Uniqueness of Nonnegative Tensor Approximations

Yang Qi, Pierre Comon, *Fellow, IEEE*, and Lek-Heng Lim

***Abstract*— We show that for a nonnegative tensor, a best nonnegative rank-*r* approximation is almost always unique, its best rank-one approximation may always be chosen to be a best nonnegative rank-one approximation, and the set of nonnegative tensors with nonunique best rank-one approximations forms an algebraic hypersurface. We show that the last part holds true more generally for real tensors and, thereby, determine a polynomial equation, so that a real or nonnegative tensor that does not satisfy this equation is guaranteed to have a unique best rank-one approximation. We also establish an analogue for real or nonnegative symmetric tensors. In addition, we prove a singular vector variant of the Perron–Frobenius theorem for positive tensors and apply it to show that a best nonnegative rank-*r* approximation of a positive tensor can never be obtained by deflation. As an aside, we verify that the Euclidean distance (ED) discriminants of the Segre variety and the Veronese variety are hypersurfaces and give defining equations of these ED discriminants.**

***Index Terms*— approximation methods, eigenvalues and eigenfunctions, geometry, modules, polynomials, Bayes methods, fluorescence, remote sensing, source separation.**

I. Introduction

Nonnegative tensor decomposition, i.e., a decomposition of a tensor with nonnegative entries (with respect to a fixed choice of bases) into a sum of tensor products of nonnegative vectors, arises in a wide range of applications. These include hyperspectral imaging, spectroscopy, statistics, phylogenetics, data mining, pattern recognition, among other areas; see [45], [52], [54], [62] and the references therein. One important reason for its prevalence is that such a decomposition shows how a joint distribution of discrete random variables decomposes when they are independent conditional on a discrete latent random variable [45], [64] — a ubiquitous model that underlies many applications. This is in fact one of the simplest Bayesian network [28], [33], [37], a local expression of the joint distribution of a set of random variables x_i as

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

where θ is some unknown latent random variable. The relation expressed in (1) is often called the *naive Bayes hypothesis*. In the case when the random variables $x_1, \dots, x_d$ and the latent variable θ take only a finite number of values, the decomposition becomes one of the form

$$t_{i_1,\dots,i_d} = \sum_{p=1}^{r} \lambda_r u_{i_1,p} \cdots u_{i_d,p}. \tag{2}$$

One can show [45] that *any* decomposition of a nonnegative tensor of the form in (2) may, upon normalization by a suitable constant, be regarded as (1), i.e., a marginal decomposition of a joint probability mass function into conditional probabilities under the naive Bayes hypothesis. In the event when the latent variable θ is not discrete or finite, one may argue that (2) becomes an approximation with '$\approx$' in place of '$=$'.

In this article, we investigate several questions regarding nonnegative tensor decompositions and approximations, focusing in particular on uniqueness issues. In Section II, we define nonnegative tensors in a way that parallels the usual abstract definition of tensors in algebra. We will view them as elements in a tensor product of cones, i.e., tensors in $C_1 \otimes \cdots \otimes C_d$ where $C_1, \dots, C_d$ are cones and the tensor product is that of $\mathbb{R}_+$-semimodules (we write $\mathbb{R}_+ := [0, \infty)$ for the nonnegative reals). The special case $C_1 = \mathbb{R}_+^{n_1}, \dots, C_d = \mathbb{R}_+^{n_d}$ then reduces to nonnegative tensors.

It has been established in [45] that every nonnegative tensor has a best nonnegative rank-*r* approximation. In Section IV we will show that this best approximation is almost always unique. Furthermore, the set of nonnegative tensors of nonnegative rank $> r$ that do not have a unique best rank-*r* approximation form a semialgebraic set contained in a hypersurface. For the special case when $r = 1$, we first show in Section V that for a nonnegative tensor, the best nonnegative rank-one and best rank-one approximations coincide. In Section VII, by exploring normalized singular pairs, we find an explicit polynomial expression describing the hypersurface of real (or nonnegative) tensors that admit non-unique best rank-one approximations, which allows one to check whether a given tensor has a unique best rank-one approximation. This polynomial expression also gives a defining equation of the Euclidean distance discriminant of the Segre variety [22]. In Section VI, we find results analogous to those in Section VII for real (or nonnegative) symmetric tensors. We prove an analogue of the Perron–Frobenius theorem for singular values/vectors of positive tensors in Section V and, among other things, deduce that one cannot obtain a best nonnegative rank-*r* approximation of a positive tensor by 'deflation', i.e., by finding *r* successive best nonnegative rank-one approximations.

Manuscript received October 29, 2014; revised November 16, 2015; accepted February 9, 2016. Date of publication February 29, 2016; date of current version March 16, 2016. Y. Qi and P. Comon were supported by the European Research Council within the European Community's Seventh Framework Program FP7/2007-2013 under Grant 320594. L.-H. Lim was supported in part by AFOSR under Grant FA9550-13-1-0133, in part by DARPA under Grant D15AP00109, and in part by NSF under Grant IIS 1546413, Grant DMS 1209136, and Grant DMS 1057064.

Y. Qi and P. Comon are with the Gipsa Laboratory, Centre National de la Recherche Scientifique, University of Grenoble Alpes, Grenoble 38000, France (e-mail: yang.qi@gipsa-lab.fr; p.comon@ieee.org).

L.-H. Lim is with the Computational and Applied Mathematics Initiative, Department of Statistics, The University of Chicago, Chicago, IL 60637 USA (e-mail: lekheng@uchicago.edu).

Communicated by G. Matz, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2016.2532906

These results would likely shed light on the large number of computational methods for nonnegative matrix factorizations and nonnegative tensor decompositions [2], [6], [11]–[13], [27], [32], [34]–[36], [40], [59], [63].

II. NONNEGATIVE TENSORS

A tensor of order d (d-tensor for short) may be represented as a d-dimensional hypermatrix, i.e., a d-dimensional array of (usually) real or complex values. This is a higher-order generalization of the fact that a 2-tensor, i.e., a linear operator, a bilinear form, or a dyad, can always be represented as a matrix. Such a coordinate representation sometimes hides intrinsic properties — in particular, this array of coordinates is meaningful only if the bases of underlying vector spaces have been specified in the first place. With this in mind, we prefer to define tensors properly rather than simply regarding them as d-dimensional arrays of numbers.

The following is the standard definition of tensors. We will see later how we may obtain an analogous definition for nonnegative tensors.

Definition 1: Let V_i be a vector space of finite dimension n_i over a field $\mathbb{K}$, $i = 1, \dots, d$, and let $V_1 \times \cdots \times V_d$ be the set of d-tuples of vectors. Then the *tensor product* $V = V_1 \otimes \cdots \otimes V_d$ is the free linear space spanned by $V_1 \times \cdots \times V_d$ quotient by the equivalence relation

$$\begin{aligned}(v_1, \dots, \alpha v_i + \beta v_i', \dots, v_d) \\ \sim \alpha(v_1, \dots, v_i, \dots, v_d) + \beta(v_1, \dots, v_i', \dots, v_d)\end{aligned} \tag{3}$$

for every $v_i, v_i' \in V_i$, $\alpha_i, \beta_i \in \mathbb{K}$, $i = 1, \dots, d$. A *tensor* is an element of $V_1 \otimes \cdots \otimes V_d$.

In particular, (3) gives

$$(\alpha_1 v_1, \alpha_2 v_2, \dots, \alpha_d v_d) = \left(\prod_{i=1}^{d} \alpha_i\right) (v_1, v_2, \dots, v_d) \tag{4}$$

More details on the definition of tensor spaces may be found in [14], [30], [39], and [43].

A *decomposable tensor* is one of the form $v_1 \otimes \cdots \otimes v_d$, $v_i \in V_i$, $i = 1, \dots, d$. It represents the equivalence class of tuples up to scaling as in (4), i.e.,

$$v_1 \otimes \cdots \otimes v_d = \left\{(\alpha_1 v_1, \dots, \alpha_d v_d) : \prod_{i=1}^{d} \alpha_i = 1\right\}.$$

By (4), it is clear that a decomposable tensor cannot in general be uniquely represented by a d-tuple of vectors, what is often called a "scaling indeterminacy" in the engineering literature. When we use the term 'unique' in this article, it is implicit that the uniqueness is only up to scaling of this nature.

From the way a tensor is defined in Definition 1, it is immediate that a nonzero tensor can always be expressed as a finite sum of nonzero decomposable tensors. When the number of summands is minimal, this decomposition is called a *rank decomposition* (the term "canonical polyadic" or CP is often also used) and the number of summands in such a decomposition is called the *rank* of the tensor. In other words, we have the following:

For every $T \in V_1 \otimes \cdots \otimes V_d$, there exist $v_{i,p} \in V_i$, $i = 1, \dots, d$, $p = 1, \dots, \operatorname{rank}(T)$, such that

$$T = \sum_{p=1}^{\operatorname{rank}(T)} v_{1,p} \otimes \cdots \otimes v_{d,p}. \tag{5}$$

We present the above material, which is largely standard knowledge, to motivate an analogous construction for real nonnegative tensors. We will first define nonnegative tensors in a coordinate-dependent manner (i.e., depending on a choice of bases on $V_1, \dots, V_d$), and then in a coordinate-independent manner.

Definition 2: For each $i = 1, \dots, d$, let V_i be a real vector space with $\dim V_i = n_i$. For any fixed choice of basis $\{v_{i,1}, \dots, v_{i,n_i}\}$ for V_i, we denote by V_i^+ the subset of vectors with nonnegative coefficients in V_i, i.e.,

$$V_i^+ = \left\{\sum_{p=1}^{n_i} \alpha_p v_{i,p} \in V_i : \alpha_1, \dots, \alpha_{n_i} \in \mathbb{R}_+\right\}.$$

We will call an element in $V := V_1 \otimes \cdots \otimes V_d$ of the form $u_1 \otimes \cdots \otimes u_d$ where $u_i \in V_i^+$ for $i = 1, \dots, d$, a *nonnegatively decomposable tensor*. The set of *nonnegative tensors* V^+ is then the subset of V defined by

$$\begin{aligned}V^+ = \Bigg\{\sum_{p=1}^{r} u_{1,p} \otimes \cdots \otimes u_{d,p} \in V : u_{i,p} \in V_i^+, \\ i = 1, \dots, d,\ p = 1, \dots, r,\ r \in \mathbb{N}\Bigg\}.\end{aligned}$$

By its definition, every element of V^+ has a representation as a finite sum of nonnegatively decomposable tensors. A decomposition of minimal length then yields the notions of nonnegative tensor rank and nonnegative tensor rank decomposition.

Definition 3: For every $T \in V^+$, there exist $v_{i,p} \in V_i^+$, $i = 1, \dots, d$, $p = 1, \dots, r$, such that

$$T = \sum_{p=1}^{\operatorname{rank}_+(T)} v_{1,p} \otimes \cdots \otimes v_{d,p} \tag{6}$$

where

$$\operatorname{rank}_+(T) := \left\{r : T = \sum_{p=1}^{r} v_{1,p} \otimes \cdots \otimes v_{d,p}\right\}. \tag{7}$$

We will call (7) *nonnegative tensor rank* or *nonnegative rank* for short and (6) a *nonnegative rank decomposition* of the nonnegative tensor T.

An obvious property is that $\operatorname{rank}_+(T) \geq \operatorname{rank}(T)$ for any $T \in V^+$.

We now examine an alternative coordinate-free approach to defining nonnegative tensors and nonnegative rank. This approach is also more general, yielding a notion of *conic rank* for a tensor product of any convex cones. We first recall the definition of a tensor product of semimodules. See [4] for details on the existence and a construction of such a tensor product.

Definition 4: Let R be a commutative semiring and M, N be R-semimodules (cf. Appendix for the definitions of semirings and semimodules). A *tensor product* $M \otimes_R N$ of M and N is an R-semimodule satisfying the universal property: There is an R-bilinear map $\varphi : M \times N \to M \otimes_R N$ such that given any other R-semimodule S together with an

R-bilinear map $h : M \times N \to S$, there is a unique R-linear map $\tilde{h} : M \otimes_R N \to S$ satisfying $h = \tilde{h} \circ \varphi$.

Recall that a *convex cone* C is a subset of a vector space over an ordered field that is closed under linear combinations with nonnegative coefficients, i.e., $\alpha x + \beta y$ belongs to C for all $x, y \in C$ and any nonnegative scalars α, β.

Since any convex cone $C_i \subset V_i$ is a semimodule over the semiring $\mathbb{R}_+$, we have the unique tensor product of these convex cones $C_1 \otimes \cdots \otimes C_d$ as an $\mathbb{R}_+$-semimodule up to isomorphism. More precisely, the *tensor product of cones* $C_1 \otimes \cdots \otimes C_d$ is the quotient monoid $F(C_1, \ldots, C_d)/\sim$, where $F(C_1, \ldots, C_d)$ is the free monoid generated by all n-tuples $(v_1, \ldots, v_d) \in C_1 \times \cdots \times C_d$, and $\sim$ is the equivalence relation on $F(C_1, \ldots, C_d)$ defined by

$$\begin{aligned}(v_1, \ldots, \alpha v_i + \beta v_i', \ldots, v_d)\\ \sim \alpha(v_1, \ldots, v_i, \ldots, v_d) + \beta(v_1, \ldots, v_i', \ldots, v_d)\end{aligned}$$

for every $v_i, v_i' \in C_i$, $\alpha, \beta \in \mathbb{R}_+$, and $i = 1, \ldots, d$. The commutative monoid $C_1 \otimes \cdots \otimes C_d$ is an $\mathbb{R}_+$-semimodule. We write $v_1 \otimes \cdots \otimes v_d$ for the equivalence class representing $(v_1, \ldots, v_d)$ in $F(C_1, \ldots, C_d)/\sim$.

A *multiconic map* from $C_1 \times \cdots \times C_d$ to a convex cone C is a map $\varphi : C_1 \times \cdots \times C_d \to C$ with the property that

$$\begin{aligned}&\varphi(u_1, \ldots, \alpha v_i + \beta w_i, \ldots, u_d)\\ &\quad = \alpha\varphi(u_1, \ldots, v_i, \ldots, u_d) + \beta\varphi(u_1, \ldots, w_i, \ldots, u_d)\end{aligned}$$

for all $\alpha, \beta \in \mathbb{R}_+$, $i = 1, \ldots, d$.

The multiconic map $\nu : C_1 \times \cdots \times C_m \to C_1 \otimes \cdots \otimes C_d$ defined by

$$\nu(v_1, \ldots, v_d) = v_1 \otimes \cdots \otimes v_d \in F(C_1, \ldots, C_d)/\sim$$

and extended nonnegative linearly to all of $C_1 \times \cdots \times C_d$ satisfies the *universal factorization property* often used to define tensor product spaces: If φ is a multiconic map from $C_1 \times \cdots \times C_d$ into a convex cone C, then there exists a unique $\mathbb{R}_+$-linear map ψ from $C_1 \otimes \cdots \otimes C_d$ into C, that makes the following diagram commutative:

$$\begin{tikzcd} C_1 \times \cdots \times C_d \arrow[r, "\nu"] \arrow[dr, "\varphi"'] & C_1 \otimes \cdots \otimes C_d \arrow[d, "\psi"] \\ & C \end{tikzcd}$$

i.e., $\psi\nu = \varphi$. Strictly speaking we should have written $C_1 \otimes_{\mathbb{R}_+} \cdots \otimes_{\mathbb{R}_+} C_d$ to indicate that the tensor product is one of $\mathbb{R}_+$-semimodules but this is obvious from context. Note that Definition 4 is consistent with our earlier definition of nonnegative tensors since $V^+ = V_1^+ \otimes \cdots \otimes V_d^+$ as tensor product of cones over $\mathbb{R}_+$.

In [60], the tensor product of $C_1, \ldots, C_d$ is defined to be the convex cone in $V_1 \otimes \cdots \otimes V_d$ formed by $v_1 \otimes \cdots \otimes v_d \in V_1 \otimes \cdots \otimes V_d$, where $v_i \in C_i$, and showed that this tensor product satisfies the above universal factorization property. By the uniqueness of the $\mathbb{R}_+$-semimodule satisfying the universal property, our construction and the one in [60] are equivalent.

If $C_1 = \mathbb{R}_+^{n_1}, \ldots, C_d = \mathbb{R}_+^{n_d}$, we may identify

$$\mathbb{R}_+^{n_1} \otimes \cdots \otimes \mathbb{R}_+^{n_d} = \mathbb{R}_+^{n_1 \times \cdots \times n_d}$$

through the interpretation of the tensor product of vectors as a hypermatrix via the Segre outer product

$$\begin{aligned}&[v_1(1), \ldots, v_1(n_1)]^\mathsf{T} \otimes \cdots \otimes [v_d(1), \ldots, v_d(n_d)]^\mathsf{T}\\ &\qquad = [v_1(i_1) \cdots v_d(i_d)]_{i_1, \ldots, i_d = 1}^{n_1, \ldots, n_d}.\end{aligned}$$

Here we write $v(j)$ for the jth coordinate of $v \in \mathbb{R}^n$.

We note that one may easily extend the notion of nonnegative rank and nonnegative rank decomposition to tensor product of other cones.

Definition 5: A tensor $T \in C_1 \otimes \cdots \otimes C_d$ is said to be *decomposable* if T is of the form $u_1 \otimes \cdots \otimes u_d$, where $u_i \in C_i$. For $T \in C_1 \otimes \cdots \otimes C_d$, the *conic rank* of T, denoted by $\operatorname{rank}_+(T)$, is the minimal value of r such that $T = \sum_{p=1}^r u_{1,p} \otimes \cdots \otimes u_{d,p}$, where $u_{i,p} \in C_i$, i.e., T is contained in the convex cone generated by $u_{1,1} \otimes \cdots \otimes u_{d,1}, \ldots, u_{1,r} \otimes \cdots \otimes u_{d,r}$. Such a decomposition will be called a *conic rank decomposition*.

In the remainder of this paper, we focus our attention on the case $V^+ = V_1^+ \otimes \cdots \otimes V_d^+$, the convex cone of nonnegative d-tensors although we will point out whenever a result holds more generally for arbitrary cones. For any given positive integer r, we let

$$D_r^+ = \{X \in V_1^+ \otimes \cdots \otimes V_d^+ : \operatorname{rank}_+(X) \le r\}$$

denote the set of tensors of nonnegative rank not more than r.

III. Uniqueness of Rank Decompositions

From the standpoints of both identifiability and well-posedness, an important issue is whether a rank decomposition of the form (5) is unique. It is clear that such decompositions can never be unique when $d = 2$, i.e., for matrices. But when $d > 2$, rank decompositions are often unique, which is probably the strongest reason for their utility in applications. There are well-known sufficient conditions ensuring uniqueness of rank decomposition [20], [21], [38], [53]and many recent works on the uniqueness of generic tensors of certain ranks [5], [9], [10], [56]. We highlight three notable results.

Theorem 6 (Kruskal): The rank decomposition of a d-tensor T is unique if

$$\operatorname{rank}(T) \le \frac{1 + \sum_{i=1}^d (\kappa_i - 1)}{2}$$

where κ_i denote the Kruskal rank of the factors $u_{i,1}, \ldots, u_{i,\operatorname{rank}(T)}$, which is generically equal to the dimension n_i when $n_i \le \operatorname{rank}(T)$.

Theorem 7 (Bocci–Chiantini–Ottaviani): The rank decomposition of a generic d-tensor T of rank-r is unique when

$$r \le \frac{\prod_{i=1}^d n_i - (n_1 + n_2 + n_3 - 2)\prod_{i=3}^d n_i}{1 + \sum_{i=1}^d (n_i - 1)}.$$

Theorem 8 (Chiantini–Ottaviani–Vannieuwenhoven): The rank decomposition of a generic d-tensor T of rank-r is unique when

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

if $\prod_{i=1}^d n_i \le 15000$, with some exceptional cases.

The authors of [10] also strengthened the above result by a prior compression of tensor T. The consequence is that the dimensions n_i in Theorem 8 may be replaced by the multilinear rank of T, which allows significant tightening of the upper bound for low multilinear rank tensors. The maximum R_{smax} where a generic tensor with rank $\leq R_{\text{smax}}$ has a unique rank decomposition has been called the maximum stable rank in [57]. Theorem 8 implies that if $\prod_{i=1}^{d} n_i \leq 15000$, then aside from the exceptional cases, the maximum stable rank is $\left\lceil \prod_{i=1}^{d} n_i/[1+\sum_{i=1}^{d}(n_i-1)]\right\rceil - 1$, which is one less than the (expected) generic rank [1], [16], [42], [56].

Nevertheless these results do not apply directly to *nonnegative* decompositions over $\mathbb{R}_+$ (as opposed to decompositions over $\mathbb{C}$) nor to rank-r *approximations* (as opposed to rank-r decompositions). The purpose of this paper is to provide some of the first results in these directions. In particular, it will be necessary to distinguish between an *exact* nonnegative rank-r decomposition and a best nonnegative rank-r *approximation*. Note that when a best nonnegative rank-r approximation to a nonnegative tensor T is unique, it means that

$$\min_{\operatorname{rank}_+(X)\leq r} \|T - X\| \tag{8}$$

has a unique minimizer X^*. The nonnegative rank-r decomposition of X^* may not however be unique.

A nonnegative rank decomposition $X = \sum_{p=1}^{r} u_{1,p} \otimes \cdots \otimes u_{d,p} \in V_1^+ \otimes \cdots \otimes V_d^+$ is said to be *unique* if for any other nonnegative rank decomposition $X = \sum_{p=1}^{r} v_{1,p} \otimes \cdots \otimes v_{d,p}$, there is a permutation σ of $\{1,\ldots,d\}$ such that $u_{1,p} \otimes \cdots \otimes u_{d,p} = v_{1,\sigma(p)} \otimes \cdots \otimes v_{d,\sigma(p)}$ for all $p = 1,\ldots,r$.

IV. Existence and Generic Uniqueness of Rank-r Approximations

Let $V_1,\ldots,V_d$ be real vector spaces. Given a nonnegative tensor $T \in V^+$, we consider the best nonnegative rank-r approximations of T, where r is less than the nonnegative rank of T. We let

$$\delta(T) = \inf_{X\in D_r^+} \|T - X\| = \inf_{\operatorname{rank}_+(X)\leq r} \|T - X\|,$$

where $\|\cdot\|$ is the *Hilbert–Schmidt norm*, i.e., the l^2-norm given by the inner product.

Henceforth any unlabelled norm $\|\cdot\|$ on $V_1 \otimes \cdots \otimes V_d$ will always denote the Hilbert–Schmidt norm. When $d = 2$, the Hilbert–Schmidt norm reduces to the Frobenius norm of matrices and when $d = 1$, it reduces to the Euclidean norm of vectors. Also, throughout this article, the notation $\langle X, Y\rangle$ will always denote tensor contraction in all possible indices for X, Y tensors of any order [43]. When X and Y are of the same order and real, $\langle X, Y\rangle$ reduces to a real inner product and our notation is consistent with the inner product notation; in particular $\langle X, X\rangle = \|X\|^2$. When X is a d-tensor and Y is a $(d-1)$-tensor, $\langle X, Y\rangle$ is a vector — this is the only other case that will arise in our discussions below. Note however that over $\mathbb{C}$, $\langle\cdot,\cdot\rangle$ is only a symmetric bilinear form and not a complex inner product (which is a sesquilinear form).

Proposition 9: Let $C_i \subseteq V_i^+$ be a closed semialgebraic cone for $i = 1,\ldots,d$. Then $D_r^+ = \{X \in C_1 \otimes \cdots \otimes C_d : \operatorname{rank}_+(X) \leq r\}$ is a closed semialgebraic set.

Proof: It follows from [45] that the set is closed and from the Tarski–Seidenberg Theorem [19] that it is semialgebraic. □

Since D_r^+ is a closed set, for any $T \notin D_r^+$, there is some $T^* \in D_r^+$ such that $\|T - T^*\| = \delta(T)$. The following result is an analogue of [25, Th. 27] for nonnegative tensors based on [25, Corollary 18].

Proposition 10: Almost every $T \in V^+$ with nonnegative rank $> r$ has a unique best nonnegative rank-r approximation.

Proof: For any $T, T' \in V_1 \otimes \cdots \otimes V_d$, $|\delta(T) - \delta(T')| \leq \|T - T'\|$, i.e., δ is Lipschitz and thus differentiable almost everywhere in $V = V_1 \otimes \cdots \otimes V_d$ by Rademacher Theorem.

Consider a general $T \in V^+$. Then in particular T lies in the interior of V^+ and there is an open neighborhood of T contained in V^+. So δ is differentiable almost everywhere in V^+ as well. Suppose δ is differentiable at $T \in V^+$. For any $U \in V$, let $\partial\delta_T^2(U)$ be the differential of δ^2 at T along the direction U. Since $\|T - T^*\| = \delta(T)$ we obtain

$$\begin{aligned}\delta^2(T + tU) &= \delta^2(T) + t\partial\delta_T^2(U) + O(t^2)\\ &\leq \left\|T + tU - T^*\right\|^2\\ &= \delta^2(T) + 2t\langle U, T - T^*\rangle + t^2\,\|U\|^2.\end{aligned}$$

Therefore, for any t, we have $t\partial\delta_T^2(U) \leq 2t\langle U, T - T^*\rangle$, which implies that

$$\partial\delta_T^2(U) = 2\langle U, T - T^*\rangle.$$

If T' is another best nonnegative rank-r approximation of T, then

$$2\langle U, T - T^*\rangle = \partial\delta_T^2(U) = 2\langle U, T - T'\rangle,$$

from which it follows that $\langle T' - T^*, U\rangle = 0$ for any U, i.e., $T' = T^*$. □

We note that Proposition 10 holds more generally for arbitrary closed cones $C_1,\ldots,C_d$ in place of $V_1^+,\ldots,V_d^+$. Our next proposition holds true for arbitrary closed semialgebraic cones $C_1,\ldots,C_d$ in place of $V_1^+,\ldots,V_d^+$.

Proposition 11: The nonnegative tensors satisfying (i) nonnegative rank $> r$, and (ii) do not have a unique best rank-r approximation, form a semialgebraic set that is contained in some hypersurface.

Proof: Observe that D_r^+ is the image of the polynomial map

$$\begin{aligned}&\varphi_r\colon (V_1^+ \times \cdots \times V_d^+)^r \to V^+,\\ &(u_{1,1},\ldots,u_{d,1},\ldots,u_{1,r},\ldots,u_{d,r})\\ &\qquad\qquad \mapsto \sum_{j=1}^{r} u_{1,j} \otimes \cdots \otimes u_{d,j}.\end{aligned}$$

Hence D_r^+ is semialgebraic by the Tarski–Seidenberg Theorem [19] and the required result follows from [26, Th. 3.4]. □

Now we examine a useful necessary condition for $\sum_{p=1}^{r} T_p$ to be a best rank-r approximation of $T \in V_1 \otimes \cdots \otimes V_d$. For a vector $u \in V_i$, we denote by $u(j)$ the jth coordinate of u, i.e., $u = (u(1),\ldots,u(n_i))$, and we will borrow a standard notation from algebraic topology where a hat over a quantity means that quantity is omitted. So for example,

$$\begin{aligned}
\widehat{u_1} \otimes u_2 \otimes u_3 &= u_2 \otimes u_3,\\
u_1 \otimes \widehat{u_2} \otimes u_3 &= u_1 \otimes u_3,\\
u_1 \otimes u_2 \otimes \widehat{u_3} &= u_1 \otimes u_2,\\
u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d &= u_1 \otimes \cdots \otimes u_{i-1} \otimes u_{i+1}\\
&\quad \otimes \cdots \otimes u_d.
\end{aligned}$$

Let us recall the following well-known fact, which has been used to develop algorithms for nonnegative matrix factorization and nonnegative tensor decomposition.

Lemma 12: Let $V_1, \ldots, V_d$ be real vector spaces and let $T \in V_1 \otimes \cdots \otimes V_d$. Let $\operatorname{rank}(T) > r$ and $\lambda \sum_{j=1}^r T_j$ be a best rank-r approximation, where $T_j = u_{1,j} \otimes \cdots \otimes u_{d,j}$ and $\left\| \sum_{j=1}^r T_j \right\| = 1$. Then for all $i = 1, \ldots, d$, and $p = 1, \ldots, r$,

$$\begin{aligned}
&\langle T, u_{1,p} \otimes \cdots \otimes \widehat{u_{i,p}} \otimes \cdots \otimes u_{d,p} \rangle\\
&\quad = \lambda \left\langle \sum_{j=1}^r T_j, u_{1,p} \otimes \cdots \otimes \widehat{u_{i,p}} \otimes \cdots \otimes u_{d,p} \right\rangle, \qquad (9)
\end{aligned}$$

where $\lambda = \langle T, \sum_{j=1}^r T_j \rangle$.

Proof: Let L denote the line in $V_1 \otimes \cdots \otimes V_d$ spanned by $\sum_{j=1}^r v_{1,j} \otimes \cdots \otimes v_{d,j}$, and $L^\perp$ denote the orthogonal complement of L. Denote the orthogonal projection of T onto L by $\operatorname{Proj}_L(T)$. Then

$$\|T\|^2 = \left\|\operatorname{Proj}_L(T)\right\|^2 + \left\|\operatorname{Proj}_{L^\perp}(T)\right\|^2,$$

and thus

$$\begin{aligned}
&\min_{\alpha \geq 0} \left\| T - \alpha \sum_{p=1}^r v_{1,p} \otimes \cdots \otimes v_{d,p} \right\|^2\\
&\quad = \left\| T - \operatorname{Proj}_L(T) \right\|^2 = \left\| \operatorname{Proj}_{L^\perp}(T) \right\|^2\\
&\quad = \|T\|^2 - \left\| \operatorname{Proj}_L(T) \right\|^2.
\end{aligned}$$

So computing

$$\min_{v_{1,1},\ldots,v_{d,r}} \min_{\alpha \geq 0} \left\| T - \alpha \sum_{j=1}^r v_{1,j} \otimes \cdots \otimes v_{d,j} \right\|$$

is equivalent to computing

$$\max_{v_{1,1},\ldots,v_{d,r}} \operatorname{Proj}_L(T) = \max_{v_{1,1},\ldots,v_{d,r}} \left\langle T, \sum_{j=1}^r v_{1,j} \otimes \cdots \otimes v_{d,j} \right\rangle.$$

Since $\left\| \sum_{j=1}^r T_j \right\| = 1$, we must have

$$\left\langle \sum_{j=1}^r T_j, u_{1,p} \otimes \cdots \otimes \widehat{u_{i,p}} \otimes \cdots \otimes u_{d,p} \right\rangle \neq 0$$

for some p. The Jacobian matrix of the hypersurface defined by $\left\| \sum_{j=1}^r v_{1,j} \otimes \cdots \otimes v_{d,j} \right\| = 1$ has constant rank 1 around $(u_{1,1}, \ldots, u_{d,1}, \ldots, u_{1,r}, \ldots, u_{d,r})$, i.e., this real hypersurface is smooth at the point $(u_{1,1}, \ldots, u_{d,1}, \ldots, u_{1,r}, \ldots, u_{d,r})$. Hence we may consider the Lagrangian

$$\begin{aligned}
\mathcal{L} &= \left\langle T, \sum_{p=1}^r v_{1,p} \otimes \cdots \otimes v_{d,p} \right\rangle\\
&\quad - \lambda \left(\left\| \sum_{p=1}^r v_{1,p} \otimes \cdots \otimes v_{d,p} \right\| - 1 \right). \qquad (10)
\end{aligned}$$

Setting $\partial \mathcal{L} / \partial v_{i,p} = 0$ at $(u_{1,1}, \ldots, u_{d,1}, \ldots, u_{1,r}, \ldots, u_{d,r})$ gives

$$\begin{aligned}
&\langle T, u_{1,p} \otimes \cdots \otimes \widehat{u_{i,p}} \otimes \cdots \otimes u_{d,p} \rangle\\
&\quad = \lambda \left\langle \sum_{j=1}^r T_j, u_{1,p} \otimes \cdots \otimes \widehat{u_{i,p}} \otimes \cdots \otimes u_{d,p} \right\rangle \qquad (11)
\end{aligned}$$

with $\lambda = \left\langle T, \sum_{j=1}^r T_j \right\rangle$ for all $i = 1, \ldots, d,\ p = 1, \ldots, r$. □

Lemma 12 has a nice geometric interpretation as follows. Let $\widehat{\sigma}_r(\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_d)$ be the cone of the rth secant variety of the Segre variety $\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_d$. Suppose $\lambda \sum_{j=1}^r T_j$ is a smooth point. Then $T - \lambda \sum_{j=1}^r T_j$ is perpendicular to the tangent space of $\widehat{\sigma}_r(\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_d)$ at $\lambda \sum_{j=1}^r T_j$.

We presented Lemma 12 in a concrete affine (as opposed to projective) manner so that there will be no ambiguity when discussing λ and $u_{i,j}$. We will see later in Definition 17 that when $r = 1$, these are *normalized singular values* and *normalized singular vector tuples* of T.

For a nonnegative tensor T with $\operatorname{rank}_+(T) > r$, we have an inequality in place of the equality in (9). First we define the *support* of a vector $v \in V$ to be

$$\operatorname{supp}(v) := \{i \in \{1, \ldots, \dim V\} : v_i \neq 0\}.$$

Lemma 13: Let $T \in V^+$ with $\operatorname{rank}_+(T) > r$ and $X = \sum_{p=1}^{r'} u_{1,p} \otimes \cdots \otimes u_{d,p}$ be a solution of the optimization problem (8)*. Then*

$$\begin{aligned}
&\langle T, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,p} \rangle\\
&\quad \leq \left\langle X, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,p} \right\rangle \qquad (12)
\end{aligned}$$

where $v_{i,p} \in V_i^+$, $i = 1, \ldots, d$, and $p = 1, \ldots, r'$. For each pair (i, p), consider the subspace

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

Then

$$\begin{aligned}
&\langle T, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,p} \rangle\\
&\quad = \left\langle X, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,p} \right\rangle \qquad (13)
\end{aligned}$$

for $v_{i,p} \in \widetilde{V}_{i,p}$.

Proof: Fix a pair (i, p) and consider a curve $X(t) = u_{1,p} \otimes \cdots \otimes (u_{i,p} + t v_{i,p}) \otimes \cdots \otimes u_{d,p} + \sum_{j \neq p} u_{1,j} \otimes \cdots \otimes u_{d,j}$, where $v_{i,p} \in V_i^+$. Since for $t \geq 0$, $\|T - X(t)\|$ achieves a local minimum at $t = 0$, i.e., nondecreasing in $[0, \varepsilon)$ for some small $\varepsilon > 0$, the right derivative

$$\lim_{t \to 0+} \frac{d}{dt} \|T - X(t)\| \geq 0.$$

In other words, we have

$$\begin{aligned}
&\langle T, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,p} \rangle\\
&\quad \leq \left\langle X, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,p} \right\rangle.
\end{aligned}$$

In particular, if $v_{i,p} \in \widetilde{V}_{i,p}$, $X(t)$ is nonnegative for $t \in (-\varepsilon, \varepsilon)$, then the local minimality of $\|T - X(t)\|$ at 0 implies that

$$\left. \frac{d}{dt} \|T - X(t)\| \right|_{t=0} = 0,$$

which gives us

$$\langle T, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,p}\rangle = \langle X, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,p}\rangle,$$

as required. □

Recall that a choice of bases is always implicit when we discuss V^+ (cf. Definition 2) and we may refer to coordinates (or entries) of a nonnegative tensor T without ambiguity.

Lemma 14: Let $T \in V^+$ with $\operatorname{rank}_+(T) > r$ and X be a solution of the optimization problem (8). *Then there exist $i_1, \dots, i_d$ such that the coordinate $(T - X)_{i_1,\dots,i_d} > 0$.*

Proof: Let $X = \sum_{p=1}^{r'} u_{1,p} \otimes \cdots \otimes u_{d,p}$. Suppose $(T - X)_{i_1,\dots,i_d} \le 0$ for all $i_1, \dots, i_d$. Then there is some $p \in \{1, \dots, r'\}$ such that $u_{1,p}(i_1) \cdots u_{d,p}(i_d) > 0$. So

$$\langle T - X, u_{1,p} \otimes \cdots \otimes u_{d,p}\rangle \le (T - X)_{i_1,\dots,i_d} u_{1,p}(i_1) \cdots u_{d,p}(i_d) < 0,$$

which contradicts (13). □

Proposition 15: Let $T \in V^+$ with $\operatorname{rank}_+(T) > r$ and X be a solution to the optimization problem (8). Then $\operatorname{rank}_+(X) = r$.

Proof: Suppose that $\operatorname{rank}_+(X) \le r - 1$. By Lemma 14 there is some coordinate $(T - X)_{i_1,\dots,i_d} > 0$. Let X' be the rank-one tensor whose only nonzero coordinate $X'_{i_1,\dots,i_d} = (T - X)_{i_1,\dots,i_d}$. Then $\|T - X - X'\| < \|T - X\|$ and $\operatorname{rank}_+(X + X') \le r$, which contradicts X being a solution of (8). □

Proposition 15 shows that a solution X of (8) indeed has nonnegative rank exactly r; so it is in fact appropriate to call X a best nonnegative rank-r approximation of T.

V. Rank-One Approximations for Nonnegative Tensors and the Perron–Frobenius Theorem

We have established in Section IV that a best nonnegative rank-r approximation of a nonnegative tensor is generically unique. In this section we focus on the case $r = 1$ and find sufficient conditions that guarantee the uniqueness of best nonnegative rank-one approximations. We begin with the following simple but useful observation: For a nonnegative tensor, a best rank-one approximation can always be chosen to be a best nonnegative rank-one approximation.

Theorem 16: Given $T \in V^+$, let $u_1 \otimes \cdots \otimes u_d \in V_1 \otimes \cdots \otimes V_d$ be a best rank-one approximation of T. Then $u_1, \dots, u_d$ can be chosen to be nonnegative, i.e., $u_1 \in V_1^+, \dots, u_d \in V_d^+$.

Proof: Let $T = (T_{i_1,\dots,i_d})$ and $u_i = (u_i(1), \dots, u_i(n_i))$. Then

$$\begin{aligned}\|T - u_1 \otimes \cdots \otimes u_d\|^2 &= \sum_{i_1,\dots,i_d=1}^{n_1,\dots,n_d} \big(T_{i_1,\dots,i_d} - u_1(i_1) \cdots u_d(i_d)\big)^2 \\ &\ge \sum_{i_1,\dots,i_d=1}^{n_1,\dots,n_d} \big(T_{i_1,\dots,i_d} - |u_1(i_1)| \cdots |u_d(i_d)|\big)^2.\end{aligned}$$

Since $u_1 \otimes \cdots \otimes u_d$ is a best rank-one approximation, we can choose $u_j(i_j) = |u_j(i_j)|$, i.e., $u_1 \in V_1^+, \dots, u_d \in V_d^+$. □

By Theorem 16, there is no need to distinguish between a best rank-one and a best nonnegative rank-one approximation of a nonnegative tensor. This allows us to treat best rank-one approximations of a real tensor in a unified way, i.e., we will look for sufficient conditions to ensure a unique best rank-one approximation of a real tensor. Motivated in part by the notion of singular pairs of a tensor [44] and by the case $r = 1$ in Lemma 12, we propose the following definition.

Definition 17: Let $V_1, \dots, V_d$ be vector spaces over $\mathbb{K}$ of dimensions $n_1, \dots, n_d$. For $T \in V_1 \otimes \cdots \otimes V_d$, we call $(\lambda, u_1, \dots, u_d) \in \mathbb{K} \times V_1 \times \cdots \times V_d$ a *normalized singular pair* of T if

$$\begin{cases}\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d\rangle = \lambda u_i, \\ \langle u_i, u_i\rangle = 1,\end{cases} \tag{14}$$

for all $i = 1, \dots, d$. We call λ a *normalized singular value* and $(u_1, \dots, u_d)$ is called a *normalized singular vector tuple* corresponding to λ. If $\mathbb{K} = \mathbb{R}$, $\lambda \ge 0$, and $u_i \in V_i^+$, we call $(\lambda, u_1, \dots, u_d)$ a *nonnegative normalized singular pair* of T.

The reader is reminded that the contraction product $\langle\cdot,\cdot\rangle$ is only an inner product over $\mathbb{R}$ but not $\mathbb{C}$. In particular, $\langle u, u\rangle \ne \|u\|^2$ over $\mathbb{C}$. In Definition 17 we require that $\langle u_i, u_i\rangle = 1$ instead of $\|u_i\| = 1$ because $\langle u_i, u_i\rangle = 1$ is an algebraic condition, i.e., it is defined by a polynomial equation. However imposing the condition $\langle u_i, u_i\rangle = 1$ would exclude *isotropic* complex singular vector tuples with $\langle u_i, u_i\rangle = 0$ — note that over $\mathbb{C}$ this can happen for $u_i \ne 0$. As such, the following projective variant introduced in [25] is useful when we would like to include such isotropic cases.

Definition 18: Let $W_1, \dots, W_d$ be complex vector space. For $T \in W_1 \otimes \cdots \otimes W_d$, $([u_1], \dots, [u_d]) \in \mathbb{P}W_1 \times \cdots \times \mathbb{P}W_d$ is called a *projective singular vector tuple* if

$$\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d\rangle = \lambda_i u_i \tag{15}$$

for some $\lambda_i \in \mathbb{C}$, $i = 1, \dots, d$.

The number of projective singular vector tuples of a generic tensor has been calculated in [25]. In the sense of [22], this number is the Euclidean distance degree of the Segre variety.

Note that as Definition 18 is over projective spaces, the λ_i's are not well-defined complex numbers, and neither is $\prod_{i=1}^d \lambda_i$, but this product corresponds in an appropriate sense to a singular value as we will see next.

Definitions 17 and 18 are related over $\mathbb{C}$ as follows. Suppose $([u_1], \dots, [u_d]) \in \mathbb{P}W_1 \times \cdots \times \mathbb{P}W_d$ is a projective singular vector tuple. We first choose a representative $(u_1, \dots, u_d)$ of $([u_1], \dots, [u_d])$ that satisfies (15) and has $\|u_i\| = 1$. Note that we may assume $\prod_{i=1}^d \lambda_i$ to be a nonnegative real number: If $(v_1, \dots, v_d)$ is such that $v_j = e^{i\theta_j} u_j$, then $\langle T, v_1 \otimes \cdots \otimes \widehat{v_j} \otimes \cdots \otimes v_d\rangle = \mu_j v_j$ and we may choose appropriate $\theta_1, \dots, \theta_d$ so that

$$\prod_{i=1}^d \mu_i = e^{i(d-2)(\theta_1+\cdots+\theta_d)} \prod_{i=1}^d \lambda_i \in \mathbb{R}_+.$$

For a nonnegative $\prod_{i=1}^d \lambda_i$,

$$\lambda := \Big(\prod_{i=1}^d \lambda_i\Big)^{1/d}$$

is 'almost' a normalized singular value of T with corresponding normalized singular vector tuple $(u_1, \dots, u_d)$ — 'almost' because the condition $\langle u_i, u_i\rangle = 1$ in Definition 17 has to be replaced by $\|u_i\| = 1$.

It has been shown in [25] that a generic T does not have a zero singular value nor a projective singular vector tuple $([u_1], \ldots, [u_d])$ such that $\langle u_i, u_i \rangle = 0$ for some i. Thus, for a generic T, both definitions above are equivalent. We may use the two definitions interchangeably depending on the situation. In this article, we will mainly consider the normalized singular pairs of a tensor as defined in Definition 17.

The next three results give an analogue of the tensorial Perron–Frobenius Theorem [8], [24], [44], [61] for nonnegative normalized singular pairs (as opposed to nonnegative eigenpairs [44]). The proof of Lemma 19 in particular will require the l^1-norm. Again recall that a choice of bases is always implicit when we discuss V^+ (cf. Definition 2) and the l^1-norm is with respect to this choice of bases.

Lemma 19 (Existence): A nonnegative tensor $T \in V^+$ has at least one nonnegative normalized singular pair.

Proof: Consider the compact convex set

$$D = \left\{ (u_1, \ldots, u_d) \in V_1^+ \times \cdots \times V_d^+ : \sum_{i=1}^d \|u_i\|_1 = 1 \right\}.$$

If $\sum_{i=1}^d \|\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle\|_1 = 0$ for some $(u_1, \ldots, u_d)$, then $\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle = 0$ for all i, which implies that $\lambda = 0$. On the other hand, if $\sum_{i=1}^d \|\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle\|_1 > 0$, we define the map $\psi : D \to D$ by

$$\psi(u_1, \ldots, u_d) = \left(\frac{\langle T, u_2 \otimes \cdots \otimes u_d \rangle}{\sum_{i=1}^d \|\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle\|_1}, \ldots \right.$$
$$\left. \ldots, \frac{\langle T, u_1 \otimes \cdots \otimes u_{d-1} \rangle}{\sum_{i=1}^d \|\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle\|_1} \right).$$

Note that each term $\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle$ in the denominator is the contraction of a d-tensor with a $(d-1)$-tensor and therefore the result is a vector. We then normalize by the sum of the l^1-norms of these vectors so that $\|\psi\|_1 = 1$.

By Brouwer's Fixed Point Theorem, there is some $u_1 \otimes \cdots \otimes u_d$ such that $\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle = \lambda u_i$ where

$$\lambda = \sum_{i=1}^d \|\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle\|_1.$$

Since $\langle T, u_1 \otimes \cdots \otimes u_d \rangle = \lambda \|u_i\|^2$ for $i = 1, \ldots, d$, $\|u_1\| = \cdots = \|u_d\|$. Let $u_i' = u_i / \|u_i\|$ and $\lambda' = \langle T, u_1' \otimes \cdots \otimes u_d' \rangle$. Then $(\lambda', u_1', \ldots, u_d')$ is a nonnegative normalized singular pair. □

One of our reviewers has pointed out to us that Lemma 19 may also be obtained from Lemma 12 and Theorem 16.

Definition 20: We say that a tensor $T \in V^+$ is *positive* if all its coordinates (with respect to the implicit choice of bases when we specify V^+, cf. Definition 2) are positive.

Lemma 21 (Positivity): If T is positive, then T has a positive normalized singular pair $(\lambda, u_1, \ldots, u_d)$ with $\lambda > 0$.

Proof: By Lemma 19, T has a nonnegative normalized singular pair $(\lambda, u_1, \ldots, u_d)$. Suppose that a choice of bases has been fixed for $V_1, \ldots, V_d$. We let $v_i(j)$ denote the jth coordinate of a vector $v_i \in V_i$, $j = 1, \ldots, n_i$. Let

$$\alpha = \min\{u_i(j) : i = 1, \ldots, d,\ j \in \operatorname{supp}(u_i)\}.$$

For any i and j,

$$\begin{aligned} \lambda u_i(j) &= \langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle(j) \\ &\geq \alpha^{d-1} \sum_{k_j \in \operatorname{supp}(u_j)} T_{k_1 \ldots k_{i-1} j k_{i+1} \ldots k_d} > 0, \end{aligned}$$

implying that λ and all coordinates of u_i are positive. □

We recall the definition of *spectral norm* for a tensor, which is known [31] to be NP-hard to compute or even approximate.

Definition 22: 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.

We may deduce the following from [25, Th. 20] and Lemma 12.

Corollary 23 (Generic Uniqueness): A general real tensor T has a unique normalized singular pair $(\lambda, u_1, \ldots, u_d)$ with $\lambda = \|T\|_\sigma$.

The relation between best rank-r and best rank-one approximations of a matrix over $\mathbb{R}$ or $\mathbb{C}$ is well-known: A best rank-r approximation can be obtained from r successive best rank-one approximations—a consequence of the Eckart–Young Theorem. It has been shown in [55] that this 'deflation procedure' does not work for real or complex d-tensors of order $d > 2$. In fact, more recently, it has been shown in [58] that the property almost never holds when $d > 2$.

We will see here that the 'deflatability' property does not hold for nonnegative tensor rank either.

Proposition 24: A best nonnegative rank-r approximation of a positive tensor with nonnegative rank $> r$ cannot be obtained by a sequence of best nonnegative rank-one approximations.

Proof: It suffices to show that a best nonnegative rank-2 approximation cannot be obtained by two best nonnegative rank-one approximations. Let $T \in V^+$ be a positive tensor with $\operatorname{rank}_+(T) > 2$. Suppose $u_1 \otimes \cdots \otimes u_d$ is a best rank-one approximation of T, and $u_1 \otimes \cdots \otimes u_d + v_1 \otimes \cdots \otimes v_d$ is a best nonnegative rank-2 approximation of T. By the proof of Lemma 21, $u_k > 0$ for all $k = 1, \ldots, d$, then by Lemma 13, we have

$$\langle T - u_1 \otimes \cdots \otimes u_d,\ u_1 \otimes \cdots \otimes u_d \rangle = 0,$$
$$\langle T - u_1 \otimes \cdots \otimes u_d - v_1 \otimes \cdots \otimes v_d,\ u_1 \otimes \cdots \otimes u_d \rangle = 0.$$

We subtract the second equation from the first to get

$$\langle v_1 \otimes \cdots \otimes v_d,\ u_1 \otimes \cdots \otimes u_d \rangle = 0,$$

which contradicts the non-negativity of each v_k and the positivity of each u_k. □

Following [58], we say that a tensor $T \in V^+$ with nonnegative rank s admits a *Schmidt–Eckart–Young decomposition* if it can be written as a linear combination of nonnegatively decomposable tensors $T = \sum_{p=1}^s u_{1,p} \otimes \cdots \otimes u_{d,p}$, and such that $\sum_{p=1}^r u_{1,p} \otimes \cdots \otimes u_{d,p}$ is a best nonnegative rank-r approximation of T for all $r = 1, \ldots, s$. Proposition 24 shows that a general nonnegative tensor does not admit a Schmidt–Eckart–Young decomposition.

We point out that methods in [18], [48], and [49] (for real/complex) [12], [35], and [63] (nonnegative) rely on deflation.

VI. Uniqueness of Best Rank-One Approximations for Real Symmetric Tensors

Not every tensor has a unique best rank-one approximation [55, Proposition 1]. For example, the symmetric 3-tensor $x \otimes x \otimes x + y \otimes y \otimes y$, where x and y are orthonormal, has two best rank-one approximations: $x \otimes x \otimes x$ and $y \otimes y \otimes y$. It is known that a best rank-one approximation of a symmetric tensor can be chosen to be symmetric over $\mathbb{R}$ and $\mathbb{C}$ [3], [23]. In this section we study various properties of the set of symmetric tensors that do not have unique best symmetric rank-one approximations. Before we get to these we will have to first introduce analogues/generalizations of *eigenpairs* and *characteristic polynomials* for higher-order symmetric tensors.

In the following, for a real or complex vector space V, $\mathsf{S}^d(V)$ denotes the symmetric d-tensors over V. For any $u \in V$, we write $u^{\otimes d} = u \otimes \cdots \otimes u \in \mathsf{S}^d(V)$ for the d-fold tensor product of u with itself.

Let V^* be the dual space of V. For any group G acting on V, G also acts naturally on $\mathsf{S}^d(V)$ and $\mathsf{S}^d(V^*)$ such that

$$\langle S, T\rangle = \langle g \cdot S, g \cdot T\rangle$$

for all $g \in G$, $T \in \mathsf{S}^d(V)$, and $S \in \mathsf{S}^d(V^*)$. If we fix an inner product $(\cdot,\cdot)$ on V, then V becomes self dual and we may identify $V^* = V$. In which case $\langle\cdot,\cdot\rangle$ may be regarded the inner product on $\mathsf{S}^d(V)$ defined by

$$\langle u^{\otimes d}, v^{\otimes d}\rangle := (u, v)^d$$

and extended linearly to any $S, T \in \mathsf{S}^d(V)$ (since any element of $\mathsf{S}^d(V)$ may be expressed as a linear combination of $u^{\otimes d}$'s [15]). The inner product $\langle\cdot,\cdot\rangle$ is clearly invariant under the group that preserves the inner product $(\cdot,\cdot)$. In particular, if $V = \mathbb{R}^n$, then $\langle\cdot,\cdot\rangle$ is invariant under the orthogonal group[1] $\mathrm{O}(n)$.

The following definition of symmetric tensor eigenpairs is based on [7], [44], and [50].

Definition 25: For $T \in \mathsf{S}^d(V)$ over $\mathbb{C}$, $(\lambda, u) \in \mathbb{C} \times V$ is called a *normalized eigenpair* of T if

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

λ is the *normalized eigenvalue* and v the corresponding *normalized eigenvector* of T. Two normalized eigenpairs (λ, u) and (μ, v) of T are *equivalent* if $(\lambda, u) = (\mu, v)$ or if $(-1)^{d-2}\lambda = \mu$ and $u = -v$. A normalized eigenvalue λ is said to be *simple* if it has only one corresponding normalized eigenvector up to equivalence.

The number of eigenpairs of a tensor over $\mathbb{C}$ has been determined in [7] and [47]; one may view this as the ED degree of the Veronese variety [22]. Definition 25 also applies to a real vector space V. In this case, normalized eigenpairs of $T \in \mathsf{S}^d(V)$ are invariant $\mathrm{O}(n)$.

It is easy to see that for a symmetric tensor $T \in \mathsf{S}^d(V)$, the spectral norm $\|T\|_\sigma$ is the largest eigenvalue of T in absolute value. Let $\mathbb{S}^{n-1}$ denote the unit sphere in $\mathbb{R}^n$. The subset $\{u \in \mathbb{S}^{n-1} : \langle T, u^{\otimes d}\rangle = \|T\|_\sigma\}$ is non-empty and closed in $\mathbb{S}^{n-1}$ and invariant under $\mathrm{O}(n)$.

To introduce the characteristic polynomial of a symmetric tensor, we first recall the definition and some basic properties of the *multipolynomial resultant* [17], [29]. For any given $n+1$ homogeneous polynomials $F_0, \dots, F_n \in \mathbb{C}[x_0, \dots, x_n]$ with positive total degrees $d_0, \dots, d_n$, let $F_i = \sum_{|\alpha|=d_i} c_{i,\alpha} x_0^{\alpha_0} \cdots x_n^{\alpha_n}$, where $\alpha = (\alpha_0, \dots, \alpha_n)$ and $|\alpha| = \alpha_0 + \cdots + \alpha_n$. We will associate each pair (i, α) with a variable $u_{i,\alpha}$. Now given a polynomial P in the variables $u_{i,\alpha}$ where $i = 0, \dots, n$, and $|\alpha| \in \{d_0, \dots, d_n\}$, we denote by $P(F_0, \dots, F_n)$ the result obtained by substituting each $u_{i,\alpha}$ in P with $c_{i,\alpha}$. The following is a classical result in invariant theory [17], [29].

Theorem 26: There is a unique polynomial Res *with integer coefficients in the variables $u_{i,\alpha}$ where $i = 0, \dots, n$, and $|\alpha| \in \{d_0, \dots, d_n\}$, that has the following properties:*

(i) $F_0 = \cdots = F_n = 0$ *have a nonzero solution over* $\mathbb{C}$ *if and only if* $\mathrm{Res}(F_0, \dots, F_n) = 0$.

(ii) $\mathrm{Res}(x_0^{d_0}, \dots, x_n^{d_n}) = 1$.

(iii) Res *is irreducible over* $\mathbb{C}$.

Definition 27: $\mathrm{Res}(F_0, \dots, F_n) \in \mathbb{C}$ is called the *resultant* of the polynomials $F_0, \dots, F_n$. Often we will also say that it is the resultant of the system of polynomial equations $F_0 = 0, \dots, F_n = 0$.

The following definition was first proposed in [51] and called an E-characteristic polynomial.

Definition 28: The *characteristic polynomial* of a symmetric tensor T is the resultant $\psi_T(\lambda)$ of the following systems of polynomial equations in $n+1$ variables u and x (note that u has n entries). (i)

1) For $T \in \mathsf{S}^{2d-1}(V)$,

$$\langle T, u^{\otimes(d-1)}\rangle - \lambda x^{d-2} u = 0 \text{ and } x^2 - \langle u, u\rangle = 0.$$

2) For $T \in \mathsf{S}^{2d}(V)$,

$$\langle T, u^{\otimes(2d-1)}\rangle - \lambda \langle u, u\rangle^{d-1} u = 0.$$

Note that we regard λ as a parameter and not one of the variables. One may show that the resultant $\psi_T(\lambda)$ is a (univariate) polynomial in λ.

In the following, for $u, v, w \in V$, we write

$$u \odot v \odot w := \frac{1}{6}(u \otimes v \otimes w + u \otimes w \otimes v + v \otimes u \otimes w + v \otimes w \otimes u + w \otimes v \otimes u + w \otimes u \otimes v)$$

for the *symmetric tensor product* [15]. Note that $u \odot v \odot w = v \odot u \odot w = \cdots = w \odot v \odot u$, i.e., symmetric tensor product is independent of order and in particular $u \odot v \odot w \in \mathsf{S}^3(V)$. It is easy to extend this to arbitrary order

$$u_1 \odot \cdots \odot u_d = \frac{1}{d!} \sum_{\tau \in \mathfrak{S}_d} u_{\tau(1)} \otimes \cdots \otimes u_{\tau(d)} \in \mathsf{S}^d(V).$$

For $u \in V$, we may write $u^{\odot d} = u \odot \cdots \odot u$ for the d-fold symmetric tensor product of u with itself but we clearly always have

$$u^{\otimes d} = u^{\odot d}.$$

[1] Henceforth we assume that our vector spaces are equipped with inner products and we write $\mathrm{O}(n)$ for the group that preserves the inner product.

Proposition 29: Let V be a real vector space of dimension n. Let $\rho = \|T\|_\sigma$ and define

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

Then H_ρ is an algebraic hypersurface in $\mathsf{S}^d(V)$.

Proof: For notation convenience, we prove the result for $d = 3$; extending to $d > 3$ is straightforward. Let $T \in \mathsf{S}^3(V)$. Suppose $T \in H_\rho$, i.e., there exist $u \neq v \in V$ with $\|u\| = \|v\| = 1$ such that

$$\langle T, u^{\otimes 2}\rangle = \rho u, \quad \langle T, v^{\otimes 2}\rangle = \rho v.$$

Let $u_1 := u$ and extend it to $\{u_1, \ldots, u_n\}$, an orthonormal basis of V. By an action of the orthogonal group $\mathrm{O}(n)$ on V, we may assume that $v = u_1 \cos\theta + u_2 \sin\theta$ for some $\theta \in (0, \pi)$. Let $T_{ijk} := \langle T, u_i \odot u_j \odot u_k\rangle$. Then

$$\begin{cases} T_{111} = \rho, & \\ T_{i11} = 0, & (16)\\ T_{111}\cos^2\theta + T_{122}\sin^2\theta = T_{111}\cos\theta, & (17)\\ 2T_{122}\sin\theta\cos\theta + T_{222}\sin^2\theta = T_{111}\sin\theta, & \\ 2T_{j12}\cos\theta + T_{j22}\sin\theta = 0, & \end{cases}$$

for $i \neq 1$ and $j > 2$.

By eliminating θ, we may obtain equations in T_{ijk}'s. For example, (17) implies $\cos\theta = 1$ or $(T_{111} - T_{122})\cos\theta = T_{122}$, and (18) implies $\sin\theta = 0$ or $2T_{122}\cos\theta + T_{222}\sin\theta = T_{111}$. Since $\sin^2\theta + \cos^2\theta = 1$ and $\theta \neq 0$ or π, we have

$$\begin{cases} [T_{111}(T_{111} - T_{122}) - 2T_{122}^2]^2 + T_{222}^2 T_{122}^2 = T_{222}^2(T_{111} - T_{122})^2, \\ (T_{111}T_{122} + 2T_{122}^2 - T_{111}^2)T_{j22} = 2T_{j12}T_{222}(T_{111} - T_{122}). \end{cases} \tag{18}$$

Let $J := \{(T, [u_1, \ldots, u_n]) \in \mathsf{S}^3(V) \times \mathrm{O}(n) : T_{ijk}$ satisfies (18)$\}$. Consider the projections

$$\begin{tikzcd} & J \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ \mathsf{S}^3(V) & & \mathrm{O}(n) \end{tikzcd} \tag{19}$$

where $\pi_1(T, [u_1, \ldots, u_n]) = T$ and $\pi_2(T, [u_1, \ldots, u_n]) = [u_1, \ldots, u_n]$. By [51], ρ is a root of the E-characteristic polynomial $\psi_T(\lambda)$ of T. So ρ and any of its corresponding normalized eigenvectors must depend algebraically on T, implying that J is a variety in $\mathsf{S}^3(V) \times \mathrm{O}(n)$. Hence T has more than one inequivalent normalized eigenvectors corresponding to ρ if and only if T is in the image of π_1, i.e., $H_\rho = \pi_1(J)$.

Now define $T' \in \mathsf{S}^3(V)$ by

$$T'_{111} = 1, \quad T'_{122} = 2\sqrt{3} - 3, \ T'_{222} = 6\sqrt{3} - 10,$$

and set all other terms $T'_{ijk} = 0$. Then T' has two normalized eigenvectors corresponding to its normalized eigenvalue $\rho = \|T'\|_\sigma = 1$. Hence $T' \in \pi_1(J)$. Since T' has a finite number of eigenvectors, a generic $T \in \pi_1(J)$ must also have a finite number of eigenvectors by semicontinuity. Hence $\dim \pi_1^{-1}(T) = \dim \mathrm{O}(n-2)$ for a generic $T \in \pi_1(J)$. So $\dim H_\rho = \dim \pi_1(J) = \dim J - \dim \mathrm{O}(n-2) = \dim J - (n-2)(n-3)/2$.

Since π_2 is a dominant morphism, and the dimension of a generic fiber $\pi_2^{-1}([u_1, \ldots, u_n])$ is $\dim \mathsf{S}^3(V) - 2(n-1)$, we deduce that $\dim J = \dim \mathsf{S}^3(V) - 2(n-1) + \dim \mathrm{O}(n)$. Therefore $\dim H_\rho = \dim \mathsf{S}^3(V) - 1$, i.e., H_ρ is a hypersurface. $\square$

Let V be a real vector space. We specify a choice of basis on V and define the set of nonnegative symmetric tensors to be

$$\mathsf{S}^d(V^+) := \mathsf{S}^d(V) \cap (V^{\otimes d})^+.$$

Recall also Definition 20.

Corollary 30: Let $T \in \mathsf{S}^3(V^+)$ be positive. Let $u \in V$ be such that $\langle T, u^{\otimes 3}\rangle = \rho = \|T\|_\sigma$ and

$$\sigma_2 := \min\{|\langle T, u \odot v \odot v\rangle| : \langle u, v\rangle = 0, \ \|v\| = 1\}.$$

If $\sigma_2 \geq \rho/2$, then T has a unique best nonnegative symmetric rank-one approximation.

Proof: By Lemma 12, suppose there exist $v \neq u$ such that $\|v\| = 1$, $\langle u, v\rangle = 0$, and $\langle T, (u\cos\theta + v\sin\theta)^{\otimes 3}\rangle = \rho$ for some $0 < \theta \leq \pi$. Then by Lemma 21, we must in fact have $0 < \theta < \pi/2$. By (17), $\langle T, u \odot v \odot v\rangle = \frac{\cos\theta}{1+\cos\theta}\rho$. Since $0 < \frac{\cos\theta}{1+\cos\theta} < \frac{1}{2}$ when $0 < \theta < \pi/2$, we get $0 < \langle T, u \odot v \odot v\rangle < \rho/2$, which contradicts $\sigma_2 \geq \rho/2$. $\square$

Let V be a real vector space of dimension n and $W = V \otimes_{\mathbb{R}} \mathbb{C}$ be its complexification. A generic $T \in \mathsf{S}^d(W)$ has distinct eigenvalues [7], so the resultant of the polynomial ψ_T and its derivative ψ'_T, denoted by $D_{\mathrm{eig}}(T)$, is a nonzero polynomial on $\mathsf{S}^d(W)$ called the *eigen discriminant*. The equation $D_{\mathrm{eig}}(T) = 0$ defines the complex hypersurface H_{disc} consisting of tensors $T \in \mathsf{S}^d(W)$ that do not have simple normalized eigenpairs. For $T \in \mathsf{S}^d(V)$, the hypersurface H_ρ in Proposition 29 is a union of some components of the real points of H_{disc}. In fact, if we replace $\rho = \|T\|_\sigma$ by any real normalized eigenvalue μ of T in the proof of Proposition 29, we may show that the subset of symmetric tensors whose normalized eigenvalues are not all simple is a finite union of real algebraic hypersurfaces, and these hypersurfaces are the real points of H_{disc}. We summarize this discussion as follows.

Theorem 31: $D_{\mathrm{eig}}(T) = 0$ is a defining equation of the hypersurface

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

For $T \in \mathsf{S}^d(V)$, if $D_{\mathrm{eig}}(T) \neq 0$, then by definition, either (i) there is a unique eigenvector v_λ corresponding to each eigenvalue λ of T when d is odd, or (ii) there are two eigenvectors $\pm v_\lambda$ corresponding to each eigenvalue λ of T when d is even. Hence we have the following.

Corollary 32: Let $T \in \mathsf{S}^d(V)$. If $D_{\mathrm{eig}}(T) \neq 0$, then T has a unique best symmetric rank-one approximation.

We deduce the following analogue for nonnegative tensors from Banach's Theorem that the best rank-one approximation of a symmetric tensor can be chosen to be symmetric [3], [23], Theorem 16, and Corollary 32.

Corollary 33: Let $T \in \mathsf{S}^d(V^+)$. If $D_{\mathrm{eig}}(T) \neq 0$, then T has a unique best symmetric nonnegative rank-one approximation.

Let $X \subset \mathbb{C}^n$ be a complex variety. For $x \in X$ and $u \notin X$, let $d_u(x) = \sum_{i=1}^n (u(i) - x(i))^2$. The *Euclidean distance degree* (ED degree) of X is the number of nonsingular critical points of d_u for a generic u, and the *ED discriminant* is the set of u such that at least two critical points of d_u coincide [22]. Hence Theorem 31 shows that the ED discriminant of the cone

$$G = \begin{bmatrix} T_{111} & T_{122} & 0 & 2T_{112} & -\lambda & 0 \\ T_{112} & T_{222} & 0 & 2T_{122} & 0 & -\lambda \\ 1 & 1 & -1 & 0 & 0 & 0 \\ 12T_{122} & 4T_{111}\lambda - 8T_{122}\lambda & 4T_{111}\lambda + 4T_{122}\lambda & 8T_{222}\lambda - 16T_{112}\lambda & 16T_{112}^2 - 4\lambda^2 - 16T_{111}T_{122} & 8T_{112}T_{122} - 8T_{111}T_{222} \\ 4T_{222}\lambda - 8T_{112}\lambda & 12T_{112}\lambda & 4T_{112}\lambda + 4T_{222}\lambda & 8T_{111}\lambda - 16T_{122}\lambda & 8T_{112}T_{122} - 8T_{111}T_{222} & 16T_{122}^2 - 4\lambda^2 - 16T_{112}T_{222} \\ 8T_{112}^2 - 8T_{111}T_{122} - 2\lambda^2 & 8T_{122}^2 - 8T_{222}T_{112} - 2\lambda^2 & -6\lambda^2 & 8T_{112}T_{122} - 8T_{111}T_{222} & 8T_{122}\lambda + 8T_{111}\lambda & 8T_{112}\lambda + 8T_{222}\lambda \end{bmatrix} \tag{20}$$

over the Veronese variety (in both the real and complex case) is a hypersurface, and $D_{\text{eig}}(T) = 0$ gives its defining equation.

Example 34: Let $T = [T_{ijk}] \in \mathsf{S}^3(\mathbb{R}^2)$. Then $\psi_T(\lambda)$ is the resultant of the polynomials

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

Consider the Jacobian determinant J of F_0, F_1, F_2. Then

$$\begin{aligned} J &= \det \begin{bmatrix} \partial F_0/\partial x & \partial F_0/\partial y & \partial F_0/\partial z \\ \partial F_1/\partial x & \partial F_1/\partial y & \partial F_1/\partial z \\ \partial F_2/\partial x & \partial F_2/\partial y & \partial F_2/\partial z \end{bmatrix} \\ &= (8T_{112}^2 - 8T_{111}T_{122} - 2\lambda^2)x^2z + 4T_{112}\lambda y^3 \\ &\quad + (8T_{122}^2 - 8T_{222}T_{112} - 2\lambda^2)y^2z + 4T_{122}\lambda x^3 \\ &\quad + (4T_{111}\lambda - 8T_{122}\lambda)xy^2 + (4T_{122}\lambda + 4T_{111}\lambda)xz^2 \\ &\quad + (4T_{112}\lambda + 4T_{222}\lambda)yz^2 + (4T_{222}\lambda - 8T_{112}\lambda)x^2y \\ &\quad - 2\lambda^2z^3 + (8T_{112}T_{122} - 8T_{111}T_{222})xyz, \\ \frac{\partial J}{\partial x} &= 12T_{122}\lambda x^2 + (4T_{111}\lambda - 8T_{122}\lambda)y^2 \\ &\quad + (4T_{111}\lambda + 4T_{122}\lambda)z^2 + (8T_{222}\lambda - 16T_{112}\lambda)xy \\ &\quad + (16T_{112}^2 - 4\lambda^2 - 16T_{111}T_{122})xz \\ &\quad + (8T_{112}T_{122} - 8T_{111}T_{222})yz, \\ \frac{\partial J}{\partial y} &= (4T_{222}\lambda - 8T_{112}\lambda)x^2 + 12T_{112}\lambda y^2 \\ &\quad + (4T_{112}\lambda + 4T_{222}\lambda)z^2 + (8T_{111}\lambda - 16T_{122}\lambda)xy \\ &\quad + (8T_{112}T_{122} - 8T_{111}T_{222})xz \\ &\quad + (16T_{122}^2 - 4\lambda^2 - 16T_{112}T_{222})yz, \\ \frac{\partial J}{\partial z} &= (8T_{112}^2 - 8T_{111}T_{122} - 2\lambda^2)x^2 + (8T_{122}\lambda + 8T_{111}\lambda)xz \\ &\quad + (8T_{122}^2 - 8T_{222}T_{112} - 2\lambda^2)y^2 - 6\lambda^2z^2 \\ &\quad + (8T_{112}T_{122} - 8T_{111}T_{222})xy + (8T_{112}\lambda + 8T_{222}\lambda)yz. \end{aligned}$$

By Salmon's formula [17], $\psi_T(\lambda) = \frac{1}{512}\det(G)$, where G is defined by (20), as shown at the top of this page. Thus $\psi_T(\lambda) = p_2\lambda^6 + p_4\lambda^4 + p_6\lambda^2 + p_8$ for some homogeneous polynomials p_m of degree m in T_{ijk}, $m = 2, 4, 6, 8$. See also [7] and [41]. Therefore $D_{\text{eig}}(T)$ is the determinant of some 11×11 matrix in T_{ijk}.

For a generic $T \in \mathsf{S}^3(\mathbb{R}^2)$, $\psi_T(\lambda) = c(\lambda^2 - \gamma_1)(\lambda^2 - \gamma_2)(\lambda^2 - \gamma_3)$ for some $c \in \mathbb{C}$ and distinct $\gamma_i \in \mathbb{C}$, and so $D_{\text{eig}}(T) \neq 0$.

For $T \in H_{\text{disc}}$, $\psi_T(\lambda)$ has multiple roots. For a specific example, let $S \in \mathsf{S}^3(\mathbb{R}^2)$ be defined by $S_{111} = S_{222} = 1$ and set other $S_{ijk} = 0$. Then $D_{\text{eig}}(S) = 0$, implying that S has at least one nonsimple eigenpairs. In fact, $\psi_S(\lambda) = (\lambda + 1)^2(\lambda - 1)^2(2\lambda^2 - 1)$ and so S has two eigenvectors $(1, 0)$, $(0, 1)$ corresponding to the eigenvalue 1, and two eigenvectors $(-1, 0)$, $(0, -1)$ corresponding to the eigenvalue -1. Note that S is, up to a change of coordinates, the same example mentioned at the beginning of this section, i.e., $S = x^{\otimes 3} + y^{\otimes 3}$ has two best rank-one approximations $x^{\otimes 3}$ and $y^{\otimes 3}$.

VII. Uniqueness of Best Rank-One Approximations for Real Tensors

In this section, V and W, with or without subscripts, would generally denote real and complex vector spaces respectively.

Let $W_1, \dots, W_d$ 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 by $\varphi_T(\lambda)$ the resultant of the following homogeneous polynomial equations

$$\begin{cases} \alpha_i \langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle = \lambda (\prod_{j \neq i} \alpha_j) u_i, \\ \langle u_i, u_i \rangle = \alpha_i^2, \end{cases} \tag{21}$$

for $i = 1, \dots, d$. Again by standard theory of resultants [17], [29], $\varphi_T(\lambda)$ vanishes if and only if (21) has a nontrivial solution, and we obtain the following analogue of Definition 28.

Definition 35: $\varphi_T(\lambda)$ is called the *singular characteristic polynomial* of $T \in W_1 \otimes \cdots \otimes W_d$.

Clearly the roots of $\varphi_T(\lambda)$ are the normalized singular values of T. We also have an analogue of Definition 25.

Definition 36: Let $T \in W_1 \otimes \cdots \otimes W_d$. Two normalized singular pairs $(\lambda, u_1, \dots, u_d)$ and $(\mu, v_1, \dots, v_d)$ of T are called *equivalent* if $(\lambda, u_1, \dots, u_d) = (\mu, v_1, \dots, v_d)$, or $(-1)^{d-2}\lambda = \mu$ and $u_i = -v_i$ for $i = 1, \dots, d$. A normalized singular value λ of T is said to be *simple* if it has only one corresponding normalized singular pair up to equivalence.

For real vector spaces $V_1, \dots, V_d$, and $T \in V_1 \otimes \cdots \otimes V_d$, normalized singular pairs are invariant under the product of orthogonal groups $\mathrm{O}(n_1) \times \cdots \times \mathrm{O}(n_d)$.

It follows from [26] that the subset $X \subseteq V_1 \otimes \cdots \otimes V_d$ consisting of tensors without unique best rank-one approximations is contained in a hypersurface. We will show that this can be strengthened to an *algebraic* hypersurface.

Proposition 37: The following subset is an algebraic hypersurface in $V_1 \otimes \cdots \otimes V_d$,

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

Proof: By Lemma 12, X comprises tensors T for which $\|T\|_\sigma$ is not a simple normalized singular value. Let $d = 3$ for notational simplicity. Let $T \in X$. Then there exist some v_1, v_2, v_3 with $\|v_i\| = 1$ and $\{u_{1,1}, u_{2,1}, u_{3,1}\} \neq \{v_1, v_2, v_3\}$ with $\|u_{i,1}\| = 1$ such that

$$\langle T, u_{1,1} \otimes u_{2,1} \otimes u_{3,1} \rangle = \|T\|_\sigma = \langle T, v_1 \otimes v_2 \otimes v_3 \rangle.$$

For each $i = 1, 2, 3$, extend $u_{i,1}$ to an orthonormal basis $\{u_{i,1}, \dots, u_{i,n_i}\}$ of V_i. By an action of $\mathrm{O}(n_1) \times \mathrm{O}(n_2) \times \mathrm{O}(n_3)$

on $V_1 \otimes V_2 \otimes V_3$, we may assume that $v_i = \cos\theta_i u_{i,1} + \sin\theta_i u_{i,2}$. Let $T_{ijk} = \langle T, u_{1,i} \otimes u_{2,j} \otimes u_{3,k}\rangle$. Then we have

$$\begin{cases} T_{111} = \|T\|_\sigma, \\ T_{i11} = T_{1i1} = T_{11i} = 0, \\ T_{111}\cos\theta_2\cos\theta_3 + T_{122}\sin\theta_2\sin\theta_3 = T_{111}\cos\theta_1, \\ T_{212}\cos\theta_2\sin\theta_3 + T_{221}\sin\theta_2\cos\theta_3 + T_{222}\sin\theta_2\sin\theta_3 = T_{111}\sin\theta_1, \\ T_{j12}\cos\theta_2\sin\theta_3 + T_{j21}\sin\theta_2\cos\theta_3 + T_{j22}\sin\theta_2\sin\theta_3 = 0, \\ T_{111}\cos\theta_1\cos\theta_3 + T_{212}\sin\theta_1\sin\theta_3 = T_{111}\cos\theta_2, \\ T_{122}\cos\theta_1\sin\theta_3 + T_{221}\sin\theta_1\cos\theta_3 + T_{222}\sin\theta_1\sin\theta_3 = T_{111}\sin\theta_2, \\ T_{1j2}\cos\theta_1\sin\theta_3 + T_{2j1}\sin\theta_1\cos\theta_3 + T_{2j2}\sin\theta_1\sin\theta_3 = 0, \\ T_{111}\cos\theta_1\cos\theta_2 + T_{221}\sin\theta_1\sin\theta_2 = T_{111}\cos\theta_3, \\ T_{122}\cos\theta_1\sin\theta_2 + T_{212}\sin\theta_1\cos\theta_2 + T_{222}\sin\theta_1\sin\theta_2 = T_{111}\sin\theta_3, \\ T_{12j}\cos\theta_1\sin\theta_2 + T_{21j}\sin\theta_1\cos\theta_2 + T_{22j}\sin\theta_1\sin\theta_2 = 0, \end{cases} \tag{22}$$

for $i \neq 1$ and $j > 2$. By eliminating the parameter θ, we obtain a system of polynomial equations that the T_{ijk}'s satisfy.

Let J be the *incidence variety* in $V_1 \otimes V_2 \otimes V_3 \times \mathrm{O}(n_1) \times \mathrm{O}(n_2) \times \mathrm{O}(n_3)$, i.e., for each $(T, g_1, g_2, g_3) \in J$ where $g_i = [u_{i,1}, \dots, u_{i,n_i}] \in \mathrm{O}(n_i)$, there is some $(\theta_1, \theta_2, \theta_3)$ such that the T_{ijk}'s satisfy (22). Define the projections

$$\begin{tikzcd} & J \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ V_1 \otimes V_2 \otimes V_3 & & \mathrm{O}(n_1) \times \mathrm{O}(n_2) \times \mathrm{O}(n_3) \end{tikzcd} \tag{23}$$

by $\pi_1(T, g_1, g_2, g_3) = T$ and $\pi_2(T, g_1, g_2, g_3) = (g_1, g_2, g_3)$. Since $\|T\|_\sigma$ is a root of $\varphi_T(\lambda)$, $\|T\|_\sigma$ and its normalized singular vector tuples depend algebraically on T, implying that J is an algebraic variety. $\|T\|_\sigma$ is simple if and only if T is in the image of π_1, i.e., $X = \pi_1(J)$.

Define $T' \in V_1 \otimes V_2 \otimes V_3$ by $T'_{111} = T'_{222} = 1$ and set all other terms $T'_{ijk} = 0$. Then T' has two normalized singular vector tuples corresponding to its normalized singular value$\|T\|_\sigma$. So $T' \in \pi_1(J)$. Since T' has a finite number of singular pairs, a generic $T \in \pi_1(J)$ must also have a finite number of singular pairs by semicontinuity. So $\dim \pi_1^{-1}(T) = \dim \mathrm{O}(n_1 - 2) + \dim \mathrm{O}(n_2 - 2) + \dim \mathrm{O}(n_3 - 2)$ for a generic $T \in \pi_1(J)$, and $\dim X = \dim \pi_1(J) = \dim J - \dim \mathrm{O}(n_1 - 2) - \dim \mathrm{O}(n_2 - 2) - \dim \mathrm{O}(n_3 - 2)$.

Since π_2 is a dominant morphism, and the dimension of a generic fiber $\pi_2^{-1}(g_1, g_2, g_3)$ is $\dim V_1 \otimes V_2 \otimes V_3 - 2(n_1 + n_2 + n_3) + 8$, it follows that $\dim J = \dim V_1 \otimes V_2 \otimes V_3 - 2(n_1 + n_2 + n_3) + 8 + \dim \mathrm{O}(n_1) + \dim \mathrm{O}(n_2) + \dim \mathrm{O}(n_3)$. Therefore the codimension of X is 1. $\square$

We will show that normalized singular vector tuples of a generic tensor are distinct, a result that we will need later but is also of independent interest.

Proposition 38: Let $W_1, \dots, W_d$ be vector spaces over $\mathbb{C}$. A generic $T \in W_1 \otimes \cdots \otimes W_d$ has distinct equivalence classes of normalized singular pairs.

Our proof of Proposition 38 will rely on the next three intermediate results. The first required result is a 'Bertini-type' theorem introduced in [25].

Theorem 39 (Friedland–Ottaviani): Let E be a vector bundle on a smooth variety B. Let $S \subseteq H^0(B, E)$ generate E. If $\operatorname{rank}(E) > \dim B$, then the zero locus of a generic $\zeta \in S$ is empty.

Lemma 40: Let $T \in W_1 \otimes \cdots \otimes W_d$ be generic and let $(u_1, \dots, u_d)$ be a normalized singular vector tuple of T. If v_d is not a scalar multiple of u_d, then $(u_1, \dots, u_{d-1}, v_d)$ is not a normalized singular vector tuple of T.

Proof: Suppose $\lambda u_d = \langle T, u_1 \otimes \cdots \otimes u_{d-1}\rangle = \mu v_d$ for some v_d not a scalar multiple of u_d. Then λ or μ must be 0, contradicting the fact that 0 cannot be a singular value of a generic T [25, Th. 1]. $\square$

Lemma 41: Let $u_i, v_i, w_i \in W_i$ with $\langle u_i, u_i\rangle = \langle v_i, v_i\rangle = 1$, $i = 1, 2, 3$. For $x \in W_i$, we write $[x]_i$ for the corresponding element in the quotient space $W_i / \operatorname{span}(u_i)$. Suppose $u_i = v_i$ for at most one i. Then

(i) *the system of linear equations*

$$\begin{cases} \langle T, u_2 \otimes u_3\rangle = \langle T, v_1 \otimes v_2 \otimes v_3\rangle u_1 + w_1, \\ \langle T, u_1 \otimes u_3\rangle = w_2, \\ \langle T, u_1 \otimes u_2\rangle = w_3, \end{cases} \tag{24}$$

has a solution $T \in W_1 \otimes W_2 \otimes W_3$ if and only if $\langle u_2, w_2\rangle = \langle u_3, w_3\rangle$;

(ii) *the system of linear equations*

$$\begin{cases} \langle T, u_2 \otimes u_3\rangle = \langle T, v_1 \otimes v_2 \otimes v_3\rangle u_1 + w_1, \\ [\langle T, u_1 \otimes u_3\rangle]_2 = [w_2]_2, \\ [\langle T, u_1 \otimes u_2\rangle]_3 = [w_3]_3, \end{cases} \tag{25}$$

always has a solution $T \in W_1 \otimes W_2 \otimes W_3$.

Proof: Note that the variables in these linear equations are T_{ijk}'s, the coordinates of T.

(i) Let A be the coefficient matrix in (24) and b be the right-hand side. The system has a solution if and only if A and the augmented matrix $[A \mid b]$ have the same rank, i.e., if there is some $x_i \in W_i$, $i = 1, 2, 3$, such that $x_1 \otimes u_2 \otimes u_3 + u_1 \otimes x_2 \otimes u_3 + u_1 \otimes u_2 \otimes x_3 - \langle x_1, u_1\rangle \cdot v_1 \otimes v_2 \otimes v_3 = 0$, then $\langle x_1, w_1\rangle + \langle x_2, w_2\rangle + \langle x_3, w_3\rangle = 0$. Since $x_1 \otimes u_2 \otimes u_3 + u_1 \otimes x_2 \otimes u_3 + u_1 \otimes u_2 \otimes x_3 - \langle x_1, u_1\rangle \cdot v_1 \otimes v_2 \otimes v_3 = 0$ if and only if $x_1 = 0$, $x_2 = \alpha u_2$, $x_3 = -\alpha u_3$ or $x_1 = 0$, $x_2 = -\alpha u_2$, $x_3 = \alpha u_3$ for some α, the system (24) has a solution if and only if $\langle u_2, w_2\rangle = \langle u_3, w_3\rangle$.

(ii) The system (25) has a solution if and only if $\langle u_2, w_2 + t_2 u_2\rangle = \langle u_3, w_3 + t_3 u_3\rangle$ for some $t_2, t_3 \in \mathbb{C}$. Choose any t_2, t_3 such that $t_3 - t_2 = \langle u_2, w_2\rangle - \langle u_3, w_3\rangle$. $\square$

Proof of Proposition 38: Let $d = 3$ for notational convenience. For $i = 1, 2, 3$, let $C_i = \{u_i \in W_i : \langle u_i, u_i\rangle = 1\}$, F_i be the trivial vector bundle on C_i with fiber isomorphic to W_i, L_i be the tautological line bundle on C_i, and Q_i be the quotient bundle F_i / L_i on C_i. Consider the exact sequence of vector bundles

$$0 \to L_i \to F_i \to Q_i \to 0$$

over C_i. Let $M = C_1 \times C_2 \times C_3$. We will need to discuss vector bundles over $M \times M$ and for clarity, we distinguish the two copies of M. So we write $M_1 \times M_2$ where $M_1 = M_2 = M$. Let $\pi_{i,j} : M_i \to C_j$ be the natural projection for $i = 1, 2$ and $j = 1, 2, 3$. Let $p_i : M_1 \times M_2 \to M_i$ be the natural projection for $i = 1, 2$. Then we have the following diagram:

$$\begin{tikzcd} & & & M_1 \times M_2 \arrow[dll, "p_1"'] \arrow[drr, "p_2"] & & & \\ & M_1 \arrow[dl] \arrow[d] \arrow[dr] & & & & M_2 \arrow[dl] \arrow[d] \arrow[dr] & \\ C_1 & C_2 & C_3 & & C_1 & C_2 & C_3 \end{tikzcd}.$$

Consider the vector bundle on $M_1 \times M_2$,

$$\widetilde{E} = \left(\bigoplus_{j=1}^{3} p_1^* \pi_{1,j}^*(Q_j)\right) \oplus p_2^* \pi_{2,1}^*(F_1) \oplus \left(\bigoplus_{j=2}^{3} p_2^* \pi_{2,j}^*(Q_j)\right),$$

where f^* denotes the pullback induced by f. Let

$$X_i = \{(v_1, v_2, v_3, u_1, u_2, u_3) \in M_1 \times M_2 : u_j = v_j \text{ for all } j \neq i\}.$$

By Lemma 40, to study the behavior of normalized singular pairs of a generic tensor, we need only consider the following open subset of the affine variety $M_1 \times M_2$,

$$B = M_1 \times M_2 \setminus (X_1 \cup X_2 \cup X_3),$$

and its corresponding vector bundle $E = \widetilde{E}|_B$ over the base space B. Then

$$\operatorname{rank}(E) = 2\sum_{i=1}^{3} \dim W_i - 5 > \dim B = 2\sum_{i=1}^{3} \dim W_i - 6.$$

So the inequality in Theorem 39 holds for our choice of E and B. Now let

$$\begin{aligned} S = \{s \in H^0(B, E) : s(&v_1, v_2, v_3, u_1, u_2, u_3) \\ = ([\langle T, v_2 \otimes v_3\rangle]_1, &[\langle T, v_1 \otimes v_3\rangle]_2, [\langle T, v_1 \otimes v_2\rangle]_3, \\ \langle T, u_2 \otimes u_3\rangle &- \langle T, v_1 \otimes v_2 \otimes v_3\rangle u_1, \\ [\langle T, u_1 \otimes u_3\rangle]_2, &[\langle T, u_1 \otimes u_2\rangle]_3)\}. \end{aligned}$$

By Lemma 41 and [25, Lemma 8], S generates E. By Theorem 39, a generic section of E does not vanish on B, implying tha each normalized singular value of a generic tensor is distinct and simple. $\square$

Let $D_{\text{sing}}(T)$ be the *singular discriminant*, the resultant of the singular characteristic polynomial φ_T and its derivative φ_T'. Since a generic T has distinct equivalence classes of normalized singular pairs, φ_T has simple roots, and so $D_{\text{sing}}(T)$ does not vanish identically. As $D_{\text{sing}}(T)$ vanishes on X, the hypersurface defined in Theorem 37, $D_{\text{sing}}(T) = 0$ indeed defines a hypersurface in $W_1 \otimes \cdots \otimes W_d$. Note that X is a union of some components of the real points of X_{disc}. Finally, we arrive at our main result of this section, singular value analogues of Theorem 31 and Corollaries 32 and 33.

Theorem 42: $D_{\text{sing}}(T) = 0$ *is a defining equation of the hypersurface*

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

In the following, let V_i be a real vector space and $W_i = V_i \otimes_{\mathbb{R}} \mathbb{C}$ be its complexification, $i = 1, \ldots, d$.

Corollary 43: Let $T \in V_1 \otimes \cdots \otimes V_d$ be real. If $D_{\text{sing}}(T) \neq 0$, then T has a unique best rank-one approximation.

We deduce the following analogue for nonnegative tensors from Theorem 16 and Corollary 43.

Corollary 44: Let $T \in V_1 \otimes \cdots \otimes V_d$ be nonnegative. If $D_{\text{sing}}(T) \neq 0$, then T has a unique best nonnegative rank-one approximation.

Theorem 42 shows that the ED discriminant X_{disc} of the cone over the Segre variety $\mathbb{P}W_1 \times \cdots \times \mathbb{P}W_d$ is a hypersurface when $d \geq 3$, and $D_{\text{sing}}(T) = 0$ gives its defining equation. The discussion before Theorem 42 shows that the set of real points of X_{disc} is a real hypersurface. It is interesting to note that when $d = 2$, i.e., the matrix case, the set of real points of the ED discriminant of the Segre variety $\mathbb{P}W_1 \times \mathbb{P}W_2$ has codimension 2 [22, Example 7.6].

APPENDIX

We use semirings and semimodules instead of rings and modules to construct tensor products of cones in order to give nonnegative tensors an algebraic description and state our results in a more general setting. A semimodule over a semiring is essentially the same notion as a vector space over a field, except that the field of scalars is now replaced by a semiring of scalars like the nonnegative reals. The nonnegative reals do not form a field or even a ring since they do not have additive inverses, but aside from this, $\mathbb{R}_+$ has all the properties of scalars that makes the notion of a vector space so useful in engineering.

Definition 45: A semiring R is a set equipped with binary operations $+$ and $\cdot$ such that

- $(R, +)$ is a commutative monoid with identity element 0;
- $(R, \cdot)$ is a monoid with identity element 1;
- Multiplication left and right distributes over addition:

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c),$$
$$(a + b) \cdot c = (a \cdot c) + (b \cdot c);$$

- Multiplication by 0 annihilates R:

$$0 \cdot a = a \cdot 0 = 0.$$

Definition 46: A commutative semiring is a semiring whose multiplication is commutative.

Definition 47: A semimodule M over a commutative semiring R is a commutative monoid $(M, +)$ and an operation $\cdot : R \times M \to M$ such that for all r, s in R and $x, y \in M$, we have:

$$\begin{aligned} r \cdot (x + y) &= r \cdot x + r \cdot y, \\ (r + s) \cdot x &= r \cdot x + s \cdot x, \\ (rs) \cdot x &= r \cdot (s \cdot x), \\ 1_R \cdot x &= x. \end{aligned}$$

In our context, the set of nonnegative real numbers $\mathbb{R}_+$ is a commutative semiring and the set of nonnegative tensors is a semimodule over $\mathbb{R}_+$.

ACKNOWLEDGMENT

Y. Qi thanks Giorgio Ottaviani for careful reading and very helpful advice, especially for suggesting that we use an argument in [25] for Proposition 38. He also thanks Emil Horobeţ for helpful discussion. L.-H. Lim thanks Shmuel Friedland for a pointer to Rademacher Theorem. The authors thank Yuning Yang for pointing out an error in an earlier version. Special thanks to the three anonymous reviewers for their useful suggestions.

REFERENCES

[1] H. Abo, G. Ottaviani, and C. Peterson, "Induction for secant varieties of Segre varieties," *Trans. Amer. Math. Soc.*, vol. 361, no. 2, pp. 767–792, 2009.

[2] S. Arora, R. Ge, R. Kannan, and A. Moitra, "Computing a nonnegative matrix factorization—Provably," in *Proc. 44th Annu. ACM Symp. Theory Comput. (STOC)*, 2012, pp. 145–162.

[3] S. Banach, "Über homogene polynome in (L^2)," *Stud. Math.*, vol. 7, no. 1, pp. 36–44, 1938.

[4] M. Banagl, "The tensor product of function semimodules," *Algebra Univ.*, vol. 70, no. 3, pp. 213–226, 2013.

[5] C. Bocci, L. Chiantini, and G. Ottaviani, "Refined methods for the identifiability of tensors," *Ann. Math. Pura Appl.*, vol. 193, no. 6, pp. 1691–1702, 2014.

[6] R. Bro, "Multi-way analysis in the food industry: Models, algorithms, and applications," Ph.D. dissertation, Dept. Dairy Food Sci., Univ. Amsterdam, Amsterdam, The Netherlands, 1998.

[7] D. Cartwright and B. Sturmfels, "The number of eigenvalues of a tensor," *Linear Algebra Appl.*, vol. 438, no. 2, pp. 942–952, 2013.

[8] K. C. Chang, K. Pearson, and T. Zhang, "Perron–Frobenius theorem for nonnegative tensors," *Commun. Math. Sci.*, vol. 6, no. 2, pp. 507–520, 2008.

[9] L. Chiantini and G. Ottaviani, "On generic identifiability of 3-tensors of small rank," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 3, pp. 1018–1037, 2012.

[10] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, "An algorithm for generic and low-rank specific identifiability of complex tensors," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 4, pp. 1265–1287, 2014.

[11] M.-T. Chu and M.-M. Lin, "Low-dimensional polytope approximation and its applications to nonnegative matrix factorization," *SIAM J. Sci. Comput.*, vol. 30, no. 3, pp. 1131–1155, 2008.

[12] A. Cichocki, R. Zdunek, A. H. Phan, and S. I. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. New York, NY, USA: Wiley, 2009.

[13] J. E. Cohen, R. C. Farias, and P. Comon, "Fast decomposition of large nonnegative tensors," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 862–866, Jul. 2015.

[14] P. Comon, "Tensors: A brief introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 44–53, May 2014.

[15] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, "Symmetric tensors and symmetric tensor rank," *SIAM J. Matrix Anal. Anal.*, vol. 30, no. 3, pp. 1254–1279, 2008.

[16] P. Comon, J. M. F. ten Berge, L. De Lathauwer, and J. Castaing, "Generic and typical ranks of multi-way arrays," *Linear Algebra Appl.*, vol. 430, nos. 11–12, pp. 2997–3007, 2009.

[17] D. A. Cox, J. Little, and D. O'Shea, *Using Algebraic Geometry* (Graduate Texts in Mathematics), vol. 185, 2nd ed. New York, NY, USA: Springer, 2005.

[18] A. P. da Silva, P. Comon, and A. L. F. de Almeida. (2015). "Rank-1 tensor approximation methods and application to deflation." [Online]. Available: http://arxiv.org/abs/1508.05273.

[19] V. de Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1084–1127, 2008.

[20] I. Domanov and L. De Lathauwer, "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.*, vol. 34, no. 3, pp. 855–875, 2013.

[21] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part II: Uniqueness of the overall decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 876–903, 2013.

[22] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. R. Thomas, "The Euclidean distance degree of an algebraic variety," *Found. Comput. Math.*, vol. 16, no. 1, pp. 99–149, 2016.

[23] S. Friedland, "Best rank one approximation of real symmetric tensors can be chosen symmetric," *Frontiers Math. China*, vol. 8, no. 1, pp. 19–40, 2013.

[24] S. Friedland, S. Gaubert, and L. Han, "Perron–Frobenius theorem for nonnegative multilinear forms and extensions," *Linear Algebra Appl.*, vol. 438, no. 2, pp. 738–749, 2013.

[25] S. Friedland and G. Ottaviani, "The number of singular vector tuples and uniqueness of best rank-one approximation of tensors," *Found. Comput. Math.*, vol. 14, no. 6, pp. 1209–1242, 2014.

[26] S. Friedland and M. Stawiska, "Some approximation problems in semi-algebraic geometry," *Banach Center Pub.*, vol. 107, pp. 129–143, Dec. 2016.

[27] M. P. Friedlander and K. Hatz, "Computing nonnegative tensor factorizations," *Optim. Method Softw.*, vol. 23, no. 4, pp. 631–647, 2008.

[28] L. D. Garcia, M. Stillman, and B. Sturmfels, "Algebraic geometry of Bayesian networks," *J. Symbolic Comput.*, vol. 39, nos. 3–4, pp. 331–355, 2005.

[29] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, Resultants, and Multidimensional Determinants*. Boston, MA, USA: Birkhäuser, 1994.

[30] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*. Berlin, Germany: Springer, 2012.

[31] C. J. Hillar and L.-H. Lim, "Most tensor problems are NP-hard," *J. ACM*, vol. 60, no. 6, 2013, Art. no. 45.

[32] N.-D. Ho, "Nonnegative matrix factorization algorithms and applications," Ph.D. dissertation, Dept. Math., Ecole Polytech. Louvain, Leuven, Belgium, 2008.

[33] M. I. Jordan, "Graphical models," *Statist. Sci.*, vol. 19, no. 1, pp. 140–155, 2004.

[34] H. Kim and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 2, pp. 713–730, 2008.

[35] J. Kim, Y. He, and H. Park, "Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework," *J. Global Optim.*, vol. 58, no. 2, pp. 285–319, 2014.

[36] J. Kim and H. Park, "Fast nonnegative matrix factorization: An active-set-like method and comparisons," *SIAM J. Sci. Comput.*, vol. 33, no. 6, pp. 3261–3281, 2011.

[37] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[38] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, no. 2, pp. 95–138, 1977.

[39] J. M. Landsberg, *Tensors: Geometry and Applications* (Graduate Studies in Mathematics), vol. 128, Providence, RI, USA: AMS, 2012.

[40] D.-D. Lee and H.-S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Sys. (NIPS)*, vol. 13. 2001, pp. 556–562.

[41] A.-M. Li, L. Qi, and B. Zhang, "E-characteristic polynomials of tensors," *Commun. Math. Sci.*, vol. 11, no. 1, pp. 33–53, 2013.

[42] T. Lickteig, "Typical tensorial rank," *Linear Algebra Appl.*, vol. 69, pp. 95–120, Aug. 1985.

[43] L.-H. Lim, "Tensors and hypermatrices," *Handbook of Linear Algebra*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2013.

[44] L.-H. Lim, "Singular values and eigenvalues of tensors: A variational approach," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt.*, Puerto Vallarta, Mexico, Dec. 2005, pp. 129–132.

[45] L.-H. Lim and P. Comon, "Nonnegative approximations of nonnegative tensors," *J. Chemometrics*, vol. 23, nos. 7–8, pp. 432–441, 2009.

[46] C.-B. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Comput.*, vol. 19, no. 10, pp. 2756–2779, 2007.

[47] L. Oeding and G. Ottaviani, "Eigenvectors of tensors and algorithms for Waring decomposition," *J. Symbolic Comput.*, vol. 54, pp. 9–35, Jul. 2013.

[48] A.-H. Phan, P. Tichavský, and A. Cichocki, "Tensor deflation for CANDECOMP/PARAFAC—Part I: Alternating subspace update algorithm," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5924–5938, Nov. 2015.

[49] A.-H. Phan, P. Tichavský, and A. Cichocki, "Tensor deflation for CANDECOMP/PARAFAC—Part II: Initialization and error analysis," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5939–5950, Nov. 2015.

[50] L. Qi, "Eigenvalues of a real supersymmetric tensor," *J. Symbolic Comput.*, vol. 40, no. 6, pp. 1302–1324, 2005.

[51] L. Qi, "Eigenvalues and invariants of tensors," *J. Math. Anal. Appl.*, vol. 325, no. 2, pp. 1363–1377, 2007.

[52] A. Shashua and T. Hazan, "Non-negative tensor factorization with applications to statistics and computer vision," in *Proc. 22nd Int. Conf. Mach. Learn. (ICML)*, 2005, pp. 792–799.

[53] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chemometrics*, vol. 14, no. 3, pp. 229–239, 2000.

[54] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis*. Chichester U.K.: Wiley, 2004.

[55] A. Stegeman and P. Comon, "Subtracting a best rank-1 approximation may increase tensor rank," *Linear Algebra Appl.*, vol. 433, no. 7, pp. 1276–1300, 2010.

[56] V. Strassen, "Rank and optimal computation of generic tensors," *Linear Algebra Appl.*, vols. 52–53, pp. 645–685, Jul. 1983.

[57] P. Tichavský, A. H. Phan, and Z. Koldovsky, "Cramér–Rao-induced bounds for CANDECOMP/PARAFAC tensor decomposition," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 1986–1997, Apr. 2013.

[58] N. Vannieuwenhoven, J. Nicaise, R. Vandebril, and K. Meerbergen, "On generic nonexistence of the Schmidt–Eckart–Young decomposition for complex tensors," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 3, pp. 886–903, 2014.

[59] S. A. Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM J. Optim.*, vol. 20, no. 3, pp. 1364–1377, 2009.

[60] M. Velasco, "Linearization functors on real convex sets," *SIAM J. Optim.*, vol. 25, no. 1, pp. 1–27, 2015.

[61] Y. Yang and Q. Yang, "Further results for Perron–Frobenius theorem for nonnegative tensors," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 5, pp. 2517–2530, 2010.

[62] P. Zhang, H. Wang, R. J. Plemmons, and V. P. Pauca, "Tensor methods for hyperspectral data analysis: A space object material identification study," *J. Opt. Soc. Amer.*, vol. 25, no. 12, pp. 3001–3012, 2008.

[63] G. Zhou, A. Cichocki, and S. Xie, "Fast nonnegative matrix/tensor factorization based on low-rank approximation," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2928–2940, Jun. 2012.

[64] J. Zhou, A. Bhattacharya, A. H. Herring, and D. B. Dunson, "Bayesian factorizations of big sparse tensors," *J. Amer. Statist. Assoc.*, vol. 110, no. 512, pp. 1562–1576, 2015.

Yang Qi received his Ph.D. degree in 2013 from Texas A&M University. He has been a postdoc in Gipsa-Lab since 2013.

Pierre Comon (M'87–SM'95–F'07) graduated in 1982, and received the Doctorate degree in 1985, both from the University of Grenoble, France. He later received the Habilitation to Lead Researches in 1995, from the University of Nice, France. He has been for nearly 13 years in industry, first with Crouzet-Sextant, Valence, France, between 1982 and 1985, and then with Thomson Marconi, Sophia-Antipolis, France, between 1988 and 1997. He spent 1987 with the ISL laboratory, Stanford University, CA. He joined in 1997 the Eurecom Institute, Sophia-Antipolis, France. He is research director with CNRS since 1998, first at laboratory I3S, Sophia- Antipolis, France, until 2012, and then at Gipsa-Lab, Grenoble, France. His research interests include High-Order Statistics, Blind techniques, Statistical Signal and Array Processing, Tensor decompositions, Multi-Way Factor Analysis and its applications to Data Science.

Dr. Comon was Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1995 to 1998, and a member of the French National Committee of Scientific Research from 1995 to 2000. He was the coordinator of the European Basic Research Working Group on HOS, ATHOS, from 1992 to 1995. Between 1992 and 1998, he was a member of the Technical and Scientific Council of the Thomson Group (now Thales). Between 2001 and 2004 he acted as launching Associate Editor with the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I, in the area of Blind Techniques. Ha has also been a member of the editorial board of the ELSEVIER JOURNAL SIGNAL PROCESSING, and member of several IEEE TC. He is presently in the editorial board of the SIAM JOURNAL ON MATRIX ANALYSIS AND APPLICATIONS.

Lek-Heng Lim received his undergraduate education at the National University of Singapore and his postgraduate education at Cornell University (MS), Cambridge University (Clare Hall Fellow), and Stanford University (PhD). He was a Charles Morrey assistant professor at the University of California, Berkeley from 2007-2010 and has been an assistant professor at the University of Chicago since 2010. He is a member of the editorial boards of LINEAR ALGEBRA AND ITS APPLICATIONS and LINEAR AND MULTILINEAR ALGEBRA. His work is supported by an AFOSR Young Investigator Award, a DARPA Young Faculty Award, and an NSF CAREER Award.