( y^* \). Together with Theorem 3.2, we get that \( y_* \) is a critical point of \( f(y) \). The convergence rate estimation comes from the result in Theorem 3.1. From the relation between \( y_k \) and \( x_k \), we can conclude that \( \{ x_k \} \) converges to point \( x_* \) with \( x_* = \frac{y_*}{(\lambda_* + \alpha)^m} \). \( \square \)

When \( \alpha < -\beta(A) \), the properties and conclusions in this section can be derived in a parallel way. Here we only list the final results of the concave case.

**Theorem 4.2.** Let \( \alpha < -\beta(A) \), \( \{ \lambda_k, x_k \} \) be generated by Algorithm 2, \( \lambda_* \) be the limit of \( \{ \lambda_k \} \), \( \{ y_k \} \), \( f \) be defined by (17) and (19) respectively, then \( \{ y_k \} \) converges to a point \( y_* \) with \( \nabla f(y_*) = 0 \). The convergence rate estimation is in (17) in term of the exponent in the Lojasiewicz gradient inequality at \( y_* \). The sequence \( \{ x_k \} \) converges to \( \frac{y_*}{(-\lambda_* - \alpha)^m} \).
Theorems 4.1 and 4.2 show the point-wise convergence of SS-HOPM, and we note that the strategies used in this paper may be applied to proving the point-wise convergence of some algorithms in which shift tricks were used, for example, [18, 2, 13].

**Example 4.7.** At the end of this section, we give a simple example to show the relation between the sequences \( \{x_k\} \) and \( \{y_k\} \). The example is from [4](Example 1): Let \( A \in \mathbb{R}^{[4,3]} \) be the symmetric tensor defined by

\[
\begin{align*}
a_{1111} &= 0.2883, & a_{1112} &= -0.0031, & a_{1113} &= 0.1973, & a_{1122} &= -0.2485, \\
a_{1123} &= -0.2939, & a_{1133} &= 0.3847, & a_{1222} &= 0.2972, & a_{1223} &= 0.1862, \\
a_{1222} &= 0.0919, & a_{1333} &= -0.3619, & a_{2222} &= 0.1241, & a_{2223} &= -0.3420, \\
a_{2233} &= 0.2127, & a_{2333} &= 0.2727, & a_{3333} &= -0.3054.
\end{align*}
\]

The trajectories of two sequences are plotted in Figure 1 for two trials with shift parameter \( \alpha \) set to be 1 and -1 respectively. Every element in sequence \( \{x_k\} \) is on the unit sphere, and \( y_k \) is computed by scaling \( x_k \) by \( (Ax^m + a)^\frac{1}{m} \) or \( (-Ax^m - a)^\frac{1}{m} \).

![Figure 1](image)

**Figure 1.** Trajectories of sequences \( \{x_k\}, \{y_k\} \) generated by SS-HOPM

5. **Conclusion.** Instead of proving the point-wise convergence of SS-HOPM directly, in this paper we focused on the convergence of a sequence closed related to SS-HOPM. The problem SS-HOPM to solve is a constrained optimization problem, and we connected it to an implicit unconstrained optimization problem in some sense. The strong convexity or concavity hidden in SS-HOPM makes it possible to use the well established results based on Lojasiewicz inequality to prove the convergence of the new sequence, from which the point-wise convergence of SS-HOPM becomes clear. For the similarity of many tensor related algorithms, we believe that an enhancement of their conclusions about the related algorithms can be achieved by making use of the approach presented in our paper.

**Acknowledgments.** We are grateful to the two anonymous referees for insightful comments that greatly improved the manuscript.
REFERENCES

[1] A. Uschmajew, A new convergence proof for the higher-order power method and generalizations, *Pac. J. Optim.*, 11 (2015), 309–321.

[2] A. T. Erdogan, On the convergence of ICA algorithms with symmetric orthogonalization, *IEEE Trans. Signal Process.*, 57 (2009), 2209–2221.

[3] D. Cartwright and B. Sturmfels, The number of eigenvalues of a tensor, *Linear Alg. Appl.*, 438 (2013), 942–952.

[4] E. Kofidis and P. A. Regalia, On the best rank-1 approximation of higher-order supersymmetric tensors, *SIAM J. Matrix Anal. Appl.*, 23 (2001), 863–884.

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Fourth edition, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 2013.

[6] 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), 1324–1342.

[7] L. Lim, Singular values and eigenvalues of tensors: A variational approach, 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005., IEEE, (2005), 129–132.

[8] L. Q. Qi, Eigenvalues of a real supersymmetric tensor, *J. Symb. Comput.*, 40 (2005), 1302–1324.

[9] L. Q. Qi and K. L. Teo, Multivariate polynomial minimization and its application in signal processing, *J. Glob. Optim.*, 26 (2013), 419–433.

[10] L. Q. Qi, W. Y. Sun and Y. J. Wang, Numerical multilinear algebra and its applications, *Front. Math. China.*, 2 (2007), 501–526.

[11] M. Ng, L. Q. Qi and G. L. Zhou, Finding the largest eigenvalue of a nonnegative tensor, *SIAM J. Matrix Anal. Appl.*, 31 (2009), 1090–1099.

[12] P.-A. Absil, R. Mahony and B. Andrews, Convergence of the iterates of descent methods for analytic cost functions, *SIAM J. Optim.*, 16 (2005), 531–547.

[13] P. A. Regalia and E. Kofidis, Monotonic convergence of fixed-point algorithms for ICA, *IEEE Trans. Neural Netw.*, 14 (2003), 943–949.

[14] Q. Ni, L. Q. Qi and F. Wang, An eigenvalue method for testing positive definiteness of a multivariate form, *IEEE Trans. Automat. Contr.*, 53 (2008), 1096–1107.

[15] R. Schneider and A. Uschmajew, Convergence results for projected line-search methods on varieties of low-rank matrices via Lojasiewicz inequality, *SIAM J. Optim.*, 25 (2015), 622–646.

[16] S. Lojasiewicz, Ensembles semi-analytiques, *Lectures Notes, IHES Bures-sur-Yvette,* (1965).

[17] T. G. Kolda and J. R. Mayo, Shlifted power method for computing tensor eigenpairs, *SIAM J. Matrix Anal. Appl.*, 32 (2011), 1095–1124.

[18] Y. J. Wang, L. Q. Qi and X. Z. Zhang, A practical method for computing the largest $m$-eigenvalue of a fourth-order partially symmetric tensor, *Numer. Linear Algebra Appl.*, 16 (2009), 589–601.

[19] Y. Y. Xu and W. T. Yin, A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion, *SIAM J. Imaging Sci.*, 6 (2013), 1758–1789.

Received January 2019; revised June 2019.

E-mail address: luogangnk@gmail.com
E-mail address: qz-yang@nankai.edu.cn