SINGULAR VALUES AND EIGENVALUES OF TENSORS: A VARIATIONAL APPROACH

Lek-Heng Lim

Stanford University
Institute for Computational and Mathematical Engineering
Gates Building 2B, Room 286, Stanford, CA 94305

ABSTRACT

We propose a theory of eigenvalues, eigenvectors, singular values, and singular vectors for tensors based on a constrained variational approach much like the Rayleigh quotient for symmetric matrix eigenvalues. These notions are particularly useful in generalizing certain areas where the spectral theory of matrices has traditionally played an important role. For illustration, we will discuss a multilinear generalization of the Perron-Frobenius theorem.

1. INTRODUCTION

It is well known that the eigenvalues and eigenvectors of a symmetric matrix A are the critical values and critical points of its Rayleigh quotient, $\mathbf{x}^{\mathsf{T}}A\mathbf{x}/\|\mathbf{x}\|_2^2$, or equivalently, the critical values and points of the quadratic form $\mathbf{x}^{\mathsf{T}}A\mathbf{x}$ constrained to vectors with unit l^2-norm, $\{\mathbf{x} \mid \|\mathbf{x}\|_2 = 1\}$. If $L : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is the associated Lagrangian with Lagrange multiplier λ,

$$L(\mathbf{x}, \lambda) = \mathbf{x}^{\mathsf{T}}A\mathbf{x} - \lambda(\|\mathbf{x}\|_2^2 - 1),$$

then the vanishing of ∇L at a critical point $(\mathbf{x}_c, \lambda_c) \in \mathbb{R}^n \times \mathbb{R}$ yields the familiar defining condition for eigenpairs

$$A\mathbf{x}_c = \lambda_c \mathbf{x}_c. \tag{1}$$

Note that this approach does not work if A is nonsymmetric — the critical points of L would in general be different from the solutions of (1).

A little less widely known is an analogous variational approach to the singular values and singular vectors of a matrix $A \in \mathbb{R}^{m\times n}$, with $\mathbf{x}^{\mathsf{T}}A\mathbf{y}/\|\mathbf{x}\|_2\|\mathbf{y}\|_2$ assuming the role of the Rayleigh quotient. The associated Lagrangian function $L : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is now

$$L(\mathbf{x}, \mathbf{y}, \sigma) = \mathbf{x}^{\mathsf{T}}A\mathbf{y} - \sigma(\|\mathbf{x}\|_2\|\mathbf{y}\|_2 - 1).$$

L is continuously differentiable for non-zero $\mathbf{x}, \mathbf{y}$. The first order condition yields

$$A\mathbf{y}_c/\|\mathbf{y}_c\|_2 = \sigma_c\mathbf{x}_c/\|\mathbf{x}_c\|_2, \quad A^{\mathsf{T}}\mathbf{x}_c/\|\mathbf{x}_c\|_2 = \sigma_c\mathbf{y}_c/\|\mathbf{y}_c\|_2,$$

at a critical point $(\mathbf{x}_c, \mathbf{y}_c, \sigma_c) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}$. Writing $\mathbf{u}_c = \mathbf{x}_c/\|\mathbf{x}_c\|_2$ and $\mathbf{v}_c = \mathbf{y}_c/\|\mathbf{y}_c\|_2$, we get the familiar

$$A\mathbf{v}_c = \sigma_c\mathbf{u}_c, \quad A^{\mathsf{T}}\mathbf{u}_c = \sigma_c\mathbf{v}_c. \tag{2}$$

Although it is not immediately clear how the usual definitions of eigenvalues and singular values via (1) and (2) may be generalized to tensors of order $k \geq 3$ (a matrix is regarded as an order-2 tensor), the constrained variational approach generalizes in a straight-forward manner — one simply replaces the bilinear functional $\mathbf{x}^{\mathsf{T}}A\mathbf{y}$ (resp. quadratic form $\mathbf{x}^{\mathsf{T}}A\mathbf{x}$) by the multilinear functional (resp. homogeneous polynomial) associated with a tensor (resp. symmetric tensor) of order k. The constrained critical values/points then yield a notion of singular values/vectors (resp. eigenvalues/vectors) for order-k tensors.

An important point of distinction between the order-2 and order-k cases is in the choice of norm for the constraints. At first glance, it may appear that we should retain the l^2-norm. However, the criticality conditions so obtained are no longer *scale invariant* (ie. the property that $\mathbf{x}_c$ in (1) or $(\mathbf{u}_c, \mathbf{v}_c)$ in (2) may be replaced by $\alpha\mathbf{x}_c$ or $(\alpha\mathbf{u}_c, \alpha\mathbf{v}_c)$ without affecting the validity of the equations). To preserve the scale invariance of eigenvectors and singular vectors for tensors of order $k \geq 3$, the l^2-norm must be replaced by the l^k-norm (where k is the order of the tensor),

$$\|\mathbf{x}\|_k = (|x_1|^k + \cdots + |x_n|^k)^{1/k}.$$

The consideration of eigenvalues and singular values with respect to l^p-norms where $p \neq 2$ is prompted by recent works [1, 2] of Choulakian, who studied such notions for matrices.

Nevertheless, we shall not insist on having scale invariance. Instead, we will define eigenpairs and singular pairs of tensors with respect to any l^p-norm ($p > 1$) as they can be interesting even when $p \neq k$. For example, when $p = 2$, our defining equations for singular values/vectors (6) become the equations obtained in the best rank-1 approximations of tensors studied by Comon [3] and de Lathauwer et. al. [4]. For the special case of symmetric tensors, our equations for eigenvalues/vectors for $p = 2$ and $p = k$ define respectively, the Z-eigenvalues/vectors and H-eigenvalues/vectors in the soon-to-appear paper [5] of Qi. For simplicity, we will restrict our study to integer-valued p in this paper.

We thank Gunnar Carlsson, Pierre Comon, Lieven de Lathauwer, Vin de Silva, and Gene Golub for helpful discussions. We would also like to thank Liqun Qi for sending us an advanced copy of his very relevant preprint.

2. TENSORS AND MULTILINEAR FUNCTIONALS

A k-array of real numbers representing an order-k tensor will be denoted by $A = [\![a_{j_1\cdots j_k}]\!] \in \mathbb{R}^{d_1\times\cdots\times d_k}$. Just as an order-2 tensor (ie. matrix) may be multiplied on the left and right by a pair of matrices (of consistent dimensions), an order-k tensor may be 'multiplied on k sides' by k matrices. The *covariant multilinear matrix multiplication* of A by matrices $M_1 = [m_{j_1 i_1}^{(1)}] \in$

This work appeared in: *Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* (CAMSAP '05), **1** (2005), pp. 129–132.

$\mathbb{R}^{d_1\times s_1}, \dots, M_k = [m^{(k)}_{j_k i_k}] \in \mathbb{R}^{d_k\times s_k}$ is defined by

$$A(M_1,\dots,M_k) := \left[\!\!\left[\sum_{j_1=1}^{d_1}\cdots\sum_{j_k=1}^{d_k} a_{j_1\cdots j_k} m^{(1)}_{j_1 i_1}\cdots m^{(k)}_{j_k i_k}\right]\!\!\right] \in \mathbb{R}^{s_1\times\cdots\times s_k}.$$

This operation arises from the way a multilinear functional transforms under compositions with linear maps. In particular, the multilinear functional associated with a tensor $A \in \mathbb{R}^{d_1\times\cdots\times d_k}$ and its gradient may be succinctly expressed via covariant multilinear multiplication:

$$A(\mathbf{x}_1,\dots,\mathbf{x}_k) = \sum_{j_1=1}^{d_1}\cdots\sum_{j_k=1}^{d_k} a_{j_1\cdots j_k} x^{(1)}_{j_1}\cdots x^{(k)}_{j_k}, \quad (3)$$
$$\nabla_{\mathbf{x}_i} A(\mathbf{x}_1,\dots,\mathbf{x}_k) = A(\mathbf{x}_1,\dots,\mathbf{x}_{i-1}, I_{d_i}, \mathbf{x}_{i+1},\dots,\mathbf{x}_k).$$

Note that we have slightly abused notations by using A to denote both the tensor and its associated multilinear functional.

An order-k tensor $[\![a_{j_1\cdots j_k}]\!] \in \mathbb{R}^{n\times\cdots\times n}$ is called *symmetric* if $a_{j_{\sigma(1)}\cdots j_{\sigma(k)}} = a_{j_1\cdots j_k}$ for any permutation $\sigma \in \mathfrak{S}_k$. The homogeneous polynomial associated with a symmetric tensor $A = [\![a_{j_1\cdots j_k}]\!]$ and its gradient can again be conveniently expressed as

$$A(\mathbf{x},\dots,\mathbf{x}) = \sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n} a_{j_1\cdots j_k} x_{j_1}\cdots x_{j_k}, \quad (4)$$
$$\nabla A(\mathbf{x},\dots,\mathbf{x}) = kA(I_n,\mathbf{x},\dots,\mathbf{x}).$$

Observe that for a symmetric tensor A,

$$A(I_n,\mathbf{x},\mathbf{x},\dots,\mathbf{x}) = A(\mathbf{x},I_n,\mathbf{x},\dots,\mathbf{x}) = \cdots = A(\mathbf{x},\mathbf{x},\dots,\mathbf{x},I_n). \quad (5)$$

The preceding discussion is entirely algebraic but we will now introduce norms on the respective spaces. Let $\|\cdot\|_{\alpha_i}$ be a norm on $\mathbb{R}^{d_i}$, $i = 1,\dots,k$. Then the *norm* (cf. [6]) of the multilinear functional $A : \mathbb{R}^{d_1}\times\cdots\times\mathbb{R}^{d_k} \to \mathbb{R}$ induced by $\|\cdot\|_{\alpha_1},\dots,\|\cdot\|_{\alpha_k}$ is defined as

$$\|A\|_{\alpha_1,\dots,\alpha_k} := \sup \frac{|A(\mathbf{x}_1,\dots,\mathbf{x}_k)|}{\|\mathbf{x}_1\|_{\alpha_1}\cdots\|\mathbf{x}_k\|_{\alpha_k}}$$

where the supremum is taken over all non-zero $\mathbf{x}_i \in \mathbb{R}^{d_i}$, $i = 1,\dots,k$. We will be interested in the case where the $\|\cdot\|_{\alpha_i}$'s are l^p-norms. Recall that for $1 \le p \le \infty$, the l^p-norm is a continuously differentiable function on $\mathbb{R}^n\backslash\{\mathbf{0}\}$. For $\mathbf{x} = [x_1,\dots,x_n]^\intercal \in \mathbb{R}^n$, we will write

$$\mathbf{x}^p := [x_1^p,\dots,x_n^p]^\intercal$$

(ie. taking pth power coordinatewise) and

$$\varphi_p(\mathbf{x}) := [\operatorname{sgn}(x_1)x_1^p,\dots,\operatorname{sgn}(x_n)x_n^p]^\intercal$$

where

$$\operatorname{sgn}(x) = \begin{cases} +1 & \text{if } x > 0,\\ 0 & \text{if } x = 0,\\ -1 & \text{if } x < 0.\end{cases}$$

Observe that if p is even, then $\varphi_p(\mathbf{x}) = \mathbf{x}^p$. The gradient of the l^p-norm is given by

$$\nabla\|\mathbf{x}\|_p = \frac{\varphi_{p-1}(\mathbf{x})}{\|\mathbf{x}\|_p^{p-1}}$$

or simply $\nabla\|\mathbf{x}\|_p = \mathbf{x}^{p-1}/\|\mathbf{x}\|_p^{p-1}$ when p is even.

3. SINGULAR VALUES AND SINGULAR VECTORS

Let $A \in \mathbb{R}^{d_1\times\cdots\times d_k}$. Then A defines a multilinear functional $A : \mathbb{R}^{d_1}\times\cdots\times\mathbb{R}^{d_k} \to \mathbb{R}$ via (3). Let us equip $\mathbb{R}^{d_i}$ with the l^{p_i}-norm, $\|\cdot\|_{p_i}$, $i = 1,\dots,k$. We will define the singular values and singular vectors of A as the critical values and critical points of $A(\mathbf{x}_1,\dots,\mathbf{x}_k)/\|\mathbf{x}_1\|_{p_1}\cdots\|\mathbf{x}_k\|_{p_k}$, suitably normalized. Taking a constrained variational approach, we let $L : \mathbb{R}^{d_1}\times\cdots\times\mathbb{R}^{d_k}\times\mathbb{R} \to \mathbb{R}$ be

$$L(\mathbf{x}_1,\dots,\mathbf{x}_k,\sigma) := A(\mathbf{x}_1,\dots,\mathbf{x}_k) - \sigma(\|\mathbf{x}_1\|_{p_1}\cdots\|\mathbf{x}_k\|_{p_k} - 1).$$

L is continuously differentiable when $\mathbf{x}_i \neq \mathbf{0}$, $i = 1,\dots,k$. The vanishing of the gradient,

$$\nabla L = (\nabla_{\mathbf{x}_1}L,\dots,\nabla_{\mathbf{x}_k}L,\nabla_\sigma L) = (\mathbf{0},\dots,\mathbf{0},0)$$

gives

$$\begin{aligned} A(I_{d_1},\mathbf{x}_2,\mathbf{x}_3,\dots,\mathbf{x}_k) &= \sigma\varphi_{p_1-1}(\mathbf{x}_1),\\ A(\mathbf{x}_1,I_{d_2},\mathbf{x}_3,\dots,\mathbf{x}_k) &= \sigma\varphi_{p_2-1}(\mathbf{x}_2),\\ &\vdots\\ A(\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_{k-1},I_{d_k}) &= \sigma\varphi_{p_k-1}(\mathbf{x}_k), \end{aligned} \quad (6)$$

at a critical point $(\mathbf{x}_1,\dots,\mathbf{x}_k,\sigma)$. As in the derivation of (2), one gets also the unit norm condition

$$\|\mathbf{x}_1\|_{p_1} = \cdots = \|\mathbf{x}_k\|_{p_k} = 1.$$

The unit vector $\mathbf{x}_i$ and σ in (6), will be called the *mode-i singular vector*, $i = 1,\dots,k$, and *singular value* of A respectively. Note that the mode-i singular vectors are simply the order-k equivalent of left- and right-singular vectors for order 2 (a matrix has two 'sides' or modes while an order-k tensor has k).

We will use the name $l^{p_1,\dots,p_k}$-singular values/vectors if we wish to emphasize the dependence of these notions on $\|\cdot\|_{p_i}$, $i = 1,\dots,k$. If $p_1 = \cdots = p_k = p$, then we will use the shorter name l^p-singular values/vectors. Two particular choices of p will be of interest to us: $p = 2$ and $p = k$ — both of which reduce to the matrix case when $k = 2$ (not so for other choices of p). The former yields

$$A(\mathbf{x}_1,\dots,\mathbf{x}_i,I_{d_i},\mathbf{x}_{i+1},\dots,\mathbf{x}_k) = \sigma\mathbf{x}_i, \quad i = 1,\dots,k,$$

while the latter yields a homogeneous system of equations that is invariant under scaling of $(\mathbf{x}_1,\dots,\mathbf{x}_k)$. In fact, when k is even, the l^p-singular values/vectors are solutions to

$$A(\mathbf{x}_1,\dots,\mathbf{x}_i,I_{d_i},\mathbf{x}_{i+1},\dots,\mathbf{x}_k) = \sigma\mathbf{x}_i^{k-1}, \quad i = 1,\dots,k.$$

The following results are easy to show. The first proposition follows from the definition of norm and the observation that a maximizer in an open set must be critical. The second proposition follows from the definition of hyperdeterminant [7]; the conditions on d_i are necessary and sufficient for the existence of Δ.

Proposition 1. *The largest $l^{p_1,\dots,p_k}$-singular value is equal to the norm of the multilinear functional associated with A induced by the norms $\|\cdot\|_{p_1},\dots,\|\cdot\|_{p_k}$, ie.*

$$\sigma_{\max}(A) = \|A\|_{p_1,\dots,p_k}.$$

Proposition 2. *Let $d_1, \dots, d_k$ be such that*

$$d_i - 1 \le \sum\nolimits_{j \ne i} (d_j - 1) \quad \textit{for all } i = 1, \dots, k,$$

and Δ denote the hyperdeterminant in $\mathbb{R}^{d_1 \times \cdots \times d_k}$. Then 0 is an l^2-singular value of $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ if and only if

$$\Delta(A) = 0.$$

4. EIGENVALUES AND EIGENVECTORS OF SYMMETRIC TENSORS

Let $A \in \mathbb{R}^{n \times \cdots \times n}$ be an order-k *symmetric* tensor. Then A defines a degree-k homogeneous polynomial function $A : \mathbb{R}^n \to \mathbb{R}$ via (4). With a choice of l^p-norm on $\mathbb{R}^n$, we may consider the multilinear Rayleigh quotient $A(\mathbf{x}, \dots, \mathbf{x})/\|\mathbf{x}\|_p^k$. The Lagrangian $L : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$,

$$L(\mathbf{x}, \lambda) := A(\mathbf{x}, \dots, \mathbf{x}) - \lambda(\|\mathbf{x}\|_p^k - 1),$$

is continuously differentiable when $\mathbf{x} \ne \mathbf{0}$ and

$$\nabla L = (\nabla_{\mathbf{x}} L, \nabla_\lambda L) = (\mathbf{0}, 0)$$

gives

$$A(I_n, \mathbf{x}, \dots, \mathbf{x}) = \lambda \varphi_{p-1}(\mathbf{x}) \tag{7}$$

at a critical point $(\mathbf{x}, \lambda)$ where $\|\mathbf{x}\|_p = 1$. The unit vector $\mathbf{x}$ and scalar λ will be called an *l^p-eigenvector* and *l^p-eigenvalue* of A respectively. Note that the LHS in (7) satisfies the symmetry in (5).

As in the case of singular values/vectors, the instances where $p = 2$ and $p = k$ are of particular interest. The l^2-eigenpairs are characterized by

$$A(I_n, \mathbf{x}, \dots, \mathbf{x}) = \lambda \mathbf{x}$$

where $\|\mathbf{x}\|_2 = 1$. When the order k is even, the l^k-eigenpairs are characterized by

$$A(I_n, \mathbf{x}, \dots, \mathbf{x}) = \lambda \mathbf{x}^{k-1} \tag{8}$$

and in this case the unit-norm constraint is superfluous since (8) is a homogeneous system and $\mathbf{x}$ may be scaled by any non-zero scalar α.

We shall refer the reader to [5] for some interesting results on l^2-eigenvalues and l^k-eigenvalues for symmetric tensors — many of which mirrors familiar properties of matrix eigenvalues.

5. EIGENVALUES AND EIGENVECTORS OF NONSYMMETRIC TENSORS

We know that one cannot use the variational approach to characterize eigenvalues/vectors of *nonsymmetric* matrices. So for an *nonsymmetric* tensor $A \in \mathbb{R}^{n \times \cdots \times n}$, we will instead *define* eigenvalues/vectors by (7) — an approach that is consistent with the matrix case. As (5) no longer holds, we now have k different forms of (7):

$$\begin{aligned} A(I_n, \mathbf{x}_1, \mathbf{x}_1, \dots, \mathbf{x}_1) &= \mu_1 \varphi_{p-1}(\mathbf{x}_1), \\ A(\mathbf{x}_1, I_n, \mathbf{x}_2, \dots, \mathbf{x}_2) &= \mu_2 \varphi_{p-1}(\mathbf{x}_2), \\ &\vdots \\ A(\mathbf{x}_k, \mathbf{x}_k, \dots, \mathbf{x}_k, I_n) &= \mu_k \varphi_{p-1}(\mathbf{x}_k). \end{aligned} \tag{9}$$

We will call the unit vector $\mathbf{x}_i$ a mode-i *eigenvector* of A corresponding to the mode-i *eigenvalue* μ_i, $i = 1, \dots, k$. Note that these are nothing more than the order-k equivalent of left and right eigenvectors.

6. APPLICATIONS

Several distinct generalizations of singular values/vectors and eigenvalues/vectors from matrices to higher-order tensors have been proposed in [3, 8, 4, 9, 5]. As one can expect, there is no one single generalization that preserves all properties of matrix singular values/vectors or matrix eigenvalues/vectors. In the lack of a canonical generalization, the validity of a multilinear generalization of a bilinear concept is often measured by the extent to which it may be applied to obtain interesting or useful results.

The proposed notions of l^2- and l^k-singular/eigenvalues arise naturally in the context of several different applications. We have mentioned the relation between the l^2-singular values/vectors and the best rank-1 approximation of a tensor under the Frobenius norm obtained in [3, 4]. Another example is the appearance of l^2-eigenvalues/vectors of symmetric tensors in the approximate solutions of constraint satisfaction problems [10, 11]. A third example is the use of l^k-eigenvalues for order-k symmetric tensors (k even) for characterizing the positive definiteness of homogeneous polynomial forms — a problem that is important in automatic control and array signal processing (see [5, 12] and the references cited therein).

Here we will give an application of l^k-eigenvalues and eigenvectors of a *nonsymmetric* tensor of order k. We will show that a multilinear generalization of the Perron-Frobenius theorem [13] may be deduced from the notion of l^k-eigenvalues/vectors as defined by (9).

Let $A = [\![a_{j_1 \cdots j_k}]\!] \in \mathbb{R}^{n \times \cdots \times n}$. We write $A > 0$ if all $a_{j_1 \cdots j_k} > 0$ (likewise for $A \ge 0$). We write $A > B$ if $A - B > 0$ (likewise for $A \ge B$).

An order-k tensor A is *reducible* if there exists a permutation $\sigma \in \mathfrak{S}_n$ such that the permuted tensor

$$[\![b_{i_1 \cdots i_k}]\!] = [\![a_{\sigma(j_1) \cdots \sigma(j_k)}]\!] \in \mathbb{R}^{n \times \cdots \times n}$$

has the property that for some $m \in \{1, \dots, n-1\}$, $b_{i_1 \cdots i_k} = 0$ for all $i_1 \in \{1, \dots, n-m\}$ and all $i_2, \dots, i_k \in \{1, \dots, m\}$.

If we allow a few analogous matrix terminologies, then A is reducible if there exists a permutation matrix P so that

$$B = A(P, \dots, P) \in \mathbb{R}^{n \times \cdots \times n}$$

can be partitioned into 2^n subblocks and regarded as a $2 \times \cdots \times 2$ block-tensor with 'square diagonal blocks' $B_{00 \cdots 0} \in \mathbb{R}^{m \times \cdots \times m}$, $B_{11 \cdots 1} \in \mathbb{R}^{(n-m) \times \cdots \times (n-m)}$, and a zero 'corner block' $B_{10 \cdots 0} \in \mathbb{R}^{(n-m) \times m \times \cdots \times m}$ which we may assume without loss of generality to be in the $(1, 0, \dots, 0)$-'corner'.

We say that A is *irreducible* if it is not reducible. In particular, if $A > 0$, then it is irreducible.

Theorem 1. *Let $A = [\![a_{j_1 \cdots j_k}]\!] \in \mathbb{R}^{n \times \cdots \times n}$ be irreducible and $A \ge 0$. Then A has a positive real l^k-eigenvalue with an l^k-eigenvector $\mathbf{x}_*$ that may be chosen to have all entries non-negative. In fact, $\mathbf{x}_*$ is unique and has all entries positive.*

Proof. Let $\mathbb{S}_+^n := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \ge 0, \|\mathbf{x}\|_k = 1\}$. For any $\mathbf{x} \in \mathbb{S}_+^n$, we define

$$\mu(\mathbf{x}) := \inf\{\mu \in \mathbb{R}_+ \mid A(I, \mathbf{x}, \dots, \mathbf{x}) \le \mu \mathbf{x}^{k-1}\}.$$

Note that for $\mathbf{x} \ge 0$, $\varphi_{k-1}(\mathbf{x}) = \mathbf{x}^{k-1}$. Since $\mathbb{S}_+^n$ is compact, there exists some $\mathbf{x}_* \in \mathbb{S}_+^n$ such that

$$\mu(\mathbf{x}_*) = \inf\{\mu(\mathbf{x}) \mid \mathbf{x} \in \mathbb{S}_k^n\} =: \mu_*.$$

Clearly,

$$A(I_n, \mathbf{x}_*, \dots, \mathbf{x}_*) \le \mu_* \mathbf{x}_*^{k-1}. \tag{10}$$

We claim that $\mathbf{x}_*$ is a (mode-1) l^k-eigenvector of A, ie.

$$A(I_n, \mathbf{x}_*, \dots, \mathbf{x}_*) = \mu_* \mathbf{x}_*^{k-1}.$$

Suppose not. Then at least one of the relations in (10) must hold with strict inequality. However, not *all* the relations in (10) can hold with strictly inequality since otherwise

$$A(I_n, \mathbf{x}_*, \dots, \mathbf{x}_*) < \mu_* \mathbf{x}_*^{k-1} \tag{11}$$

would contradict the definition of μ_* as an infimum. Without loss of generality, we may assume that the first m relations in (10) are the ones that hold with strict inequality and the remaining $n-m$ relations are the ones that hold with equality. We will write $\mathbf{x}_* = [\mathbf{x}_0, \mathbf{x}_1]^\intercal$ with $\mathbf{x}_0 \in \mathbb{R}^m, \mathbf{x}_1 \in \mathbb{R}^{n-m}$. By assumption, A may be partitioned into blocks so that

$$\begin{aligned} A_{00\cdots 0}(I_m, \mathbf{x}_0, \dots, \mathbf{x}_0, \mathbf{x}_0) &+ \\ A_{00\cdots 1}(I_m, \mathbf{x}_0, \dots, \mathbf{x}_0, \mathbf{x}_1) &+ \cdots + \\ A_{01\cdots 1}(I_m, \mathbf{x}_1, \dots, \mathbf{x}_1, \mathbf{x}_1) &< \mu_* \mathbf{x}_0^{k-1}, \end{aligned} \tag{12}$$

$$\begin{aligned} A_{10\cdots 0}(I_{n-m}, \mathbf{x}_0, \dots, \mathbf{x}_0, \mathbf{x}_0) &+ \\ A_{00\cdots 1}(I_{n-m}, \mathbf{x}_0, \dots, \mathbf{x}_0, \mathbf{x}_1) &+ \cdots + \\ A_{11\cdots 1}(I_{n-m}, \mathbf{x}_1, \dots, \mathbf{x}_1, \mathbf{x}_1) &= \mu_* \mathbf{x}_1^{k-1}. \end{aligned} \tag{13}$$

Note that $\mathbf{x}_0 \ne \mathbf{0}$ since the LHS of (12) is non-negative. We will fix $\mathbf{x}_1$ and consider the following (vector-valued) functions of $\mathbf{y}$:

$$\begin{aligned} F(\mathbf{y}) := A_{00\cdots 0}(I_m, \mathbf{y}, \dots, \mathbf{y}, \mathbf{y}) &+ \\ A_{00\cdots 1}(I_m, \mathbf{y}, \dots, \mathbf{y}, \mathbf{x}_1) &+ \cdots + \\ A_{01\cdots 1}(I_m, \mathbf{x}_1, \dots, \mathbf{x}_1, \mathbf{x}_1) &- \mu_* \mathbf{y}^{k-1}, \end{aligned}$$

$$\begin{aligned} G(\mathbf{y}) := A_{10\cdots 0}(I_{n-m}, \mathbf{y}, \dots, \mathbf{y}, \mathbf{y}) &+ \\ A_{00\cdots 1}(I_{n-m}, \mathbf{y}, \dots, \mathbf{y}, \mathbf{x}_1) &+ \cdots + \\ A_{11\cdots 1}(I_{n-m}, \mathbf{x}_1, \dots, \mathbf{x}_1, \mathbf{x}_1) &- \mu_* \mathbf{x}_1^{k-1}. \end{aligned}$$

Let $f_1, \dots, f_m$ and $g_1, \dots, g_{n-m}$ be the component functions of F and G respectively, ie. f_i's and g_i's are real-valued functions of $\mathbf{y} \in \mathbb{R}^m$ such that $F(\mathbf{y}) = [f_1(\mathbf{y}), \dots, f_m(\mathbf{y})]^\intercal$ and $G(\mathbf{y}) = [g_1(\mathbf{y}), \dots, g_{n-m}(\mathbf{y})]^\intercal$.

By (12), we get $f_i(\mathbf{x}_0) < 0$ for $i = 1, \dots, m$. Since f_i is continuous, there is a neighborhood $B(\mathbf{x}_0, \delta_i) \subseteq \mathbb{R}^m$ such that $f_i(\mathbf{y}) < 0$ for all $\mathbf{y} \in B(\mathbf{x}_0, \delta_i)$. Let $\delta = \min\{\delta_1, \dots, \delta_m\}$. Then $F(\mathbf{y}) < \mathbf{0}$ for all $\mathbf{y} \in B(\mathbf{x}_0, \delta)$.

By (13), we get $g_j(\mathbf{x}_0) = 0$ for $j = 1, \dots, n-m$. Observe that if g_j is not identically 0, then $g_j(\mathbf{y}) = g_j(y_1, \dots, y_m)$ is a non-constant multivariate polynomial function in the variables $y_1, \dots, y_m$. Furthermore, all coefficients of this multivariate polynomial are non-negative since $A \ge 0$. It is easy to see that such a function must be 'strictly monotone' in the following sense: if $\mathbf{0} \le \mathbf{y} \le \mathbf{z}$ and $\mathbf{z} \ne \mathbf{y}$, then $g_j(\mathbf{y}) < g_j(\mathbf{z})$. So for $\mathbf{0} \le \mathbf{y} \le \mathbf{x}_0$ and $\mathbf{y} \ne \mathbf{x}_0$, we get $g_j(\mathbf{y}) < g_j(\mathbf{x}_0) = 0$. Since A is irreducible, $A_{10\cdots 0}$ is non-zero and thus some g_j is not identically 0.

Let $\mathbf{y}_0$ be a point on the line joining $\mathbf{0}$ to $\mathbf{x}_0$ within a distance δ of $\mathbf{x}_0$ and $\mathbf{y}_0 \ne \mathbf{x}_0$. Then with $\mathbf{y}_0$ in place of $\mathbf{x}_0$, the m strict inequalities in (12) are retained while at least one equality in (13) will have become a strict inequality. Note that the homogeneity of (12) and (13) allows us to scale $[\mathbf{y}_0, \mathbf{x}_1]^\intercal$ to unit l^k-norm without affecting the validity of the inequalities and equalities. Thus we have obtained a solution with at least $m+1$ relations in (10) being strict inequalities. Repeating the same arguments inductively, we can eventually replace all the equalities in (12) with strict inequalities, leaving us with (11), a contradiction. [*We defer the proof of uniqueness and positivity of* $\mathbf{x}_*$ *to the full paper.*] $\square$

A proposal to use the multilinear Perron-Frobenius theorem in the ranking of linked objects may be found in [14]. A symmetric version of this result can be used to study hypergraphs [15].

7. REFERENCES

[1] V. Choulakian, "Transposition invariant principal component analysis in l_1 for long tailed data," *Statist. Probab. Lett.*, vol. 71, no. 1, pp. 23–31, 2005.

[2] V. Choulakian, "l_1-norm projection pursuit principal component analysis," *Comput. Statist. Data Anal.*, 2005, to appear.

[3] P. Comon, "Tensor decompostions: state of the art and applications," Oxford, UK, 2002, number 71 in Inst. Math. Appl. Conf. Ser., pp. 1–24, Oxford Univ. Press.

[4] L. de Lathauwer, B. de Moor, and J. Vandewalle, "On the best rank-1 and rank-$(r_1, \dots, r_N)$ approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, 2000.

[5] L. Qi, "Eigenvalues of a real supersymmetric tensor," *J. Symbolic Comput.*, 2005, to appear.

[6] A. Defant and K. Floret, *Tensor norms and operator ideals*, Number 176 in North-Holland Mathematics Studies. North-Holland, Amsterdam, 1993.

[7] I.M. Gelfand, M.M. Kapranov, and A.V. Zelevinsky, "Hyperdeterminants," *Adv. Math.*, vol. 96, no. 2, pp. 226–263, 1992.

[8] L. de Lathauwer, B. de Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[9] J.R. Ruíz-Tolosa and E. Castillo, *From vectors to tensors*, Universitext. Springer-Verlag, Berlin, 2005.

[10] N. Alon, W.F. de la Vega, R. Kannan, and M. Karpinski, "Random sampling and approximation of max-csps," *J. Comput. System Sci.*, vol. 67, no. 2, pp. 212–243, 2003.

[11] W.F. de la Vega, R. Kannan, M. Karpinski, and S. Vempala, "Tensor decomposition and approximation algorithms for constraint satisfaction problems," New York, NY, USA, 2005, pp. 747–754, ACM Press.

[12] L. Qi and K.L. Teo, "Multivariate polynomial minimization and its application in signal processing," *J. Global Optim.*, vol. 26, no. 4, pp. 419–433, 2003.

[13] A. Berman and R.J. Plemmons, *Nonnegative matrices in the mathematical sciences*, Number 9 in Classics in applied mathematics. SIAM, Philadelphia, PA, 1994.

[14] L.-H. Lim, "Multilinear pagerank: measuring higher order connectivity in linked objects," The Internet: Today & Tomorrow, July 2005.

[15] P. Drineas and L.-H. Lim, "A multilinear spectral theory of hypergraphs and expander hypergraphs," work in progress, 2005.