mathematics

MDPI

Review

A Very Brief Introduction to Nonnegative Tensors from the Geometric Viewpoint

Yang Qi

Department of Mathematics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, USA; yangqi@math.uchicago.edu

Received: 27 September 2018; Accepted: 24 October 2018; Published: 30 October 2018

Abstract: This note is a short survey of nonnegative tensors, primarily from the geometric point of view. In addition to basic definitions, we discuss properties of and questions about nonnegative tensors, which may be of interest to geometers.

Keywords: nonnegative tensors; low-rank approximations; uniqueness and identifiability; spectral theory; EM algorithm; semialgebraic geometry

1. Introduction

Tensors are ubiquitous in mathematics and sciences. In the study of complex and real tensors, algebraic geometry has demonstrated its power [1,2]. On the other hand, tensor computations also help people understand classical algebraic varieties, such as the secant varieties of Segre varieties and Veronese varieties, and raise interesting and challenging questions in algebraic geometry [3,4]. Traditionally, geometers tend to study tensors in a coordinate-free way. However, in applications, practitioners must work with coordinates. Among those tensors widely used in practice, a large number of them are nonnegative tensors, i.e., tensors with nonnegative entries. In this case, most powerful geometric tools developed for complex tensors can not be applied directly due to the fact that the Euclidean closure of tensors with rank no greater than a fixed integer is no longer a variety, but a semialgebraic set. This forces us to investigate the semialgebraic geometry of nonnegative tensors. In this note, we will review some important properties of nonnegative tensors obtained by studying the semialgebraic geometry, and propose several open problems which are pivotal in understanding nonnegative tensors and also may be interesting to geometers.

2. Definitions

Nonnegative tensors arise naturally in many areas, such as hyperspectral imaging, statistics, spectroscopy, computer vision, phylogenetics, and so on. See [5–8] and the references therein. Before further investigations, let us recall basic definitions of tensors.

Definition 1. *Let $V_1, \ldots, V_d$ be vector spaces over a field $\mathbb{K}$. The tensor product $\mathbb{V} = V_1 \otimes \cdots \otimes V_d$ is the free linear space spanned by $V_1 \times \cdots \times V_d$ quotient by the equivalence relation:*

$$(v_1, \ldots, \alpha v_i + \beta v_i', \ldots, v_d) \sim \alpha(v_1, \ldots, v_i, \ldots, v_d) + \beta(v_1, \ldots, v_i', \ldots, v_d) \tag{1}$$

for every $v_i, v_i' \in V_i$, $\alpha_i, \beta_i \in \mathbb{K}$, and $i = 1, \ldots, d$. An element of $V_1 \otimes \cdots \otimes V_d$ is called a tensor.

Equivalently, $V_1 \otimes \cdots \otimes V_d$ is the vector space of multilinear functions:

$$V_1^* \times \cdots \times V_d^* \to \mathbb{K},$$

where V_i^* is the dual space of V_i for $i = 1, \ldots, d$. A representative of the equivalence class of $(v_1, \ldots, v_d)$ is called a *decomposable tensor* and denoted by $v_1 \otimes \cdots \otimes v_d$.

The *rank* of a given tensor $T \in V_1 \otimes \cdots \otimes V_d$ is the minimum integer r such that T is a sum of r decomposable tensors, i.e.,

$$T = \sum_{i=1}^{r} v_{1,i} \otimes \cdots \otimes v_{d,i}, \tag{2}$$

where $v_{j,i} \in V_j$ for $j = 1, \ldots, d$ and $i = 1, \ldots, r$. Such a decomposition is called a *rank decomposition* (or canonical polyadic decomposition or CP decomposition).

Now we focus on the case $\mathbb{K} = \mathbb{R}$, and for each V_i we fix a basis, which enables us to work with coordinates. Let $\mathbb{R}_+$ be the semiring of nonnegative real numbers. A nonnegative tensor in $V_1 \otimes \cdots \otimes V_d$ is a tensor whose coordinates are nonnegative. Let V_i^+ denote the set of nonnegative vectors in V_i for each $i = 1, \ldots, d$, and $\mathbb{V}^+$ denote the set of nonnegative tensors in $\mathbb{V}$.

Definition 2. *For $T \in \mathbb{V}^+$, the nonnegative rank of T is the minimum integer r so that there exist nonnegative vectors $v_{i,j} \in V_i^+$ for $i = 1, \ldots, d$ and $j = 1, \ldots, r$ making Equation (2) holds.*

It is clear that $\operatorname{rank}_+(T) \geq \operatorname{rank}(T)$ for every $T \in \mathbb{V}^+$. Besides, there exists some $T \in \mathbb{V}^+$ such that $\operatorname{rank}_+(T) > \operatorname{rank}(T)$. For example, let

$$T = e_1 \otimes e_1 \otimes e_1 + e_1 \otimes e_2 \otimes e_2 + e_2 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_2 \in \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2, \tag{3}$$

where $e_1 = [1, 0]^\top$ and $e_2 = [0, 1]^\top$, then $\operatorname{rank}_+(T) = 4 > 2 = \operatorname{rank}_{\mathbb{R}}(T)$.

3. Applications

One reason that nonnegative tensors are popular is due to the statistical interpretation behind—a Bayesian network [9–11]. More precisely, assume a joint distribution of several random variables x_i is given by:

$$p(x_1, \ldots, x_d) = \int \prod_{i=1}^{d} p(x_i \mid \theta)\, d\mu_\theta \tag{4}$$

where θ is a latent variable. When $x_1, \ldots, x_d$ and θ are discrete, (4) becomes:

$$t_{i_1,\ldots,i_d} = \sum_{p=1}^{r} \lambda_r u_{i_1,p} \cdots u_{i_d,p}, \tag{5}$$

i.e., a nonnegative rank decomposition [5,12]. Such a model, for instance, has been applied in clustering [13].

As another application, nonnegative tensors have shown their powers in image processing. Usually hyperspectral images are processed as nonnegative matrices $M \in \mathbb{R}_+^{n \times m}$, where n is the number of pixels and m denotes the number of spectral bands. By the sensor developments, it is possible to collect time series of hyperspectral data, which can be understood as nonnegative tensors, namely $A \in \mathbb{R}_+^{n \times m \times d}$, where d is the dimensionality of the time or multiangle ways [14]. A nonnegative rank decomposition of A gives a blind spectral unmixing of hyperspectral data.

Recently, tensor methods have also used in isogeometric analysis (IGA) [15–17]. A Galerkin-based approach of IGA studies tensor-product B-splines. To obtain Galerkin matrices effectively, low rank approximations of integral kernels are employed [16]. In many cases, the constructed mass tensor is a positive tensor.

4. Algorithms

Due to the broad and important applications, nonnegative tensor decomposition (NTD) and nonnegative matrix factorization (NMF) have received vast research on their algorithms. Perhaps the

most popular algorithm for NMF is the multiplicative updating rule [18], and since then, numerous more efficient algorithms have been proposed, such as the algorithms using the alternating nonnegative least squares [19–21], the algorithms using the hierarchical alternating least squares [22], the algorithms using deflation [23], and many more algorithms based on other methods, for example [24,25], etc. The main ideas of some algorithms have been naturally generalized to decompose nonnegative tensors, for example [23,26–28]. Since the main purpose of this note is to introduce the geometric properties of nonnegative tensors, we invite those readers who are interested in algorithms to read comprehensive surveys on algorithms of NTD, for example [29–31].

5. Nonnegative Rank Decompositions

It is known that when $d > 2$, rank decompositions (2) are often unique over $\mathbb{C}$ and $\mathbb{R}$, which is very important in applications. For nonnegative tensors, it is also an important issue to investigate the identifiability property. Before studying the identifiability of nonnegative tensors, let us recall fundamental definitions and known results of complex and real tensors.

For any tuple of positive integers $(n_1, \ldots, n_d)$, there is a unique integer $r_g(n_1, \ldots, n_d)$, which only depends on $n_1, \ldots, n_d$ such that the set of complex rank-$r_g(n_1, \ldots, n_d)$ tensors in $\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$ contains a Zariski open subset of $\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$. In fact, $r_g(n_1, \ldots, n_d)$ is the minimum integer r such that the rth secant variety of the Segre variety $\mathrm{Seg}(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1})$ is the ambient space $\mathbb{P}^{n_1 \cdots n_d - 1}$. $r_g(n_1, \ldots, n_d)$ is called the *generic rank* of $\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$. It is not always the case that the generic rank $r_g(n_1, \ldots, n_d)$ equals $\lceil \frac{n_1 \cdots n_d}{n_1 + \cdots + n_d - d + 1} \rceil$, which leads us to the following definition.

Definition 3. *If the $\mathbb{K}$-dimension of the set of rank-r tensors in $\mathbb{K}^{n_1} \otimes \cdots \otimes \mathbb{K}^{n_d}$ is strictly less than* $\min\{r(n_1 + \cdots + n_d - d + 1), n_1 \cdots n_d\}$*, then $\mathbb{K}^{n_1} \otimes \cdots \otimes \mathbb{K}^{n_d}$ is called* r-defective.

When $\mathbb{K}$ is algebraically closed, $\mathbb{K}^{n_1} \otimes \cdots \otimes \mathbb{K}^{n_d}$ is not r-defective, which implies a general rank-r tensor T has finitely many rank decompositions. If we require further that T has a unique decomposition, we will arrive at the following definition.

Definition 4. *If a general rank-r tensor in $\mathbb{K}^{n_1} \otimes \cdots \otimes \mathbb{K}^{n_d}$ has a unique rank-r decomposition over $\mathbb{K}$, then $\mathbb{K}^{n_1} \otimes \cdots \otimes \mathbb{K}^{n_d}$ is called* r-identifiable.

There has been a large amount of research on defectivity [32–34] and identifiability [1,35–41]. Here, we highlight three notable results.

Theorem 1 (Kruskal)**.** *Let $V_1, \ldots, V_d$ be finite dimensional vector spaces over a field $\mathbb{K}$ [35], and*

$$T = \sum_{i=1}^{r} v_{1,i} \otimes \cdots \otimes v_{d,i} \in V_1 \otimes \cdots \otimes V_d. \tag{6}$$

If $\kappa_1 + \cdots + \kappa_d \geq 2r + d - 1$, then $\mathrm{rank}(T) = r$ *and T has a unique rank-r decomposition (6), where κ_i is the maximum integer such that every subset of $\{v_{i,1}, \ldots, v_{i,r}\}$ with κ_i elements is linearly independent for $i = 1, \ldots, d$.*

Theorem 2 (Bocci–Chiantini–Ottaviani [1])**.** *Assume $n_1 \leq \cdots \leq n_d$. Then $\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$ is r-identifiable when:*

$$r \leq \frac{\prod_{j=1}^{d} n_j - (n_1 + n_2 + n_3 - 2) \prod_{j=3}^{d} n_j}{1 + \sum_{j=1}^{d} (n_j - 1)}.$$

Theorem 3 (Chiantini–Ottaviani–Vannieuwenhoven). *$\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$ is r-identifiable when [40]:*

$$r < \left\lceil \frac{\prod_{j=1}^{d} n_j}{1 + \sum_{j=1}^{d}(n_j - 1)} \right\rceil$$

and $\prod_{j=1}^{d} n_j \leq 15000$, except the following cases:

$(n_1, \ldots, n_d)$	r	*Type*
$(4,4,3)$	5	*defective*
$(4,4,4)$	6	*sporadic*
$(6,6,3)$	8	*sporadic*
$(n,n,2,2)$	$2n-1$	*defective*
$(2,2,2,2,2)$	5	*sporadic*
$n_1 > \prod_{i=2}^{d} n_i - \sum_{i=2}^{d}(n_i - 1)$	$r \geq \prod_{i=2}^{d} n_i - \sum_{i=2}^{d}(n_i - 1)$	*unbalanced*

Note that Theorems 2 and 3 focus on complex tensors; however, with the help of the following lemma, we are able to extend these results to real tensors.

Lemma 1. *If $\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$ is r-identifiable, then $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ is r-identifiable when $r < r_g(n_1, \ldots, n_d)$ [42].*

As an example, we have the following corollary.

Corollary 1. *$\mathbb{R}^{n_1 \times \cdots \times n_d}$ is r-identifiable if:*

$$r < \left\lceil \frac{\prod_{i=1}^{d} n_i}{1 + \sum_{i=1}^{d}(n_i - 1)} \right\rceil,$$

$\prod_{i=1}^{d} n_i \leq 15000$, and $(n_1, \ldots, n_d, r)$ is not one of the following cases:

$(n_1, \ldots, n_d)$	r
$(4,4,3)$	5
$(4,4,4)$	6
$(6,6,3)$	8
$(n,n,2,2)$	$2n-1$
$(2,2,2,2,2)$	5
$n_1 > \prod_{i=2}^{d} n_i - \sum_{i=2}^{d}(n_i - 1)$	$r \geq \prod_{i=2}^{d} n_i - \sum_{i=2}^{d}(n_i - 1)$

In fact, for the above exceptional cases, we can derive more information from Theorem 3.

Corollary 2.

- *$\mathbb{R}^{4\times4\times3}$ is 5-defective.*
- *For any $n \geq 2$, $\mathbb{R}^{n\times n\times 2\times 2}$ is $(2n-1)$-defective.*
- *For $n_1 \geq \cdots \geq n_d \geq 2$, $\mathbb{R}^{n_1 \times \cdots \times n_d}$ is r-defective if*

$$n_1 > \prod_{i=2}^{d} n_i - \sum_{i=2}^{d}(n_i - 1) \quad \textit{and} \quad r \geq \prod_{i=2}^{d} n_i - \sum_{i=2}^{d}(n_i - 1).$$

Recall that for a symmetric tensor $T \in \mathsf{S}^d(V)$ over $\mathbb{K}$, the *symmetric rank* of T is the minimum integer r such that

$$T = \sum_{i=1}^{r} \lambda_i v_i^{\otimes d},$$

where $\lambda_i \in \mathbb{K}, v_i \in V$ for $i = 1, \ldots, r$. For symmetric tensors, we can also have the definitions of generic rank, r-defectivity, and r-identifiability. More precisely, the generic symmetric rank, $r_g(n; d)$,

is defined to be the minimum integer r such that the rth secant variety of the Veronese variety $\nu_d(\mathbb{P}^{n-1})$ fills in the ambient space over $\mathbb{C}$. If the $\mathbb{K}$-dimension of the set of symmetric rank-r tensors in $\mathsf{S}^d(\mathbb{K}^n)$ is strictly less than $\min\{rn, \binom{n+d-1}{d}\}$, then $\mathsf{S}^d(\mathbb{K}^n)$ is called r-defective. If a general symmetric rank-r tensor has a unique decomposition over $\mathbb{K}$, $\mathsf{S}^d(\mathbb{K}^n)$ is called r-identifiable. The defectivity problem has been completely solved in [43].

Theorem 4 (Alexander–Hirschowitz)**.** *The generic rank [43]:*

$$r_g(n;d) = \left\lceil \frac{\binom{n+d-1}{d}}{n} \right\rceil$$

except the following case:

- *When $d = 2$, $r_g(n;d) = n$.*
- *When $(d,n) = (3,5), (4,3), (4,4), (4,5)$, $r_g(n;d) = \lceil \frac{\binom{n+d-1}{d}}{n} \rceil + 1$.*

The identifiability problem of $\mathsf{S}^d(\mathbb{C}^n)$ has been addressed in [2,44–46], and the complete solution was given in [2].

Theorem 5 (Chiantini–Ottaviani–Vannieuwenhoven)**.** *$\mathsf{S}^d(\mathbb{C}^{n+1})$ is r-identifiable when [2]*

$$r < \left\lceil \frac{\binom{n+d}{d}}{n+1} \right\rceil$$

and $d \geq 3$, except that $(d,n,r) \in \{(6,2,9),(4,3,8),(3,5,9)\}$ where a general complex symmetric rank-r tensor has exactly two symmetric rank decompositions.

Similar to Lemma 1, we have the following lemma for real symmetric tensors.

Lemma 2. *Let $r < r_g(n;d)$. If $\mathsf{S}^d(\mathbb{C}^n)$ is r-identifiable, then $\mathsf{S}^d(\mathbb{R}^n)$ is r-identifiable [42].*

As an example, we have

Corollary 3. *$\mathsf{S}^d(\mathbb{R}^{n+1})$ is r-identifiable when:*

$$r < \left\lceil \frac{\binom{n+d}{d}}{n+1} \right\rceil$$

and if $(d,n,r) \notin \{(6,2,9),(4,3,8),(3,5,9)\}$.

Now, we are in a position to study the relations among complex, real, and nonnegative ranks. Given real vector spaces $V_1, \ldots, V_d$ of dimensions $n_1, \ldots, n_d$, respectively, let $\mathbb{V} := V_1 \otimes \cdots \otimes V_d$ and $\mathbb{V}_{\mathbb{C}}$ be the complexification of $\mathbb{V}$. For any positive integer r, let

$$D_r^+ = \{X \in \mathbb{V}^+ \mid \mathrm{rank}_+(X) \leq r\}$$

denote the set of nonnegative tensors with nonnegative ranks not greater than r.

Theorem 6. *Let $r < r_g(n_1, \ldots, n_d)$. For a general $T \in D_r^+$, its real rank and complex rank are also r. If $\mathbb{V}_{\mathbb{C}}$ is r-identifiable, then T has a unique rank-r decomposition, which is nonnegative [42].*

For nonnegative tensors, when $r \geq r_g(n_1, \ldots, n_d)$, the set of nonnegative rank-r tensors may contain a nonempty open subset of $\mathbb{V}^+$ under the Euclidean topology. If so, r is called a *nonnegative*

typical rank. A Similar phenomenon happens in the real case which motivates the definition of *real typical rank*, i.e., r is a real typical rank if the set of real rank-r tensors contains a nonempty open subset of $\mathbb{V}$. We will illustrate the difference between nonnegative ranks and real ranks by the following example, where $\mathbb{R}_+^{n\times n\times n}$ denotes the set of nonnegative tensors in $\mathbb{R}^{n\times n\times n}$.

Proposition 1. *[42]*

- *The nonnegative typical ranks of* $\mathbb{R}_+^{2\times 2\times 2}$ *are* $2, 3, 4$.
- *The nonnegative typical ranks of* $\mathbb{R}_+^{3\times 3\times 3}$ *are all integers* m *satisfying:*

$$5 \leq m \leq 9.$$

- *When* $n \geq 4$, *the nonnegative typical ranks of* $\mathbb{R}_+^{n\times n\times n}$ *consist of all integers* m *satisfying:*

$$\left\lceil \frac{n^3}{3n-2} \right\rceil \leq m \leq n^2.$$

Theorem 6 and Proposition 1 reveal that for a general nonnegative rank-r tensor T, the true difference among its complex, real, and nonnegative ranks appears when $r \geq r_g(n_1, \ldots, n_d)$, namely when $r < r_g(n_1, \ldots, n_d)$, the complex and real ranks of T are also r, but when $r > r_g(n_1, \ldots, n_d)$, D_r^+ contains a nonempty open subset $\mathcal{U}$ of $\mathbb{V}^+$ such that for each $T \in \mathcal{U}$,

$$r_g(n_1, \ldots, n_d) = \mathrm{rank}_{\mathbb{C}}(T) \leq \mathrm{rank}_{\mathbb{R}}(T) < \mathrm{rank}_+(T) = r.$$

More concretely, let T be the tensor defined in (3). Then there exists a nonempty open neighborhood $\mathcal{U}$ of T in $\mathbb{R}_+^{2\times 2\times 2}$ such that for any $A \in \mathcal{U}$,

$$\mathrm{rank}_{\mathbb{C}}(A) = 2 < 4 = \mathrm{rank}_+(A).$$

6. Low Rank Approximations

Given $T \in \mathbb{V}^+$ with $r \leq \mathrm{rank}_+(T)$, let:

$$\delta(T) = \inf_{X \in D_r^+} \|T - X\|,$$

where $\|\cdot\|$ is the *Hilbert–Schmidt norm*.

It is known that the set $D_r = \{X \in V_1 \otimes \cdots \otimes V_d \mid \mathrm{rank}(X) \leq r\}$ is not closed under the Euclidean topology over $\mathbb{R}$ or $\mathbb{C}$ when $r > 1$ [47]. However, for nonnegative tensors, we can show:

Proposition 2. D_r^+ *is a closed semialgebraic set under the Euclidean topology [48].*

Since D_r^+ is closed, for any $T \notin D_r^+$, there is always some $T_0 \in D_r^+$ such that $\|T - T_0\| = \delta(T)$, i.e., the optimization problem:

$$\min_{\mathrm{rank}_+(X) \leq r} \|T - X\| \tag{7}$$

makes sense. Furthermore, we can have the following result.

Proposition 3. *A general* $T \in \mathbb{V}^+$ *has a unique best nonnegative low-rank approximation [48].*

Before studying nonnegative rank approximations, let us recall the following useful lemma.

Lemma 3. *For $T \in \mathbb{V}$ over $\mathbb{R}$, assume $\operatorname{rank}(T) > r$ and $\lambda \sum_{l=1}^{r} T_l$ is a best rank-r approximation, where $T_l = v_{1,l} \otimes \cdots \otimes v_{d,l}$ and $\left\|\sum_{l=1}^{r} T_l\right\| = 1$. Then:*

$$\langle T, v_{1,j} \otimes \cdots \otimes \widehat{v_{i,j}} \otimes \cdots \otimes v_{d,j} \rangle = \lambda \left\langle \sum_{l=1}^{r} T_l, v_{1,j} \otimes \cdots \otimes \widehat{v_{i,j}} \otimes \cdots \otimes v_{d,j} \right\rangle, \tag{8}$$

where $i = 1, \ldots, d$, and $j = 1, \ldots, r$, where $\lambda = \langle T, \sum_{l=1}^{r} T_l \rangle$, and $\langle , \rangle$ denotes tensor contraction.

The *support* of a vector $u \in V$ is defined to be:

$$\operatorname{supp}(u) := \{j = \{1, \ldots, \dim_{\mathbb{R}} V\} \mid \text{ the } j\text{th coordinate of } u \text{ is nonzero}\}.$$

Then for a nonnegative tensor T, Lemma 3 becomes

Lemma 4. *Let $T \in \mathbb{V}^+$ with $\operatorname{rank}_+(T) > r$ and $Y = \sum_{j=1}^{s} v_{1,j} \otimes \cdots \otimes v_{d,j}$ be a solution of (7). Then:*

$$\langle T, v_{1,j} \otimes \cdots \otimes w_{i,j} \otimes \cdots \otimes v_{d,j} \rangle \leq \left\langle Y, v_{1,j} \otimes \cdots \otimes w_{i,j} \otimes \cdots \otimes v_{d,j} \right\rangle \tag{9}$$

where $w_{i,j} \in V_i^+$, $i = 1, \ldots, d$, and $j = 1, \ldots, s$. For every pair (i, j), define:

$$\widetilde{V}_{i,j} := \{v \in V_i : \operatorname{supp}(v) \subseteq \operatorname{supp}(v_{i,j})\}.$$

Then:

$$\langle T, v_{1,j} \otimes \cdots \otimes w_{i,j} \otimes \cdots \otimes v_{d,j} \rangle = \left\langle Y, v_{1,j} \otimes \cdots \otimes w_{i,j} \otimes \cdots \otimes v_{d,j} \right\rangle \tag{10}$$

for $w_{i,j} \in \widetilde{V}_{i,j}$.

Lemma 4 guarantees us the following result.

Proposition 4. *Let $T \in \mathbb{V}^+$ with $\operatorname{rank}_+(T) > r$ and X be a solution of (7). Then $\operatorname{rank}_+(X) = r$ [48].*

Proposition 4 shows that it is indeed appropriate to call a solution of (7) a best nonnegative rank-r approximation.

By Proposition 3, we know a general nonnegative tensor has a unique best nonnegative rank-r approximation. However, it is still unclear if this best approximation has a unique nonnegative rank-r decomposition. Below is an example where we have the uniqueness. On the other hand, the general case is not known yet.

Proposition 5. *Let $r = 2$ or 3 and let $n_1, \ldots, n_d \geq 3$. Then for a general $T \in \mathbb{R}_+^{n_1 \times \cdots \times n_d}$, its unique best nonnegative rank-r approximation has a unique nonnegative rank-r decomposition [42].*

Question 1. *Assume $\mathbb{V}$ is r-identifiable. Given a general $T \in \mathbb{V}^+$, is it true that the unique best nonnegative rank-r approximation of T has a unique nonnegative decomposition?*

7. Spectral Theory

In this section, we start with nonnegative rank-one approximations, which lead us to the spectral theory of nonnegative tensors.

Proposition 6. *Given $T \in \mathbb{V}^+$, let $u_1 \otimes \cdots \otimes u_d \in \mathbb{V}$ be a best real rank-one approximation of T. Then $u_1, \ldots, u_d$ can be chosen in the form $u_1 \in V_1^+, \ldots, u_d \in V_d^+$.*

By Proposition 6, for a nonnegative tensor T, we will not distinguish a best real rank-one approximation and a best nonnegative rank-one approximation. By Lemma 3, a best real rank-one approximation of a real tensor is a solution of (8), which motivates us the following definition.

Definition 5. *Let $V_1, \ldots, V_d$ be vector spaces over $\mathbb{K}$ of dimensions $n_1, \ldots, n_d$. For $T \in V_1 \otimes \cdots \otimes V_d$, we call $(\lambda, u_1, \ldots, u_d) \in \mathbb{K} \times V_1 \times \cdots \times V_d$ a normalized singular pair of T if:*

$$\begin{cases} \langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle = \lambda u_i, \\ \langle u_i, u_i \rangle = 1, \end{cases} \tag{11}$$

where $i = 1, \ldots, d$. Then, λ is called a normalized singular value and $(u_1, \ldots, u_d)$ is called a normalized singular vector tuple. When $\mathbb{K} = \mathbb{R}$, $\lambda \geq 0$, and $u_i \in V_i^+$, we call $(\lambda, u_1, \ldots, u_d)$ a nonnegative normalized singular pair of T.

Similar definitions have been proposed by several authors; for example, the following projective variant was introduced in [49].

Definition 6. *Given vector spaces $W_1, \ldots, W_d$ over $\mathbb{K}$ of dimensions $n_1, \ldots, n_d$, for $T \in W_1 \otimes \cdots \otimes W_d$, $([v_1], \ldots, [v_d]) \in \mathbb{P}W_1 \times \cdots \times \mathbb{P}W_d$ is called a projective singular vector tuple if [49]:*

$$\langle T, v_1 \otimes \cdots \otimes \widehat{v_i} \otimes \cdots \otimes v_d \rangle = \lambda_i v_i \tag{12}$$

for some $\lambda_i \in \mathbb{K}$, where $1 \leq i \leq d$.

The number of projective singular vector tuples of a general complex tensor was calculated in [49], which is the Euclidean Distance (ED) degree of $\mathbb{P}W_1 \times \cdots \times \mathbb{P}W_d$ by [50].

Theorem 7. *Let T be a generic tensor in $W_1 \otimes \cdots \otimes W_d$ over $\mathbb{C}$. Then T has exactly $c(n_1, \ldots, n_d)$ simple projective singular vector tuples corresponding to nonzero singular values, where $c(n_1, \ldots, n_d)$ is the coefficient of the monomial $\prod_{i=1}^d t_i^{n_i - 1}$ in the polynomial [49]:*

$$\prod_{i=1}^d \frac{\widehat{t_i}^{n_i} - t_i^{n_i}}{\widehat{t_i} - t_i}, \text{ where } \widehat{t_i} = \sum_{j \neq i} t_j, i = 1, \ldots, d.$$

Over $\mathbb{R}$, there are several nonempty open subsets $\mathcal{U}_1, \ldots, \mathcal{U}_k$ of $V_1 \otimes \cdots \otimes V_d$ such that the number of projective singular vector tuples is constant on each $\mathcal{U}_i$, denoted by m_i, for $i = 1, \ldots, k$, but $m_i \neq m_j$ if $i \neq j$. One way to describe the number of projective singular vector tuples by using a single number is to impose certain probability distribution on $V_1 \otimes \cdots \otimes V_d$ and compute the expected number of projective singular vector tuples of T when T is randomly drawn under the given distribution.

Theorem 8. *Let $T \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be a random tensor drawn under the Gaussian distribution. Then the expected number of projective singular vector tuples of T is given by [51]:*

$$\frac{(2\pi)^{d/2}}{2^{n/2}} \frac{1}{\prod_{i=1}^d \Gamma(\frac{n_i}{2})} \int_W |\det \mathcal{C}| \, d\mu_W,$$

where $n = \sum_i n_i$, Γ is Euler's gamma function,

$$\mathcal{C} = \begin{bmatrix} \lambda\, \mathrm{I}_{n_1-1} & A_{1,2} & \cdots & A_{1,d} \\ A_{1,2}^\top & \lambda\, \mathrm{I}_{n_2-1} & \cdots & A_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1,d}^\top & A_{2,d}^\top & \cdots & \lambda\, \mathrm{I}_{n_d-1} \end{bmatrix},$$

and W is the vector space formed by λ and $A_{i,j}$ with $i < j$.

Coming back to nonnegative tensors, we may have more information about their singular pairs than real tensors. Before studying singular pairs of nonnegative tensors, let us recall the following well-known Perron–Frobenius Theorem. See for example [52] for more details.

Theorem 9. *Given a nonnegative square matrix M,*

- *its spectral radius $r(M)$ is an eigenvalue.*
- *there is some nonnegative vector $v \neq 0$ such that $Mv = r(M)v$.*
- *$r(M) > 0$ if M is irreducible.*
- *there is some positive vector $u > 0$ such that $Mu = r(M)u$ if M is irreducible.*
- *if M is irreducible, then λ is an eigenvalue of M with a nonnegative eigenvector if and only if $\lambda = r(M)$.*
- *$r(M)$ is simple if M is irreducible.*
- *every eigenvalue λ satisfies $|\lambda| \leq r(M)$ if M is irreducible.*

The next three results, namely Lemmas 5–7, give an analogue of the tensorial Perron–Frobenius Theorem [53–56] for nonnegative normalized singular pairs, which will help us learn more about best rank-one approximations.

Lemma 5 (Existence)**.** *Any nonnegative tensor has (at least) a nonnegative normalized singular pair.*

Definition 7. *A tensor is called positive if all its entries are positive.*

Lemma 6 (Positivity)**.** *A positive tensor has a positive normalized singular pair.*

Recall that the *spectral norm* for a tensor, which is NP-hard to compute or approximate [57], is defined as follows.

Definition 8. *For $T \in V_1 \otimes \cdots \otimes V_d$ over $\mathbb{R}$, let $\|T\|_\sigma := \max\{|\langle T, u_1 \otimes \cdots \otimes u_d\rangle| : \|u_1\| = \cdots = \|u_d\| = 1\}$ be the spectral norm of T.*

Lemma 7 (Generic Uniqueness)**.** *Let T be a general real tensor. Then T has a unique normalized singular pair $(\lambda, u_1, \ldots, u_d)$ such that $\lambda = \|T\|_\sigma$.*

Lemma 7 motivates the following open question.

Question 2. *Can we give a sufficient condition such that any nonnegative tensor satisfying this condition has a unique normalized singular pair with $\lambda = \|T\|_\sigma$ and this condition can be satisfied by a general nonnegative tensor?*

For matrices over $\mathbb{R}$ or $\mathbb{C}$, by the Eckart–Young Theorem, best low-rank approximations can be obtained from successive best rank-one approximations. However, for tensors, this 'deflation procedure' does not work [48,58,59].

Besides singular vector tuples, eigenvalues and eigenvectors of a tensor can be also defined. Unlike matrices, there are several ways to define eigenvalues and eigenvectors of a tensor [60]. In this note, we will use the following one which was firstly introduced in [55,61].

Definition 9. *For $T \in V^{\otimes d}$ over $\mathbb{K}$, if:*

$$\langle T, u^{\otimes(d-1)} \rangle = \lambda u,$$

then $\lambda \in \mathbb{K}$ is called an eigenvalue of T, and $u \in V$ is called an eigenvector. The pair (λ, u) is called an eigenpair. Two eigenpairs (λ, u) and (μ, v) of T is said to be equivalent if $t^{d-2}\lambda = \mu$ and $tu = v$ for some $t \in \mathbb{K}$.

When T is a real or complex symmetric tensor, a best rank-one approximation of T can be always chosen to be symmetric [62,63], and thus is an eigenvector of T.

Theorem 10. *Let $T \in V^{\otimes(d+1)}$ be a real random tensor under the Gaussian distribution. Then the expected number of equivalence classes of eigenpairs of T is given by [64]:*

$$\frac{2^{n-1}\sqrt{d}^n\Gamma(n-\frac{1}{2})}{\sqrt{\pi}(d+1)^{n-\frac{1}{2}}\Gamma(n)}\left[2(n-1)\,{}_2F_1\left(1, n-\frac{1}{2}; \frac{3}{2}; \frac{d-1}{d+1}\right) + {}_2F_1\left(1, n-\frac{1}{2}; \frac{n+1}{2}; \frac{1}{d+1}\right)\right],$$

where ${}_2F_1$ is the Gaussian hypergeometric function.

The number of equivalence classes of eigenpairs of a generic complex symmetric tensor has been calculated in [65]. See [66] for another proof. This number is the ED degree of the Veronese variety [50].

Theorem 11. *Let V be an n-dimensional complex vector space. Let $T \in \mathsf{S}^d(V)$ be a symmetric tensor whose equivalence classes of eigenpairs are finitely many. Then, T has [65]:*

$$\frac{(d-1)^n - 1}{d-2}$$

equivalence classes of eigenpairs, counted with multiplicities.

For the real case, again, we may impose a probability distribution on $\mathsf{S}^d(V)$ and compute the expected number of equivalence classes of eigenvalues of a random symmetric tensor.

Theorem 12. *Let V be an n-dimensional real vector space of dimension n and $T \in \mathsf{S}^d(V)$ be drawn under the Gaussian distribution. Then, the expected number of equivalence classes of eigenpairs of T is [51]:*

$$\frac{1}{2^{(n^2+3n-2)/4}\prod_{j=1}^n \Gamma(j/2)} \int_{\mu_2 \le \cdots \le \mu_n} \int_{-\infty}^{+\infty} \left(\prod_{j=2}^n |\sqrt{d}\lambda - \sqrt{d-1}\mu_j|\right) \left(\prod_{i<j}(\mu_i - \mu_j)\right) e^{-\lambda^2/2 - \sum_{j=2}^n \mu_j^2/4}\, d\lambda\, d\mu_2 \cdots d\mu_d.$$

A closed formula of the above integral was given in [67]. Other variants of eigenvalues and eigenvectors can be found in [60,61]. For our purpose, we will focus on the following definition introduced in [48].

Definition 10. *For $T \in \mathsf{S}^d(V)$ over $\mathbb{K}$, $(\lambda, v) \in \mathbb{K} \times V$ is called a normalized eigenpair of T if the following equations hold:*

Cases 1.

$$\langle T, v^{\otimes(d-1)}\rangle = \lambda v, \langle v, v\rangle = 1.$$

In particular, λ is called a normalized eigenvalue. We say two normalized eigenpairs (α, u) and (β, v) of T are equivalent if:

$$(\alpha, u) = (\beta, v)$$

or if

$$(-1)^{d-2}\alpha = \beta \text{ and } u = -v.$$

In the following, we would like to investigate sufficient conditions to ensure a tensor to have a unique rank-one approximation. First we recall the definition of the *multipolynomial resultant* [68,69]. For any given $n+1$ homogeneous polynomials $F_0, \ldots, F_n \in \mathbb{C}[x_0, \ldots, x_n]$ with positive total degrees $d_0, \ldots, d_n$, let $F_i = \sum_{|\alpha|=d_i} c_{i,\alpha} x_0^{\alpha_0} \cdots x_n^{\alpha_n}$, where $\alpha = (\alpha_0, \ldots, \alpha_n)$ and $|\alpha| = \alpha_0 + \cdots + \alpha_n$. Associate every pair (i, α) with a variable $u_{i,\alpha}$. For a polynomial P in the variables $u_{i,\alpha}$, denote by $P(F_0, \ldots, F_n)$ the result obtained by letting $u_{i,\alpha} = c_{i,\alpha}$. Then we have the following classical result [68,69].

Theorem 13. *There is a unique polynomial, denoted by* Res*, in $u_{i,\alpha}$'s with integer coefficients, where $i = 0, \ldots, n$, and $|\alpha| \in \{d_0, \ldots, d_n\}$, that has the following properties:*

- *$F_0 = \cdots = F_n = 0$ has a nonzero solution over $\mathbb{C}$ if and only if* $\mathrm{Res}\,(F_0, \ldots, F_n) = 0$.
- $\mathrm{Res}\,(x_0^{d_0}, \ldots, x_n^{d_n}) = 1$.
- Res *is irreducible over $\mathbb{C}$.*

Definition 11. $\mathrm{Res}\,(F_0, \ldots, F_n) \in \mathbb{C}$ *is called the resultant of $F_0, \ldots, F_n$.*

Definition 12. *For a symmetric tensor T, the resultant $\psi_T(\lambda)$ of the following polynomials is called the characteristic polynomial of T [70].*

- *For $T \in \mathsf{S}^{2d-1}(V)$,*
$$\langle T, v^{\otimes(d-1)}\rangle - \lambda x^{d-2} v = 0 \quad \text{and} \quad x^2 - \langle v, v\rangle = 0.$$
- *For $T \in \mathsf{S}^{2d}(V)$,*
$$\langle T, v^{\otimes(2d-1)}\rangle - \lambda \langle v, v\rangle^{d-1} v = 0.$$

Note the resultant $\psi_T(\lambda)$ is a (univariate) polynomial in λ. We call the resultant of $\psi_T(\lambda)$ and its derivative $\psi_T'(\lambda)$, denoted by $D_{\mathrm{eig}}(T)$, the *eigen discriminant*.

Proposition 7. *Let V be a real vector space, and $\rho = \|T\|_\sigma$ [48]. Define*

$$H_\rho := \{T \in \mathsf{S}^d(V) \mid \rho \text{ is not a simple eigenvalue of } T\}.$$

Then, H_ρ is a real hypersurface in $\mathsf{S}^d(V)$.

Let $W = V \otimes_{\mathbb{R}} \mathbb{C}$ be the complexification of V. Then we have:

Theorem 14. $D_{\mathrm{eig}}(T) = 0$ *is a defining equation of the complex hypersurface [48]*

$$H_{\mathrm{disc}} := \{T \in \mathsf{S}^d(W) \mid T \text{ has a non-simple normalized eigenvalue}\}.$$

In fact H_ρ consists of some components of the real points of H_{disc}. In the sense of [50], Theorem 14 shows the ED discriminant of the Veronese variety is a hypersurface.

Corollary 4. *For $T \in \mathsf{S}^d(V)$, if $D_{\mathrm{eig}}(T) \neq 0$, then T has a unique best rank-one approximation.*

Corollary 5. *Let $T \in \mathsf{S}^d(V)$ be a nonnegative tensor. If $D_{\mathrm{eig}}(T) \neq 0$, then T has a unique best rank-one approximation, which is nonnegative and symmetric.*

Example 1. *Let $T = [T_{ijk}] \in \mathsf{S}^3(\mathbb{R}^2)$. Then $\psi_T(\lambda)$ is the resultant of the polynomials:*

Cases 2.

$$F_0 = T_{111}x^2 + 2T_{112}xy + T_{122}y^2 - \lambda xz, F_1 = T_{112}x^2 + 2T_{122}xy + T_{222}y^2 - \lambda yz, F_2 = x^2 + y^2 - z^2.$$

In fact, $\psi_T(\lambda) = \frac{1}{512}\det(G)$, where G is defined by:

$$G = \begin{bmatrix} T_{111} & T_{122} & 0 & 2T_{112} & -\lambda & 0 \\ T_{112} & T_{222} & 0 & 2T_{122} & 0 & -\lambda \\ 1 & 1 & -1 & 0 & 0 & 0 \\ 12T_{122} & 4T_{111}\lambda - 8T_{122}\lambda & 4T_{111}\lambda + 4T_{122}\lambda & 8T_{222}\lambda - 16T_{112}\lambda & 16T_{112}^2 - 4\lambda^2 - 16T_{111}T_{122} & 8T_{112}T_{122} - 8T_{111}T_{222} \\ 4T_{222}\lambda - 8T_{112}\lambda & 12T_{112}\lambda & 4T_{112}\lambda + 4T_{222}\lambda & 8T_{111}\lambda - 16T_{122}\lambda & 8T_{112}T_{122} - 8T_{111}T_{222} & 16T_{122}^2 - 4\lambda^2 - 16T_{112}T_{222} \\ 8T_{112}^2 - 8T_{111}T_{122} - 2\lambda^2 & 8T_{122}^2 - 8T_{222}T_{112} - 2\lambda^2 & -6\lambda^2 & 8T_{112}T_{122} - 8T_{111}T_{222} & 8T_{122}\lambda + 8T_{111}\lambda & 8T_{112}\lambda + 8T_{222}\lambda \end{bmatrix}.$$

Thus, $\psi_T(\lambda) = p_2\lambda^6 + p_4\lambda^4 + p_6\lambda^2 + p_8$, where each p_m is a homogeneous polynomial of degree m in T_{ijk}. See also [65,71].

For a general $T \in \mathsf{S}^3(\mathbb{R}^2)$, $\psi_T(\lambda) = \alpha(\lambda^2 - \gamma_1)(\lambda^2 - \gamma_2)(\lambda^2 - \gamma_3)$ for some $\alpha \in \mathbb{C}$, where $\gamma_1, \gamma_2, \gamma_3$ are distinct. So $D_{\mathrm{eig}}(T) \neq 0$.

For $T \in H_{\mathrm{disc}}$, $\psi_T(\lambda)$ has multiple roots. For example, let $A \in \mathsf{S}^3(\mathbb{R}^2)$ be defined by $A_{111} = A_{222} = 1$ and set other $A_{ijk} = 0$. Then $D_{\mathrm{eig}}(A) = 0$, which implies that A has a nonsimple eigenpair. Here $\psi_A(\lambda) = (\lambda+1)^2(\lambda-1)^2(2\lambda^2-1)$. So A has two eigenvectors $(1,0)$ and $(0,1)$ with eigenvalue 1, and two eigenvectors $(-1,0)$ and $(0,-1)$ with eigenvalue -1. This computation coincides with the fact that $A = a^{\otimes 3} + b^{\otimes 3}$ has two best rank-one approximations, namely $a^{\otimes 3}$ and $b^{\otimes 3}$, where a and b are two orthonormal vectors in $\mathbb{R}^2$.

Similarly, we can define characteristic polynomials for non-symmetric tensors. Let $W_1, \ldots, W_d$ be complex vector spaces. For $T \in W_1 \otimes \cdots \otimes W_d$, $u_i \in W_i$, and $\alpha_i \in \mathbb{C}$, we denote the resultant of the following equations by $\varphi_T(\lambda)$.

$$\begin{cases} \alpha_i \langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle = \lambda(\prod_{j \neq i} \alpha_j) u_i, \\ \langle u_i, u_i \rangle = \alpha_i^2, \end{cases} \tag{13}$$

where $i = 1, \ldots, d$. Then, $\varphi_T(\lambda)$ vanishes if and only if (13) has a nontrivial solution.

Definition 13. *$\varphi_T(\lambda)$ is called the singular characteristic polynomial of T.*

The following is an analogue of Definition 10.

Definition 14. *Let $T \in W_1 \otimes \cdots \otimes W_d$. Two normalized singular pairs $(\lambda, u_1, \ldots, u_d)$ and $(\mu, v_1, \ldots, v_d)$ of T are called equivalent if $(\lambda, u_1, \ldots, u_d) = (\mu, v_1, \ldots, v_d)$, or $(-1)^{d-2}\lambda = \mu$ and $u_i = -v_i$ for $i = 1, \ldots, d$.*

It follows from [72] that the subset $X \subseteq V_1 \otimes \cdots \otimes V_d$ consisting of tensors which do not have unique best rank-one approximations is contained in some hypersurface. In fact we can strengthen the result by showing that X *is* a hypersurface.

Theorem 15. *The following subset is an algebraic hypersurface in $V_1 \otimes \cdots \otimes V_d$ [48],*

$$X := \{T \in V_1 \otimes \cdots \otimes V_d : T \text{ has non-unique best rank-one approximations}\}.$$

Besides, we have the following property.

Proposition 8. *Let $W_1, \ldots, W_d$ be complex vector spaces. Then for a general $T \in W_1 \otimes \cdots \otimes W_d$, the equivalence classes of normalized singular pairs of T are distinct [48].*

Definition 15. *The resultant of φ_T and its derivative φ'_T is called the singular discriminant and denoted by $D_{\text{sing}}(T)$.*

Theorem 16. *$D_{\text{sing}}(T) = 0$ is a defining equation of the hypersurface [48]*

$$X_{\text{disc}} := \{T \in W_1 \otimes \cdots \otimes W_d \mid T \text{ has a non-simple normalized singular value}\},$$

and X consists of some components of the real points of X_{disc}.

Corollary 6. *For a real tensor T, if $D_{\text{sing}}(T) \neq 0$, then T has a unique best rank-one approximation.*

Corollary 7. *For a nonnegative tensor T, if $D_{\text{sing}}(T) \neq 0$, then T has a unique best rank-one approximation, which is nonnegative.*

Theorem 16 shows that the ED discriminant X_{disc} of $\text{Seg}(\mathbb{P}W_1 \times \cdots \times \mathbb{P}W_d)$ is a complex hypersurface when $d \geq 3$, and the set of real points of X_{disc} is a real hypersurface. It is worth noting that when $d = 2$, the set of real points of the ED discriminant of $\text{Seg}(\mathbb{P}W_1 \times \mathbb{P}W_2)$ has codimension 2 ([50], Example 7.6).

8. EM Algorithm

Expectation–Maximization (EM) algorithm, as a classical technique, has been used in nonnegative matrix factorizations, and its performance and geometry has been carefully studied. See [73] and the references therein. However, to the best of our knowledge, such an analysis for nonnegative tensors has not been written down. In this section, we routinely apply the EM algorithm to nonnegative tensor decompositions and give a description of the EM fixed points.

Given a real function:

$$f(p_1, \ldots, p_n) = \sum_{i=1}^{n} u_i \log p_i$$

of $p_1, \ldots, p_n$ with parameters $u_1, \ldots, u_n$, where $p_1, \ldots, p_n, u_1, \ldots, u_n$ satisfy

$$0 \leq u_i, p_i \leq 1 \text{ for } i = 1, \ldots, n, \quad \text{and } \sum_{i=1}^{n} u_i = \sum_{i=1}^{n} p_i = 1.$$

Then, the maximum of f is obtained when $p_i = u_i$ for $i = 1, \ldots, n$. In fact $(p_1 = u_1, \ldots, p_n = u_n)$ is a critical point of the Lagrangian:

$$\sum_{i=1}^{n} u_i \log p_i - \lambda(1 - \sum_{i=1}^{n} p_i).$$

Hence, for a given nonnegative rank-r tensor $u = (u_{i_1,\ldots,i_d})$ with:

$$u_+ = \sum_{i_1,\ldots,i_d} u_{i_1,\ldots,i_d} = 1,$$

a nonnegative rank decomposition:

$$u = \sum_{l=1}^{r} \lambda_l v_1^{(l)} \otimes \cdots \otimes v_d^{(l)},$$

in coordinates:

$$u_{i_1,\ldots,i_d} = \sum_{l=1}^{r} \lambda_l v_{1,i_1}^{(l)} \cdots v_{d,i_d}^{(l)}$$

where $v_{j,i_j}^{(l)}$ is the i_jth entry of the vector $v_j^{(l)}$, gives a maximum of the likelihood function:

$$\mathcal{L}(\lambda, v_i^{(l)}) = \sum_{i_1,\ldots,i_d} u_{i_1,\ldots,i_d} \log(\sum_{l=1}^{r} \lambda_l v_{1,i_1}^{(l)} \cdots v_{d,i_d}^{(l)}). \tag{14}$$

This is a hidden model in statistics, and a classical way to optimize (14) is that we first use the EM algorithm to maximize the following likelihood function

$$L(\lambda, v_i^{(l)}) = \sum_{i_1,\ldots,i_d} \sum_{l=1}^{r} w_{i_1,\ldots,i_d}^{(l)} \log(\lambda_l v_{1,i_1}^{(l)} \cdots v_{d,i_d}^{(l)}), \tag{15}$$

where:

$$u_{i_1,\ldots,i_d} = \sum_{l=1}^{r} w_{i_1,\ldots,i_d}^{(l)}.$$

By ([74], Theorem 1.15), the value of the likelihood function (15) weakly increases during every iteration of the EM algorithm, and the local maxima of $\mathcal{L}$ are among the EM fixed points (final outputs) of L [73,74].

Remark 1. *EM algorithm and its analogues have been widely used in nonnegative matrix factorizations, by maximizing different likelihood functions, i.e., finding critical points of different divergences, for example Kullback-Leibler divergence, β-divergence and so on. Similarly we can obtain other algorithms for nonnegative tensor decompositions as well by using different divergences. Usually the fixed points of these algorithms contain critical points, and the local maxima are among the critical points.*

A fixed point of EM algorithm need satisfy the following equations:

$$\lambda_k = \frac{1}{u_+} \sum_{i_1,\ldots,i_d} \frac{\lambda_k v_{1,i_1}^{(k)} \cdots v_{d,i_d}^{(k)}}{\sum_{l=1}^{r} \lambda_l v_{1,i_1}^{(l)} \cdots v_{d,i_d}^{(l)}} u_{i_1,\ldots,i_d} \tag{16}$$

$$v_{j,i_j}^{(k)} = \frac{1}{\lambda_k u_+} \sum_{i_1,\ldots,i_{j-1},i_{j+1},\ldots,i_d} \frac{\lambda_k v_{1,i_1}^{(k)} \cdots v_{d,i_d}^{(k)}}{\sum_{l=1}^{r} \lambda_l v_{1,i_1}^{(l)} \cdots v_{d,i_d}^{(l)}} u_{i_1,\ldots,i_d} \tag{17}$$

Since $\sum_{i_j} v_{j,i_j}^{(l)} = 1$, $\lambda_k > 0$. By canceling λ_k in (16) and (17), we have:

$$1 = \frac{1}{u_+} \sum_{i_1,\ldots,i_d} \frac{v_{1,i_1}^{(k)} \cdots v_{d,i_d}^{(k)}}{\sum_{l=1}^{r} \lambda_l v_{1,i_1}^{(l)} \cdots v_{d,i_d}^{(l)}} u_{i_1,\ldots,i_d}, \tag{18}$$

$$v_{j,i_j}^{(k)} = \frac{1}{u_+} \sum_{i_1,\ldots,i_{j-1},i_{j+1},\ldots,i_d} \frac{v_{1,i_1}^{(k)} \cdots v_{d,i_d}^{(k)}}{\sum_{l=1}^{r} \lambda_l v_{1,i_1}^{(l)} \cdots v_{d,i_d}^{(l)}} u_{i_1,\ldots,i_d}, \tag{19}$$

and (18) can be obtained from (19). By (19) we have:

$$v_{j,i_j}^{(k)}(\sum_{i_1,\ldots,i_{j-1},i_{j+1},\ldots,i_d}(u_+ - \frac{u_{i_1,\ldots,i_d}}{p_{i_1,\ldots,i_d}})v_{1,i_1}^{(k)}\cdots\widehat{v_{j,i_j}^{(k)}}\cdots v_{d,i_d}^{(k)}) = 0, \tag{20}$$

where $p_{i_1,\ldots,i_d}$ is the output in Algorithm 1. Let:

$$R_{i_1,\ldots,i_d} = u_+ - \frac{u_{i_1,\ldots,i_d}}{p_{i_1,\ldots,i_d}},$$

then (20) is equivalent to:

$$v_j^{(k)} \circ \langle R, v_1^{(k)} \otimes \cdots \otimes \widehat{v_j^{(k)}} \otimes \cdots \otimes v_d^{(k)} \rangle = 0, \tag{21}$$

where $\circ$ denotes the Hadamard product. Since we require $\sum_{i_1,\ldots,i_d} p_{i_1,\ldots,i_d} = 1$, the likelihood function

$$\mathcal{L} = \sum_{i_1,\ldots,i_d} u_{i_1,\ldots,i_d} \log p_{i_1,\ldots,i_d} - u_+ \log(\sum_{i_1,\ldots,i_d} p_{i_1,\ldots,i_d}),$$

then the gradient of $\mathcal{L}$ is R. Hence:

$$\langle R, v_1^{(k)} \otimes \cdots \otimes \widehat{v_j^{(k)}} \otimes \cdots \otimes v_d^{(k)} \rangle = 0$$

implies that R is orthogonal to the tangent space of $\widehat{\sigma}_r(\mathrm{Seg}(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}))$, i.e.,

$$\sum_k v_1^{(k)} \otimes \cdots \otimes v_d^{(k)}$$

is a critical point. Therefore, we arrive at the following description, which is a trivial generalization of ([73], Theorem 3).

Algorithm 1 EM Algorithm

Step 0: Given $\epsilon > 0$, select random $v_i^{(l)} \in \Delta_{n_i-1}$, $\lambda \in \Delta_{r-1}$;

E-Step: Define the expected hidden data tensor $w = (w_{i_1,\ldots,i_d}^{(k)})$ by

$$w_{i_1,\ldots,i_d}^{(k)} = \frac{\lambda_k v_{1,i_1}^{(k)} \cdots v_{d,i_d}^{(k)}}{\sum_{l=1}^r \lambda_l v_{1,i_1}^{(l)} \cdots v_{d,i_d}^{(l)}} u_{i_1,\ldots,i_d};$$

M-Step: Compute $v_i^{(l)}, \lambda$ to maximize the likelihood function:

$$\lambda_k^* = \frac{\sum_{i_1,\ldots,i_d} w_{i_1,\ldots,i_d}^{(k)}}{u_+},$$

$$v_{j,i_j}^{(k)*} = \frac{\sum_{i_1,\ldots,i_{j-1},i_{j+1},\ldots,i_d} w_{i_1,\ldots,i_d}^{(k)}}{\lambda_k^* u_+};$$

Step 3: If $|L(\lambda^*, v_i^{(l)*}) - L(\lambda, v_i^{(l)})| > \epsilon$ then set $(\lambda, v_i^{(l)}) = (\lambda^*, v_i^{(l)*})$ and go to the E-Step;

Step 4: Output $p_{i_1,\ldots,i_d} = \sum_k \lambda_k v_{1,i_1}^{(k)} \cdots v_{d,i_d}^{(k)}$.

Proposition 9. *The variety of EM fixed points is defined by:*

$$v_j^{(k)} \circ \langle R, v_1^{(k)} \otimes \cdots \otimes \widehat{v_j^{(k)}} \otimes \cdots \otimes v_d^{(k)} \rangle = 0, \quad j = 1, \ldots, d.$$

The subset that are critical is defined by:

$$\langle R, v_1^{(k)} \otimes \cdots \otimes \widehat{v_j^{(k)}} \otimes \cdots \otimes v_d^{(k)} \rangle = 0, \quad j = 1, \ldots, d.$$

A semialgebraic characterization of the set of nonnegative matrices with nonnegative ranks no greater than 3 was given in [73]. A semialgebraic characterization of the set of nonnegative tensors with nonnegative ranks no greater than 2 was given in [75].

Question 3. *Give a semialgebraic characterization of the set of nonnegative tensors with nonnegative ranks no greater than* 3.

9. Conclusions

In this short note, we give a very brief introduction to nonnegative tensors, mainly from the geometric perspective. More precisely, we review the generic uniqueness of rank decompositions of subgeneric nonnegative tensors and nonnegative typical ranks, and thus see the difference among the nonnegative, real, and complex settings. We review the generic uniqueness of nonnegative low-rank approximations. In particular, the rank-one approximation problem leads us to the spectral theory of nonnegative tensors. Finally, we describe the semialgebraic geometry of EM algorithm. Most of the results we present are obtained by studying the corresponding geometric properties of nonnegative tensors, and we have seen there are many open problems and unknown properties in this direction, which we hope would be understood better when more geometries are unveiled.

Funding: This research received no external funding.

Acknowledgments: The author is grateful to L. Chiantini and L.-H. Lim for their kind encouragements and useful discussions. Special thanks go to the anonymous referees for their careful reading and invaluable suggestions that significantly improved the presentation of this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bocci, C.; Chiantini, L.; Ottaviani, G. Refined methods for the identifiability of tensors. *Ann. Mat. Pura Appl.* **2014**, *193*, 1691–1702. [CrossRef]
2. Chiantini, L.; Ottaviani, G.; Vannieuwenhoven, N. On generic identifiability of symmetric tensors of subgeneric rank. *Trans. Am. Math. Soc.* **2017**, *369*, 4021–4042. [CrossRef]
3. Comon, P.; Golub, G.; Lim, L.H.; Mourrain, B. Symmetric tensors and symmetric tensor rank. *SIAM J. Matrix Anal. Appl.* **2008**, *30*, 1254–1279. [CrossRef]
4. Brachat, J.; Comon, P.; Mourrain, B.; Tsigaridas, E. Symmetric tensor decomposition. *Linear Algebra Appl.* **2010**, *433*, 1851–1872. [CrossRef]
5. Lim, L.H.; Comon, P. Nonnegative approximations of nonnegative tensors. *J. Chemom.* **2009**, *23*, 432–441. [CrossRef]
6. Shashua, A.; Hazan, T. Non-negative tensor factorization with applications to statistics and computer vision. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 792–799.
7. Smilde, A.; Bro, R.; Geladi, P. *Multi-Way Analysis*; Wiley: Chichester, UK, 2004.
8. Zhang, Q.; Wang, H.; Plemmons, R.J.; Pauca, V.P. Tensor methods for hyperspectral data analysis: A space object material identification study. *J. Opt. Soc. Am. A* **2008**, *25*, 3001–3012. [CrossRef]
9. Garcia, L.D.; Stillman, M.; Sturmfels, B. Algebraic geometry of Bayesian networks. *J. Symb. Comput.* **2005**, *39*, 331–355. [CrossRef]

10. Jordan, M.I. Graphical models. *Stat. Sci.* **2004**, *19*, 140–155. [CrossRef]
11. Koller, D.; Friedman, N. *Probabilistic Graphical Models: Principles and techniques*; Adaptive Computation and Machine Learning; MIT Press: Cambridge, MA, USA, 2009; p. xxxvi+1231.
12. Zhou, J.; Bhattacharya, A.; Herring, A.H.; Dunson, D.B. Bayesian factorizations of big sparse tensors. *J. Am. Stat. Assoc.* **2015**, *110*, 1562–1576. [CrossRef]
13. Shashua, A.; Zass, R.; Hazan, T. Multi-way Clustering Using Super-Symmetric Non-negative Tensor Factorization. In *Computer Vision—ECCV 2006*; Leonardis, A., Bischof, H., Pinz, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 595–608.
14. Veganzones, M.; Cohen, J.; Farias, R.; Chanussot, J.; Comon, P. Nonnegative tensor CP decomposition of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2577–2588. [CrossRef]
15. Bartoň, M.; Calo, V.M. Optimal quadrature rules for odd-degree spline spaces and their application to tensor-product-based isogeometric analysis. *Comput. Methods Appl. Mech. Energy* **2016**, *305*, 217–240. [CrossRef]
16. Mantzaflaris, A.; Jüttler, B.; Khoromskij, B.N.; Langer, U. Low rank tensor methods in Galerkin-based isogeometric analysis. *Comput. Methods Appl. Mech. Energy* **2017**, *316*, 1062–1085. [CrossRef]
17. Scholz, F.; Mantzaflaris, A.; Jüttler, B. Partial tensor decomposition for decoupling isogeometric Galerkin discretizations. *Comput. Methods Appl. Mech. Energy* **2018**, *336*, 485–506. [CrossRef]
18. Lee, D.D.; Seung, H.S. Algorithms for non-negative matrix factorization. *Adv. Neural Inf. Process. Syst.* **2001**, 556–562.
19. Chu, M.T.; Lin, M.M. Low-dimensional polytope approximation and its applications to nonnegative matrix factorization. *SIAM J. Sci. Comput.* **2008**, *30*, 1131–1155. [CrossRef]
20. Kim, H.; Park, H. Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM J. Matrix Anal. Appl.* **2008**, *30*, 713–730. [CrossRef]
21. Kim, J.; Park, H. Fast nonnegative matrix factorization: an active-set-like method and comparisons. *SIAM J. Sci. Comput.* **2011**, *33*, 3261–3281. [CrossRef]
22. Ho, N.D. Nonnegative Matrix Factorization Algorithms and Applications. Ph.D. Thesis, Ecole Polytechnique de Louvain, Ottignies-Louvain-la-Neuve, Belgium, 2008.
23. Zhou, G.; Cichocki, A.; Xie, S. Fast nonnegative matrix/tensor factorization based on low-rank approximation. *IEEE Trans. Signal Process.* **2012**, *60*, 2928–2940. [CrossRef]
24. Vavasis, S.A. On the complexity of nonnegative matrix factorization. *SIAM J. Optim.* **2009**, *20*, 1364–1377. [CrossRef]
25. Arora, S.; Ge, R.; Kannan, R.; Moitra, A. Computing a nonnegative matrix factorization—provably. In *STOC'12—Proceedings of the 2012 ACM Symposium on Theory of Computing*; ACM: New York, NY, USA, 2012; pp. 145–161, doi:10.1145/2213977.2213994.
26. Bro, R. Multi-Way Analysis in the Food Industry: Models, Algorithms, and Applications. Ph.D. Thesis, Universiteit van Amsterdam, Amsterdam, The Netherlands, 1998.
27. Friedlander, M.P.; Hatz, K. Computing non-negative tensor factorizations. *Optim. Methods Softw.* **2008**, *23*, 631–647. [CrossRef]
28. Cohen, J.; Farias, R.C.; Comon, P. Fast Decomposition of Large Nonnegative Tensors. *IEEE Signal Process. Lett.* **2015**, *22*, 862–866. [CrossRef]
29. Cichocki, A.; Zdunek, R.; Phan, A.H.; Amari, S.I. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*; Wiley Publishing: Hoboken, NJ, USA, 2009.
30. Kim, J.; He, Y.; Park, H. Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework. *J. Glob. Optim.* **2014**, *58*, 285–319. [CrossRef]
31. Zhou, G.; Cichocki, A.; Zhao, Q.; Xie, S. Nonnegative matrix and tensor factorizations: An algorithmic perspective. *IEEE Signal Process. Mag.* **2014**, *31*, 54–65. [CrossRef]
32. Strassen, V. Rank and optimal computation of generic tensors. *Linear Algebra Appl.* **1983**, *52*, 645–685. [CrossRef]
33. Lickteig, T. Typical tensorial rank. *Linear Algebra Appl.* **1985**, *69*, 95–120. [CrossRef]
34. Abo, H.; Ottaviani, G.; Peterson, C. Induction for secant varieties of Segre varieties. *Trans. Am. Math. Soc.* **2009**, *361*, 767–792. [CrossRef]
35. Kruskal, J.B. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* **1977**, *18*, 95–138. [CrossRef]

36. Stegeman, A. On uniqueness conditions for Candecomp/Parafac and Indscal with full column rank in one mode. *Linear Algebra Appl.* **2009**, *431*, 211–227. [CrossRef]
37. Chiantini, L.; Ottaviani, G. On generic identifiability of 3-tensors of small rank. *SIAM J. Matrix Anal. Appl.* **2012**, *33*, 1018–1037. [CrossRef]
38. Domanov, I.; De Lathauwer, L. On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part I: Basic results and uniqueness of one factor matrix. *SIAM J. Matrix Anal. Appl.* **2013**, *34*, 855–875. [CrossRef]
39. Domanov, I.; De Lathauwer, L. On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part II: Uniqueness of the overall decomposition. *SIAM J. Matrix Anal. Appl.* **2013**, *34*, 876–903. [CrossRef]
40. Chiantini, L.; Ottaviani, G.; Vannieuwenhoven, N. An algorithm for generic and low-rank specific identifiability of complex tensors. *SIAM J. Matrix Anal. Appl.* **2014**, *35*, 1265–1287. [CrossRef]
41. Domanov, I.; De Lathauwer, L. Generic uniqueness conditions for the canonical polyadic decomposition and INDSCAL. *SIAM J. Matrix Anal. Appl.* **2015**, *36*, 1567–1589. [CrossRef]
42. Qi, Y.; Comon, P.; Lim, L.H. Semialgebraic geometry of nonnegative tensor rank. *SIAM J. Matrix Anal. Appl.* **2016**, *37*, 1556–1580. [CrossRef]
43. Alexander, J.; Hirschowitz, A. Polynomial interpolation in several variables. *J. Algebr. Geom.* **1995**, *4*, 201–222.
44. Chiantini, L.; Ciliberto, C. On the concept of k-secant order of a variety. *J. Lond. Math. Soc.* **2006**, *73*, 436–454. [CrossRef]
45. Ballico, E. On the weak non-defectivity of Veronese embeddings of projective spaces. *Cent. Eur. J. Math.* **2005**, *3*, 183–187. [CrossRef]
46. Mella, M. Singularities of linear systems and the Waring problem. *Trans. Am. Math. Soc.* **2006**, *358*, 5523–5538. [CrossRef]
47. Landsberg, J.M. *Tensors: Geometry and Applications*; Graduate Studies in Mathematics; American Mathematical Society: Providence, RI, USA, 2012; Volume 128, p. xx+439.
48. Qi, Y.; Comon, P.; Lim, L.H. Uniqueness of nonnegative tensor approximations. *IEEE Trans. Inform. Theory* **2016**, *62*, 2170–2183. [CrossRef]
49. Friedland, S.; Ottaviani, G. The number of singular vector tuples and uniqueness of best rank-one approximation of tensors. *Found. Comput. Math.* **2014**, *14*, 1209–1242. [CrossRef]
50. Draisma, J.; Horobeţ, E.; Ottaviani, G.; Sturmfels, B.; Thomas, R.R. The Euclidean distance degree of an algebraic variety. *Found. Comput. Math.* **2016**, *16*, 99–149. [CrossRef]
51. Draisma, J.; Horobeţ, E. The average number of critical rank-one approximations to a tensor. *Linear Multilinear Algebra* **2016**, *64*, 2498–2518. [CrossRef]
52. Berman, A.; Plemmons, R.J. *Nonnegative Matrices in the Mathematical Sciences*; Classics in Applied Mathematics; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 1994; Volume 9, p. xx+340, doi:10.1137/1.9781611971262.
53. Chang, K.C.; Pearson, K.; Zhang, T. Perron-Frobenius theorem for nonnegative tensors. *Commun. Math. Sci.* **2008**, *6*, 507–520. [CrossRef]
54. Friedland, S.; Gaubert, S.; Han, L. Perron-Frobenius theorem for nonnegative multilinear forms and extensions. *Linear Algebra Appl.* **2013**, *438*, 738–749. [CrossRef]
55. Lim, L.H. Singular Values and Eigenvalues of tensors: A Variational Approach. In Proceedings of the 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Puerto Vallarta, Mexico, 13–15 December 2005; pp. 129–132.
56. Yang, Y.; Yang, Q. Further results for Perron-Frobenius theorem for nonnegative tensors. *SIAM J. Matrix Anal. Appl.* **2010**, *31*, 2517–2530. [CrossRef]
57. Hillar, C.J.; Lim, L.H. Most tensor problems are NP-hard. *J. ACM* **2013**, *60*. [CrossRef]
58. Stegeman, A.; Comon, P. Subtracting a best rank-1 approximation may increase tensor rank. *Linear Algebra Appl.* **2010**, *433*, 1276–1300. [CrossRef]
59. Vannieuwenhoven, N.; Nicaise, J.; Vandebril, R.; Meerbergen, K. On generic nonexistence of the Schmidt-Eckart-Young decomposition for complex tensors. *SIAM J. Matrix Anal. Appl.* **2014**, *35*, 886–903. [CrossRef]
60. Chang, K.; Qi, L.; Zhang, T. A survey on the spectral theory of nonnegative tensors. *Numer. Linear Algebra Appl.* **2013**, *20*, 891–912. [CrossRef]
61. Qi, L. Eigenvalues of a real supersymmetric tensor. *J. Symb. Comput.* **2005**, *40*, 1302–1324. [CrossRef]

62. Banach, S. Über homogene Polynome in (L^2). *Stud. Math.* **1938**, *7*, 36–44. [CrossRef]
63. Friedland, S. Best rank one approximation of real symmetric tensors can be chosen symmetric. *Front. Math. China* **2013**, *8*, 19–40. [CrossRef]
64. Breiding, P. The expected number of eigenvalues of a real Gaussian tensor. *SIAM J. Appl. Algebra Geom.* **2017**, *1*, 254–271. [CrossRef]
65. Cartwright, D.; Sturmfels, B. The number of eigenvalues of a tensor. *Linear Algebra Appl.* **2013**, *438*, 942–952. [CrossRef]
66. Oeding, L.; Ottaviani, G. Eigenvectors of tensors and algorithms for Waring decomposition. *J. Symb. Comput.* **2013**, *54*, 9–35. [CrossRef]
67. Breiding, P. The average number of critical rank-one-approximations to a symmetric tensor. *arXiv* **2017**, arXiv:1701.07312.
68. Gel'fand, I.M.; Kapranov, M.M.; Zelevinsky, A.V. *Discriminants, Resultants, and Multidimensional Determinants*; Mathematics: Theory & Applications; Birkhäuser Boston, Inc.: Boston, MA, USA, 1994; p. x+523.
69. Cox, D.A.; Little, J.; O'Shea, D. *Using Algebraic Geometry*, 2nd ed.; Graduate Texts in Mathematics; Springer: New York, NY, USA, 2005; Volume 185, p. xii+572.
70. Qi, L. Eigenvalues and invariants of tensors. *J. Math. Anal. Appl.* **2007**, *325*, 1363–1377. [CrossRef]
71. Li, A.M.; Qi, L.; Zhang, B. E-characteristic polynomials of tensors. *Commun. Math. Sci.* **2013**, *11*, 33–53. [CrossRef]
72. Friedland, S.; Stawiska, M.G. Some approximation problems in semi-algebraic geometry. *Constr. Approx. Funct.* **2015**, *107*, 133–147. [CrossRef]
73. Kubjas, K.; Robeva, E.; Sturmfels, B. Fixed points EM algorithm and nonnegative rank boundaries. *Ann. Stat.* **2015**, *43*, 422–461. [CrossRef]
74. Pachter, L.; Sturmfels, B. *Algebraic Statistics for Computational Biology*; Cambridge University Press: New York, NY, USA, 2005.
75. Allman, E.S.; Rhodes, J.A.; Sturmfels, B.; Zwiernik, P. Tensors of nonnegative rank two. *Linear Algebra Appl.* **2015**, *473*, 37–53. [CrossRef]

© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).