



HAL
open science

Uniqueness of Nonnegative Tensor Approximations

Yang Qi, Pierre Comon, Lek-Heng Lim

► **To cite this version:**

Yang Qi, Pierre Comon, Lek-Heng Lim. Uniqueness of Nonnegative Tensor Approximations. 2014.
hal-01015519v3

HAL Id: hal-01015519

<https://hal.science/hal-01015519v3>

Preprint submitted on 29 Oct 2014 (v3), last revised 12 Apr 2016 (v5)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Uniqueness of Nonnegative Tensor Approximations

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

Abstract

We show that the best nonnegative rank- r approximation of a nonnegative tensor is almost always unique and that nonnegative tensors with nonunique best nonnegative rank- r approximation form a semialgebraic set contained in an algebraic hypersurface. We then establish a singular vector variant of the Perron–Frobenius Theorem for positive tensors that may be used for answering various questions involving nonnegative tensor approximations. In particular, we apply it to show that the best nonnegative rank-1 approximation of a positive tensor is always unique and that a best nonnegative rank- r approximation of a positive tensor can almost never be obtained by deflation.

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 [20], [21], [23], [26] 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 [20], [27] — a ubiquitous model that underlies many applications. This is in fact one of the simplest Bayesian network [12], [14], [15], a local expression of the joint distribution of a set of random variables x_i as

$$p(x_1, \dots, x_d) = \int \prod_{i=1}^d p(x_i | \theta) d\mu_\theta \quad (\text{I.1})$$

where θ is some unknown latent random variable. The relation expressed in (I.1) is often called the *naive Bayes hypothesis*. In the case when both 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

$$t_{i_1, \dots, i_d} = \sum_{p=1}^r \lambda_p u_{i_1, p} \cdots u_{i_d, p}. \quad (\text{I.2})$$

One can show [20] that *any* decomposition of a nonnegative tensor of the form in (I.2) may, upon normalization by a suitable constant, be regarded as (I.1), a marginal decomposition of 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 (I.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. We first 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 [20] that every nonnegative tensor has a best nonnegative rank- r approximation. Here 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 some hypersurface. We then show that for $r = 1$, every best rank-1 approximation is necessarily a best nonnegative rank-1 approximation. Moreover, every positive tensor T has a unique best nonnegative rank-1 approximation and that if T has any partial symmetry under permutation of its entries, the best nonnegative rank-1 approximation can always be chosen to preserve this symmetry. These last two results depend on a ‘Perron–Frobenius Theorem’ for singular pairs of positive tensors that we will establish in this article. Lastly we show that one cannot in general obtain a best nonnegative rank- r approximation by ‘deflation’, *i.e.*, by finding r successive best nonnegative rank-1 approximations.

* Yang Qi and Pierre Comon are with GIPSA-Lab CNRS UMR5216, Grenoble Campus, BP.46, F-38402 St Martin d’Heres cedex, France. ‡ Lek-Heng Lim is with the Computational and Applied Mathematics Initiative, Department of Statistics, University of Chicago, 5734 South University Avenue, Chicago, IL 60637, USA.

II. NONNEGATIVE TENSORS

A tensor of order d (d -tensor for short) may be represented as a d -dimensional hypermatrix, *i.e.*, a d -way 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 -way 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 $V_1 \times \dots \times V_d$ be the set of d -tuples of vectors. Then the *tensor product* $V = V_1 \otimes \dots \otimes V_d$ is the free linear space spanned by $V_1 \times \dots \times V_d$ quotient by an equivalence relation that imposes the two properties below for every $\alpha_i, \beta_i \in \mathbb{K}$:

$$(\alpha_1 u_1, \alpha_2 u_2, \dots, \alpha_d u_d) \equiv \left(\prod_{i=1}^d \alpha_i \right) (u_1, u_2, \dots, u_d), \quad (\text{II.1})$$

$$(u_1, \dots, \alpha_i u_i, \dots, u_d) + (u_1, \dots, \beta_i v_i, \dots, u_d) \equiv (u_1, \dots, \alpha_i u_i + \beta_i v_i, \dots, u_d). \quad (\text{II.2})$$

A *tensor* is any element of $V_1 \otimes \dots \otimes V_d$.

More details on the definition of tensor spaces may be found in [6], [13], [17], [18].

Definition 2. A *decomposable tensor* is one of the form $u_1 \otimes \dots \otimes u_d$, $u_i \in V_i$, $i = 1, \dots, d$. It represents the equivalence class of tuples up to scaling as in (II.1), *i.e.*, $u_1 \otimes \dots \otimes u_d = \{(\alpha_1 u_1, \dots, \alpha_d u_d) : \prod_{i=1}^d \alpha_i = 1\}$.

By the above definition, 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 non-zero 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:

Definition 3 (Rank decomposition). For every $T \in V_1 \otimes \dots \otimes V_d$, there exist $u_{i,p} \in V_i$, $i = 1, \dots, d$, $p = 1, \dots, \text{rank}(T)$, such that

$$T = \sum_{p=1}^{\text{rank}(T)} u_{1,p} \otimes \dots \otimes u_{d,p}. \quad (\text{II.3})$$

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.

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 $V := V_1 \otimes \dots \otimes V_d$ of the form $u_1 \otimes \dots \otimes u_d$ 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

$$V^+ = \left\{ \sum_{p=1}^r u_{1,p} \otimes \dots \otimes u_{d,p} \in V : u_{i,p} \in V_i^+, i = 1, \dots, d, p = 1, \dots, r, r \in \mathbb{N} \right\}.$$

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*, denoted by $\text{rank}_+(T)$, and nonnegative tensor rank decomposition. For every $T \in V^+$, there exist $u_{i,p} \in V_i^+$, $i = 1, \dots, d$, $p = 1, \dots, r$, such that

$$T = \sum_{p=1}^{\text{rank}_+(T)} u_{1,p} \otimes \dots \otimes u_{d,p}. \quad (\text{II.4})$$

We will call the nonnegative tensor rank the *nonnegative rank* and the decomposition above a *nonnegative decomposition* of the nonnegative tensor T , for short. An obvious property is that $\text{rank}_+(T) \geq \text{rank}(T)$ for any $T \in V^+$.

We now examine an alternative coordinate-free approach for 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 [2] for details on the existence and construction of such a tensor product.

Definition 4. Let R be a commutative semiring and M, N be R -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 $\varphi : M \times N \rightarrow M \otimes_R N$ such that given any

other R -semimodule S together with an R -bilinear map $h : M \times N \rightarrow S$, there is a unique R -linear map $\tilde{h} : M \otimes_R N \rightarrow S$ satisfying $h = \tilde{h} \circ \varphi$.

Definition 5. 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 $F(C_1, \dots, C_d) / \sim$, where $F(C_1, \dots, C_d)$ is the free monoid generated by all n -tuples $(v_1, \dots, v_d) \in C_1 \times \cdots \times C_d$, and \sim is the equivalence relation on $F(C_1, \dots, C_d)$ generated by

$$(v_1, \dots, \alpha v_i + \beta v'_i, \dots, v_d) \sim \alpha(v_1, \dots, v_i, \dots, v_d) + \beta(v_1, \dots, v'_i, \dots, v_d)$$

for all $v_i, v'_i \in C_i$, $\alpha, \beta \in \mathbb{R}_+$, and $i = 1, \dots, d$. 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, \dots, v_d) in $F(C_1, \dots, C_d) / \sim$.

A *multiconic map* from $C_1 \times \cdots \times C_d$ to a convex cone C is a map $\varphi : C_1 \times C_2 \times \cdots \times C_d \rightarrow C$ with the property

$$\varphi(u_1, \dots, \alpha v_i + \beta w_i, \dots, u_d) = \alpha \varphi(u_1, \dots, v_i, \dots, u_d) + \beta \varphi(u_1, \dots, w_i, \dots, u_d), \quad \alpha, \beta \in \mathbb{R}_+,$$

for all $i = 1, \dots, d$.

The tensor product space $C_1 \otimes \cdots \otimes C_d$ with $\nu : C_1 \times \cdots \times C_d \rightarrow C_1 \otimes \cdots \otimes C_d$ defined by

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

and extended nonnegative linearly 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:

$$\begin{array}{ccc} C_1 \times \cdots \times C_d & \xrightarrow{\nu} & C_1 \otimes \cdots \otimes C_d \\ & \searrow \varphi & \downarrow \psi \\ & & C \end{array}$$

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}_+ .

If $C_1 = \mathbb{R}_+^{n_1}, \dots, C_d = \mathbb{R}_+^{n_d}$, 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

$$[a_1, \dots, a_l]^\top \otimes [b_1, \dots, b_m]^\top \otimes [c_1, \dots, c_n]^\top = [a_i b_j c_k]_{i,j,k=1}^{l,m,n}$$

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

Definition 6. 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 $\text{rank}_+(T)$, 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 $u_{1,1} \otimes \cdots \otimes u_{d,1}, \dots, u_{1,r} \otimes \cdots \otimes u_{d,r}$. 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

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

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 (II.3) 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 [16], [22], [3], [5], notably the two following results.

Theorem 7 (Kruskal). *The rank decomposition of a d -tensor T is unique if*

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

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

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

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

Theorem 9 (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 $\prod_{i=1}^d n_i \leq 15000$ except for some exceptional cases.

The authors of [5] also strengthened the above result by a prior compression of tensor T . The consequence is that the dimensions n_i in Theorem 9 may be replaced by the multilinear rank of T , which allows significant tightening of the upper bound for low multilinear rank tensors.

Nevertheless these results do not apply 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

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

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

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

IV. EXISTENCE AND GENERIC UNIQUENESS OF RANK- r APPROXIMATIONS

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 will assume that each V_i has a given inner product, which induces an inner product $\langle \cdot, \cdot \rangle$ on $V_1 \otimes \dots \otimes V_d$ that restricts to $C_1 \otimes \dots \otimes C_d$. We let

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

where $\|\cdot\|$ is the *Hilbert–Schmidt norm*, *i.e.*, the l^2 -norm given by the inner product. Henceforth any unlabelled norm $\|\cdot\|$ on $V_1 \otimes \dots \otimes V_d$ will always denote the Hilbert–Schmidt norm with respect to some fixed choices of orthonormal bases on V_1, \dots, V_d (which induces an orthonormal basis on $V_1 \otimes \dots \otimes V_d$). 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.

Proposition 10. $D_r^+ = \{X \in C_1 \otimes \dots \otimes C_d : \text{rank}_+(X) \leq r\}$ is a closed semialgebraic set if each $C_i \subseteq V_i^+$ is a closed semialgebraic cone.

Proof. The closedness follows from [20], and the fact that it is semialgebraic follows from Tarski–Seidenberg Theorem [7]. \square

By the closedness of D_r^+ , 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 [10, Theorem 27] for nonnegative tensors based on [10, Corollary 18].

Proposition 11. 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 $B(T, \delta)$ 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

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

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., $T' = T^*$. \square

We note that Proposition 11 holds more generally for arbitrary closed cones C_1, \dots, C_d in place of V_1^+, \dots, V_d^+ .

Proposition 12. The nonnegative tensors which have nonnegative rank $> r$ and do not have a unique best rank- r approximation form a semialgebraic set which does not contain an open set and is contained in some hypersurface.

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

$$\begin{aligned} \phi_r: (V_1^+ \times \cdots \times V_d^+)^r &\rightarrow V^+, \\ (u_{1,1}, \dots, u_{d,1}, \dots, u_{1,r}, \dots, u_{d,r}) &\mapsto \sum_{j=1}^r u_{1,j} \otimes \cdots \otimes u_{d,j}. \end{aligned}$$

Hence D_r^+ is semialgebraic by the Tarski–Seidenberg Theorem [7] and the required result follows from [11, Theorem 3.4]. \square

Proposition 12 remains true for arbitrary closed semialgebraic cones C_1, \dots, C_d .

Now we examine a useful necessary condition for $\sum_{p=1}^r T_p$ to be a best rank- r approximation of $T \in V_1 \otimes \cdots \otimes V_d$. We will borrow a standard notation from algebraic topology where a hat over a quantity means that quantity is omitted. So for example,

$$\begin{aligned} \widehat{u}_1 \otimes u_2 \otimes u_3 &= u_2 \otimes u_3, & u_1 \otimes \widehat{u}_2 \otimes u_3 &= u_1 \otimes u_3, & u_1 \otimes u_2 \otimes \widehat{u}_3 &= u_1 \otimes u_2, \\ u_1 \otimes \cdots \otimes \widehat{u}_i \otimes \cdots \otimes u_d &= u_1 \otimes \cdots \otimes u_{i-1} \otimes u_{i+1} \otimes \cdots \otimes u_d. \end{aligned}$$

The notation $\langle T, X \rangle$ denotes tensor contraction in all possible indices [18]. When T and X are of the same order, $\langle T, X \rangle$ reduces to inner product and our notation is consistent with the inner product notation. When T is a d -tensor and X is a $(d-1)$ -tensor, $\langle T, X \rangle$ is a vector — this is the only other case that will arise in our discussions below.

Lemma 13. Let $\text{rank}(T) > r$ and $\lambda \sum_{p=1}^r T_p$ be a best rank- r approximation, where $T_p = u_{1,p} \otimes \cdots \otimes u_{d,p}$ and $\|\sum_{p=1}^r T_p\| = 1$. If $\langle \sum_{p=1}^r T_p, u_{1,p} \otimes \cdots \otimes \widehat{u}_{i,p} \otimes \cdots \otimes u_{d,p} \rangle \neq 0$ for some $i \in \{1, \dots, d\}$, then

$$\langle T, u_{1,p} \otimes \cdots \otimes \widehat{u}_{i,p} \otimes \cdots \otimes u_{d,p} \rangle = \lambda \left\langle \sum_{p=1}^r T_p, u_{1,p} \otimes \cdots \otimes \widehat{u}_{i,p} \otimes \cdots \otimes u_{d,p} \right\rangle, \quad (\text{IV.1})$$

with $\lambda = \langle T, \sum_{p=1}^r T_p \rangle$.

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

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

So computing

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

is equivalent to computing

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

If at least one $\langle \sum_{p=1}^r T_p, u_{1,p} \otimes \cdots \otimes \widehat{u_{i,p}} \otimes \cdots \otimes u_{d,p} \rangle$ is nonzero, then the Jacobian matrix of the hypersurface defined by $\|\sum_{p=1}^r v_{1,p} \otimes \cdots \otimes v_{d,p}\| = 1$ around $(u_{1,1}, \dots, u_{d,1}, \dots, u_{1,r}, \dots, u_{d,r})$ has constant rank 1, *i.e.*, this real hypersurface is smooth at the point $(u_{1,1}, \dots, u_{d,1}, \dots, u_{1,r}, \dots, u_{d,r})$, so we can consider the Lagrangian

$$\mathcal{L} = \left\langle T, \sum_{p=1}^r v_{1,p} \otimes \cdots \otimes v_{d,p} \right\rangle - \lambda \left(\left\| \sum_{p=1}^r v_{1,p} \otimes \cdots \otimes v_{d,p} \right\| - 1 \right). \quad (\text{IV.2})$$

Then $\partial \mathcal{L} / \partial v_{i,p} = 0$ at the point $(u_{1,1}, \dots, u_{d,1}, \dots, u_{1,r}, \dots, u_{d,r})$ gives

$$\langle T, u_{1,p} \otimes \cdots \otimes \widehat{u_{i,p}} \otimes \cdots \otimes u_{d,p} \rangle = \lambda \left\langle \sum_{p=1}^r T_p, u_{1,p} \otimes \cdots \otimes \widehat{u_{i,p}} \otimes \cdots \otimes u_{d,p} \right\rangle \quad (\text{IV.3})$$

with $\lambda = \langle T, \sum_{p=1}^r T_p \rangle$ for all $i = 1, \dots, d$, and $p = 1, \dots, r$. □

V. UNIQUENESS OF RANK-ONE APPROXIMATIONS

We have established in Section IV that the best nonnegative rank- r approximation of a nonnegative tensor is generically unique. In this section we focus on the case $r = 1$, where we could say more. In particular, we will investigate some sufficient conditions under which a nonnegative tensor has a unique best nonnegative rank-1 approximation.

We begin with the following simple but useful observation: for a nonnegative tensor, a best rank-1 approximation can always be chosen to be a best nonnegative rank-1 approximation.

Lemma 14. *Given $T \in V^+$. Let $u_1 \otimes \cdots \otimes u_d \in V_1 \otimes \cdots \otimes V_d$ be a best rank-1 approximation of T , then u_1, \dots, u_d can be chosen to so that $u_1 \in V_1^+, \dots, u_d \in V_d^+$.*

Proof. Let $T = (T_{i_1, \dots, i_d})$ and $u_i = (u_i(1), \dots, u_i(n_i))$, where we have denoted the j th coordinate of a vector $u_i \in V_i$ by $u_i(j)$, $j = 1, \dots, n_i$. Then

$$\|T - u_1 \otimes \cdots \otimes u_d\|^2 = \sum_{i_1, \dots, i_d=1}^{n_1, \dots, n_d} (T_{i_1, \dots, i_d} - u_1(i_1) \cdots u_d(i_d))^2 \geq \sum_{i_1, \dots, i_d=1}^{n_1, \dots, n_d} (T_{i_1, \dots, i_d} - |u_1(i_1)| \cdots |u_d(i_d)|)^2.$$

Since $u_1 \otimes \cdots \otimes u_d$ is a best rank-1 approximation, then we can choose $u_i(j_i) = |u_i(j_i)|$, *i.e.*, $u_1 \otimes \cdots \otimes u_d \in V^+$. □

Motivated partly by the notion of singular pairs of a tensor [19] and partly by Lemma 13, we propose the following definition.

Definition 15. For $T \in V_1 \otimes \cdots \otimes V_d$, $(\lambda, u_1 \otimes \cdots \otimes u_d)$ is called a *nonnegative singular pair* of T if $\lambda \geq 0$ and there exist nonzero vectors $u_i \in V_i^+$ with

$$\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle = \lambda u_i, \quad \text{for } i = 1, \dots, d.$$

I would suggest to denote the singular value by σ , for two reasons: it is the usual practice for matrices, and λ has been already used earlier in the paper. If you are OK, I make the change.

The next two lemmas give an analogue of the tensorial Perron–Frobenius Theorem [4], [9], [19] for nonnegative singular pairs (as opposed to nonnegative eigenpairs).

Lemma 16 (Existence). *A nonnegative tensor T has at least one nonnegative singular pair.*

Proof. Let $D = \{(u_1, \dots, u_d) \in V_1^+ \times \cdots \times V_d^+ : \sum_{i=1}^d \|u_i\|_1 = 1\}$. Then D is a compact convex set. Define the map $\phi: D \rightarrow D$ by

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

Note that each term $\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle$ in the denominator is a 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 $\|\phi\|_1 = 1$.

Now if $\sum_{i=1}^d \|\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle\|_1 = 0$, 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 $\sum_{i=1}^d \|\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle\|_1 > 0$, 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 $\lambda = \sum_{i=1}^d \|\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle\|_1$. □

Lemma 17 (Uniqueness). *If T is positive, T has a nonnegative singular pair $(\lambda, u_1 \otimes \cdots \otimes u_d)$ with $\lambda > 0$. If $u_1 \otimes \cdots \otimes u_d$ has unit l^2 -norm, then $u_1 \otimes \cdots \otimes u_d$ is unique and $u_i > 0$, *i.e.*, every coordinate of u_i is positive, for every $i = 1, \dots, d$.*

Proof. We assume that a choice of bases has been fixed for V_1, \dots, V_d . We denote the j th coordinate of a vector $v_i \in V_i$ by $v_i(j)$, $j = 1, \dots, n_i$. Let $I_i = \{j : u_i(j) \neq 0\}$ be the support of u_i and set $\alpha = \min\{u_i(j) : i = 1, \dots, d, j \in I_i\}$. For any i and j ,

$$\lambda u_i(j) = \langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle(j) \geq \alpha^{d-1} \sum_{l_j \in I_j} T_{l_1 \dots l_{i-1} j l_{i+1} \dots l_d} > 0.$$

Suppose T has two positive singular vector tuples $u_1 \otimes \cdots \otimes u_d$ and $v_1 \otimes \cdots \otimes v_d$ corresponding to λ ,

$$\begin{aligned}\langle T, u_1 \otimes \cdots \otimes \widehat{u}_i \otimes \cdots \otimes u_d \rangle &= \lambda u_i, \quad i = 1, \dots, d, \\ \langle T, v_1 \otimes \cdots \otimes \widehat{v}_i \otimes \cdots \otimes v_d \rangle &= \lambda v_i, \quad i = 1, \dots, d.\end{aligned}$$

Let $\alpha_i = \max\{\alpha \in \mathbb{R}_+ : u_i - \alpha v_i \geq 0\}$ and $\beta_i = \max\{\beta \in \mathbb{R}_+ : v_i - \alpha u_i \geq 0\}$. Since $\|u_i\| = \|v_i\| = 1$ and $u_i, v_i > 0$, we must have $0 < \alpha_i, \beta_i \leq 1$. Therefore

$$\begin{aligned}\lambda u_i &= \langle T, u_1 \otimes \cdots \otimes \widehat{u}_i \otimes \cdots \otimes u_d \rangle \geq \langle T, \alpha_1 v_1 \otimes \cdots \otimes \widehat{\alpha_i v_i} \otimes \cdots \otimes \alpha_d v_d \rangle = \lambda \prod_{j \neq i} \alpha_j \cdot v_i, \\ \lambda v_i &= \langle T, v_1 \otimes \cdots \otimes \widehat{v}_i \otimes \cdots \otimes v_d \rangle \geq \langle T, \beta_1 u_1 \otimes \cdots \otimes \widehat{\beta_i u_i} \otimes \cdots \otimes \beta_d u_d \rangle = \lambda \prod_{j \neq i} \beta_j \cdot u_i.\end{aligned}$$

By the maximality of α_i ,

$$\frac{\prod_{j \neq i} \alpha_j}{\alpha_i} \leq 1$$

for each $i = 1, \dots, d$. Thus $\alpha_i = 1$, and similarly, $\beta_i = 1$ for $i = 1, \dots, d$. \square

A consequence of Lemmas 16 and 17 in the context of best nonnegative rank-1 approximation is as follows.

Theorem 18. *A positive tensor T of nonnegative rank at least one has a unique best nonnegative rank-1 approximation.*

Proof. Let \mathbb{S}^{n_i-1} denote the unit l^2 -sphere in V_i . Since the smooth function

$$\begin{aligned}\varphi: \mathbb{S}^{n_1-1} \times \cdots \times \mathbb{S}^{n_d-1} &\rightarrow \mathbb{R}_+, \\ (u_1, \dots, u_d) &\mapsto \langle T, u_1 \otimes \cdots \otimes u_d \rangle,\end{aligned}$$

attains its maximal value at some $(u_1, \dots, u_d) \in V_1^+ \times \cdots \times V_d^+$, the critical points of the Lagrangian

$$\mathcal{L} = \langle T, u_1 \otimes \cdots \otimes u_d \rangle - \sum_{i=1}^d \lambda_i (\|u_i\| - 1)$$

satisfy

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

for any $i = 1, \dots, d$. Furthermore, since $\langle T, u_1 \otimes \cdots \otimes u_d \rangle = \lambda_i$ for all $i = 1, \dots, d$, we get $\lambda_1 = \cdots = \lambda_d = \lambda$, where we let λ denote the common value.

Since λ is maximal, $\lambda > 0$ and $u_1 \otimes \cdots \otimes u_d$ is unique by Lemma 17. Hence this unique critical point $u_1 \otimes \cdots \otimes u_d$ yields the best rank-1 approximation, which is necessarily also the best nonnegative rank-1 approximation by Lemma 14. \square

The positivity requirement on T cannot in general be relaxed to nonnegativity. For example, the symmetric 3-tensor $x \otimes x \otimes x + y \otimes y \otimes y$, where x and y are orthonormal, has three best rank-1 approximations: $x \otimes x \otimes x$, $y \otimes y \otimes y$, and $((1/2)^{1/3}x + (1/2)^{1/3}y)^{\otimes 3}$.

Our mention of symmetric tensors naturally prompts the question: Could the best nonnegative rank-1 approximation of a nonnegative symmetric tensor be chosen to be symmetric? In the absence of nonnegativity, the answer is known to be ‘yes’ over \mathbb{R} and \mathbb{C} [1], [8].

We show below that for the case of positive tensors, a nonnegative tensor with *any* symmetry under permutation of its indices has a unique rank-1 nonnegative approximation with the same symmetry.

Definition 19. Let $d = d_1 + \cdots + d_k$ be an integer partition of d . We set $S^{(d_1, \dots, d_k)}(V) = S^{d_1}V_1 \otimes \cdots \otimes S^{d_k}V_k \subset V_1^{\otimes d_1} \otimes \cdots \otimes V_k^{\otimes d_k}$ to be the subspace of tensors that are *symmetric with respect to the partition* (d_1, \dots, d_k) , and $S^{(d_1, \dots, d_k)}(V^+)$ be the subset of nonnegative tensors of $S^{(d_1, \dots, d_k)}(V)$.

Note that for the special case $k = 1$, we get $S^d(V)$, the space of symmetric tensors on V .

Proposition 20. Each positive $T \in S^{(d_1, \dots, d_k)}(V^+)$ with $\text{rank} > 1$ has a unique best nonnegative rank-1 approximation that is (d_1, \dots, d_k) -symmetric.

Proof. By [8], T has a best rank-1 approximation that is (d_1, \dots, d_k) -symmetric. By Lemma 14 and Theorem 18, this approximation is unique. \square

VI. DEFLATABILITY

The relation between best rank- r and best rank-1 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-1 approximations — a consequence of the Eckhart–Young Theorem. It has been shown in [24] 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 [25] 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 21. Let $r \geq 1$. For almost every positive tensor T of nonnegative rank $> r + 1$, a best nonnegative rank- $(r + 1)$ approximation of T cannot be obtained from a best nonnegative rank-1 approximation of $T - T_{(r)}$, where $T_{(r)}$ is a best nonnegative rank- r approximation of T .

Proof. We will proceed by contradiction. Suppose T_p and \tilde{T}_p , $p = 1, \dots, r + 1$, are of rank 1 and satisfy

$$\begin{aligned} \left\| T - \lambda \sum_{p=1}^r T_p \right\| &= \min_{X \in D_r^+} \|T - X\|, \\ \left\| T - \mu \sum_{p=1}^{r+1} \tilde{T}_p \right\| &= \min_{X \in D_{r+1}^+} \|T - X\|, \end{aligned}$$

where $\|\sum_{p=1}^r T_p\| = \|\sum_{p=1}^{r+1} \tilde{T}_p\| = 1$, and $\lambda \sum_{p=1}^r T_p = \mu \sum_{p=1}^r \tilde{T}_p$. Hence we can assume $\tilde{T}_p = \alpha T_p$ and $T_p = u_{1,p} \otimes \cdots \otimes u_{d,p}$ for each $p \in \{1, \dots, r\}$, where $\alpha\mu = \lambda$, and $\tilde{T}_{r+1} = \alpha u_{1,r+1} \otimes \cdots \otimes u_{d,r+1}$.

By Lemma 13, we have for every $j = 1, \dots, r$, and $\ell = 1, \dots, r + 1$,

$$\left\langle T - \lambda \sum_{p=1}^r u_{1,p} \otimes \cdots \otimes u_{d,p}, u_{1,j} \otimes \cdots \otimes \widehat{u_{i,j}} \otimes \cdots \otimes u_{d,j} \right\rangle = 0, \quad (\text{VI.1})$$

$$\left\langle T - \mu \sum_{p=1}^{r+1} \alpha u_{1,p} \otimes \cdots \otimes u_{d,p}, \alpha u_{1,\ell} \otimes \cdots \otimes \widehat{u_{i,\ell}} \otimes \cdots \otimes u_{d,\ell} \right\rangle = 0. \quad (\text{VI.2})$$

We may simplify (VI.2) to get

$$\left\langle T - \lambda \sum_{p=1}^{r+1} u_{1,p} \otimes \cdots \otimes u_{d,p}, u_{1,\ell} \otimes \cdots \otimes \widehat{u_{i,\ell}} \otimes \cdots \otimes u_{d,\ell} \right\rangle = 0. \quad (\text{VI.3})$$

For $\ell = j$, we subtract (VI.1) from (VI.3) to get

$$\langle u_{1,r+1} \otimes \cdots \otimes u_{d,r+1}, u_{1,j} \otimes \cdots \otimes \widehat{u_{i,j}} \otimes \cdots \otimes u_{d,j} \rangle = 0$$

for $i = 1, \dots, d$, and $j = 1, \dots, r$. By the nonnegativity of the vector $u_{i,r+1}$,

$$\prod_{k \neq i} \langle u_{k,r+1}, u_{k,j} \rangle = 0, \quad (\text{VI.4})$$

and

$$\begin{aligned} \langle T, u_{1,r+1} \otimes \cdots \otimes \widehat{u_{i,r+1}} \otimes \cdots \otimes u_{d,r+1} \rangle &= \left\langle \lambda \sum_{p=1}^{r+1} u_{1,p} \otimes \cdots \otimes u_{d,p}, u_{1,r+1} \otimes \cdots \otimes \widehat{u_{i,r+1}} \otimes \cdots \otimes u_{d,r+1} \right\rangle \\ &= \lambda \prod_{k \neq i} \|u_{k,r+1}\|^2 \cdot u_{i,r+1}. \end{aligned}$$

In other words,

$$\left(\lambda \prod_{k=1}^d \|u_{k,r+1}\|, \frac{u_{1,r+1}}{\|u_{1,r+1}\|} \otimes \cdots \otimes \frac{u_{d,r+1}}{\|u_{d,r+1}\|} \right)$$

is a singular pair of T . So by Lemma 17, $u_{k,r+1} > 0$ for all $k = 1, \dots, d$, which contradicts (VI.4). \square

Following [25], 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 $r = 1, \dots, s$. Proposition 21 shows that a general nonnegative tensor does not admit a Schmidt–Eckart–Young decomposition.

ACKNOWLEDGMENT

LHL would like to thank Shmuel Friedland for helpful discussions (in particular for Rademacher Theorem). The work of YQ and PC is funded by the European Research Council under the European Community’s Seventh Framework Program FP7/2007–2013 Grant Agreement no. 320594. The work of LHL is funded by AFOSR Young Investigator Award FA9550-13-1-0133, NSF Collaborative Research Grant DMS 1209136, and NSF CAREER Award DMS 1057064.

REFERENCES

- [1] S. Banach, "Über homogene polynome in (L^2) ," *Studia Math.*, **7** (1938), no. 1, pp. 36–44.
- [2] M. Banagl, "The tensor product of function semimodules," *Algebra universalis*, vol. 70, no. 3, pp. 213–226, 2013.
- [3] C. Bocci, L. Chiantini, and G. Ottaviani, "Refined methods for the identifiability of tensors," *Annali di Matematica Pura ed Applicata*, May 2013.
- [4] 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.
- [5] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, "An algorithm for generic and low-rank specific identifiability of complex tensors," *SIAM J. matrix Ana. Appl.*, 2014, to appear.
- [6] P. Comon, "Tensors: a brief introduction," *IEEE Sig. Proc. Magazine*, vol. 31, no. 3, pp. 44–53, May 2014, special issue on BSS. hal-00923279.
- [7] V. De Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM Journal on Matrix Analysis Appl.*, vol. 30, no. 3, pp. 1084–1127, 2008.
- [8] S. Friedland, "Best rank-1 approximation of real symmetric tensors can be chosen symmetric," *Frontiers of Mathematics in China*, vol. 8, no. 1, pp. 19–40, 2013.
- [9] S. Friedland, S. Gaubert, and L. Han, "Perron–Frobenius theorem for nonnegative multilinear forms and extensions," *Linear Algebra Appl.*, vol. 438, pp. 738–749, 2013.
- [10] S. Friedland and G. Ottaviani, "The number of singular vector tuples and uniqueness of best rank-one approximation of tensors", Mar. 2014, arXiv:1210.8316, to appear in Foundations of Computational Mathematics.
- [11] S. Friedland and M. Stawiska, "Best approximation on semialgebraic sets and k -border rank approximation of symmetric tensors", Nov. 2013, arXiv:1311.1561.
- [12] L. D. Garcia, M. Stillman, and B. Sturmfels, "Algebraic geometry of Bayesian networks," *J. Symbolic Comput.*, **39** (2005), no. 3–4, pp. 331–355.
- [13] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, ser. Series in Computational Mathematics. Berlin, Heidelberg: Springer, 2012.
- [14] M. I. Jordan, "Graphical models," *Statistical Science*, vol. 19, no. 1, pp. 140–155, 2004.
- [15] D. Koller and N. Friedman, *Probabilistic Graphical Models*. MIT Press, 2009.
- [16] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions," *Linear Algebra and Applications*, vol. 18, pp. 95–138, 1977.
- [17] J. M. Landsberg, *Tensors: Geometry and Applications*, ser. Graduate Studies in Mathematics. AMS publ., 2012, vol. 128.
- [18] L.-H. Lim, "Tensors and hypermatrices," *Handbook of Linear Algebra*, 2nd Ed., CRC Press, Boca Raton, FL, 2013.
- [19] L.-H. Lim, "Singular values and eigenvalues of tensors: a variational approach," in *IEEE Int. Workshop on Comput. Adv. Multi-Sensor Adapt. Proc.*, Puerto Vallarta, Mexico, 13-15 Dec. 2005.
- [20] L.-H. Lim and P. Comon, "Nonnegative approximations of nonnegative tensors," *Jour. Chemometrics*, vol. 23, pp. 432–441, Aug. 2009, hal-00410056. [Online]. Available: <http://dx.doi.org/10.1002/cem.1244>
- [21] A. Shashua and T. Hazan, "Non-negative tensor factorization with applications to statistics and computer vision," in *22nd International Conference on Machine Learning*, Bonn, 2005, pp. 792–799.
- [22] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N -way arrays," *Jour. Chemo.*, vol. 14, pp. 229–239, 2000.
- [23] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis*. Chichester UK: Wiley, 2004.
- [24] A. Stegeman and P. Comon, "Subtracting a best rank-1 approximation does not necessarily decrease tensor rank," *Linear Algebra Appl.*, vol. 433, no. 7, pp. 1276–1300, Dec. 2010, hal-00512275.
- [25] N. Vannieuwenhoven, J. Nicaise, R. Vandebril, and K. Meerbergen, "On generic nonexistence of the Schmidt–Eckart–Young decomposition for complex tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 3, pp. 886–903, 2014.
- [26] P. Zhang, H. Wang, R. Plemmons, and P. Pauca, "Tensor methods for hyperspectral data analysis: A space object material identification study," *J. Optical Soc. Amer.*, vol. 25, no. 12, pp. 3001–3012, Dec. 2008.
- [27] J. Zhou, A. Bhattacharya, A. Herring, and D. Dunson, *Bayesian factorizations of big sparse tensors*, Jun. 2013, arXiv:1306.1598.