Non-Negative Matrix Factorization Revisited: Uniqueness and Algorithm for Symmetric Decomposition

Kejun Huang, *Student Member, IEEE*, Nicholas D. Sidiropoulos, *Fellow, IEEE*, and Ananthram Swami, *Fellow, IEEE*

***Abstract*—Non-negative matrix factorization (NMF) has found numerous applications, due to its ability to provide interpretable decompositions. Perhaps surprisingly, existing results regarding its uniqueness properties are rather limited, and there is much room for improvement in terms of algorithms as well. Uniqueness aspects of NMF are revisited here from a geometrical point of view. Both symmetric and asymmetric NMF are considered, the former being tantamount to element-wise non-negative square-root factorization of positive semidefinite matrices. New uniqueness results are derived, e.g., it is shown that a sufficient condition for uniqueness is that the conic hull of the latent factors is a superset of a particular second-order cone. Checking this condition is shown to be NP-complete; yet this and other results offer insights on the role of latent sparsity in this context. On the computational side, a new algorithm for symmetric NMF is proposed, which is very different from existing ones. It alternates between Procrustes rotation and projection onto the non-negative orthant to find a non-negative matrix close to the span of the dominant subspace. Simulation results show promising performance with respect to the state-of-art. Finally, the new algorithm is applied to a clustering problem for co-authorship data, yielding meaningful and interpretable results.**

***Index Terms*—Nonnegative matrix factorization (NMF), symmetric NMF, uniqueness, sparsity, Procrustes rotation.**

I. Introduction

A. Background and Motivation

NON-NEGATIVE Matrix Factorization (NMF) is the problem of (approximately) factoring $\mathbf{S} = \mathbf{W}\mathbf{H}$, where $\mathbf{S}$ is $I \times J$, $\mathbf{W}$ is $I \times K$, with $K < \min(I, J)$, $\mathbf{H}$ is $K \times J$, $\mathbf{W} \geq 0$, $\mathbf{H} \geq 0$ (inequalities interpreted element-wise). The smallest possible K for which such decomposition is possible is the *non-negative rank* of $\mathbf{S}$. Due to the non-negativity constraints, the non-negative rank can be higher than the usual matrix rank over the real field.

NMF has been studied for more than 30 years [2]–[4], originally known as *non-negative rank factorization* or *positive matrix factorization*. Non-negative matrices have many interesting properties and a long history in science and engineering [5]. Lee and Seung [6] popularized NMF when they discovered that it tends to decompose images of visual objects in meaningful parts—i.e., NMF "is able to learn the parts of objects". NMF quickly found numerous other applications in diverse disciplines—see [7] and references therein. Unfortunately, it was recently shown that (asymmetric) NMF is NP-hard [8]; yet sustained interest in NMF has produced many good algorithms, including optimization-based methods [9]–[13] and geometry-based methods [14]–[18].

NMF has been such a success story across disciplines because non-negativity is a valid constraint in so many applications, *and* NMF often provides meaningful/interpretable results, and sometimes even 'correct' results—that is, it yields the true latent factors $\mathbf{W}$, $\mathbf{H}$. As an example, matrix $\mathbf{S}$ could represent an $I \times J$ keyword by document incidence matrix, wherein entry $\mathbf{S}_{i,j}$ is the number of occurences of keyword i in document j. The K-dimensional reduction $\mathbf{S} = \mathbf{W}\mathbf{H}$ can be interpreted as follows. The columns of $\mathbf{W}$ represent K document prototypes as consisting of keywords in specific proportions. The columns of $\mathbf{H}$ represent each document as a weighted combination of these prototypes. In a clustering scenario, the columns of $\mathbf{W}$ are the cluster centroids, and the columns of $\mathbf{H}$ are the (soft) cluster membership indicators. A specific application of clustering via NMF is [19], where NMF was applied to an article by journal matrix, resulting in a soft clustering of articles by topic. See also Xu *et al.* [20] for an application of NMF to document clustering. Another interpretation of $\mathbf{S} = \mathbf{W}\mathbf{H}$ is that the i-th row of $\mathbf{W}$ is a reduced $1 \times K$ representation of the i-th 'keyword' or 'concept' in latent K-dimensional space, and likewise the j-th column of $\mathbf{H}$ is the reduced $K \times 1$ representation of the j-th document *in the same latent space*; $\mathbf{S}_{i,j}$ is the inner product of these two latent representations, measuring 'relevance'.

Uniqueness of NMF is tantamount to the question of whether or not these true latent factors are *the only* interpretation of the data, or alternative ones exist. Unfortunately, NMF is in general non-unique. One can inquire about existence and uniqueness of NMF of $\mathbf{S}$ for a given K without any other side information; or about uniqueness of a *particular factorization* $\mathbf{S} = \mathbf{W}_o\mathbf{H}_o$, i.e., given a particular pair of factors $\mathbf{W}_o$, $\mathbf{H}_o$ as side information.

Manuscript received February 06, 2013; revised June 06, 2013; accepted September 25, 2013. Date of publication October 11, 2013; date of current version December 12, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Lawrence Carin. Supported in part by ARO STIR W911NF-11-1-0500, and NSF IIS-1247632. Conference version of part of the results appears in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) Vancouver, CA, USA, May 26–31, 2013 [1].

K. Huang and N. D. Sidiropoulos are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: huang663@umn.edu; nikos@umn.edu).

A. Swami is with the Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: a.swami@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2013.2285514

Additional constraints can be added to help make the factorization unique, e.g., sparsity [21] or minimum determinant [22]. Thomas [23] and Chen [3] first gave different geometric interpretations of NMF, and stated the uniqueness problem in a geometric way. Donoho *et al.* [24] and Laurberg *et al.* [25] later provided uniqueness conditions by exploiting the aforementioned geometric viewpoint. The sufficient conditions they provided, however, require one of the two matrix factors to contain a scaled identity matrix—a particularly strict condition. Moussaoui *et al.* [26] and Laurberg *et al.* [25] also provided necessary conditions. We will review these conditions in Section III, and discuss their relationship with the uniqueness conditions we provide herein.

The symmetric version of NMF ($\mathbf{S} = \mathbf{W}\mathbf{W}^T$ with $\mathbf{W}$ having non-negative elements) is relatively less studied. Notice that $\mathbf{W}$ in symmetric NMF is a square-root factor of the symmetric positive semi-definite matrix $\mathbf{S}$ that is also element-wise non-negative. Thus symmetric NMF is an element-wise non-negative square-root factorization of positive semidefinite matrices. In the mathematical programming field, such matrices are called *completely positive matrices* [27]. It has recently been shown that checking whether a positive semidefinite matrix is completely positive is also NP-hard [28]. Algorithmic work on symmetric NMF is very limited relative to asymmetric NMF. The state-of-art in terms of symmetric NMF algorithms can be found in He *et al.* [29], which proposed three algorithms—the multiplicative update, α-SNMF and β-SNMF, and showed that the latter two outperform other alternatives. Other papers on symmetric NMF include Catral *et al.* [30] who studied the symmetric NMF problem for $\mathbf{S}$ that is not necessarily positive semidefinite, and conditions under which asymmetric NMF yields a symmetric approximation; and Ding *et al.* [31], who developed interesting links between symmetric NMF and 'soft' k-means clustering.

The contributions of the paper are as follows. Adopting a geometric viewpoint similar to [3], [23]–[25], but taking a different approach, new results and perspectives on uniqueness of NMF are derived here. A subtlety of NMF is that properties that hold in the asymmetric case do not necessarily hold in the symmetric case. Our novel necessary conditions and sufficient conditions for uniqueness hold for both the symmetric and asymmetric cases. The necessary conditions in Theorem 3 have a nice interpretation in terms of the *support* sets (the set of cells with positive weights) of the columns (rows) of $\mathbf{W}(\mathbf{H})$. In terms of the document clustering example discussed earlier, the interpretation is that two different prototype documents cannot be composed of the same set of keywords (even if in different proportions). In other words, every pair of prototype documents must differ in the inclusion of at least one keyword. The necessary conditions in Theorem 3 are easy to verify. We also establish a new sufficient condition for uniqueness (Theorem 4): we show that the conic hulls of the latent factors (the $\mathbf{W}$ and $\mathbf{H}$ matrices) must be supersets of a particular second-order cone. We demonstrate via examples that our sufficient conditions are broader than those currently known. However, checking the sufficient condition is proved to be NP-complete. Corollary 2 provides a nice connection with sparsity of the original matrix and the latent factors. For the symmetric case, we develop a novel algorithm based on an alternating approach and the use of Procrustes projections. The complexity of each iteration is $O(IK^2)$ in contrast with $O(I^2K)$ in the SNMF approaches of [29], which represent the state-of-the-art for symmetric NMF. Recall that $K \leq I$, and in typical cases $K \ll I$, so that the proposed algorithms are computationally cheap. Simulations on synthetic data indicate that the proposed algorithms converge significantly faster than SNMF. We also applied our symmetric NMF algorithm to a clustering problem for co-authorship data, yielding meaningful and interpretable results.

A conference summary of part of our results has been presented at ICASSP 2013 [1]. Relative to [1], this journal version includes proofs, additional analytical and algorithmic results, and further experiments.

B. Notation

A scalar is denoted by an italic letter, e.g., a or A. A column vector is denoted by a bold lowercase letter, e.g., $\mathbf{a}$. The i-th entry of $\mathbf{a}$ is a_i. A matrix is denoted by a bold uppercase letter, e.g., $\mathbf{A}$, where $\mathbf{A}_{i,j}$ is its i,j-th entry, $\mathbf{A}_{i,:}$ is the i-th row of $\mathbf{A}$, and $\mathbf{A}_{:,j}$ is the j-th column of $\mathbf{A}$. A set is denoted by a calligraphic uppercase letter, e.g., $\mathcal{A}$. $\mathbb{R}^n_+ = \{\mathbf{x}|x_i \geq 0, i = 1,\ldots,n\}$ is the positive orthant in $\mathbb{R}^n$. $\mathbf{e}_i, \mathbf{1}, \mathbf{0}$ are the i-th standard coordinate vector, all ones vector, and the zero vector, respectively. By using $\mathbf{e}_i$ we can also represent the i-th row and i-th column of $\mathbf{A}$ as $\mathbf{e}_i^T\mathbf{A}$ and $\mathbf{A}\mathbf{e}_i$.

Inequality marks represent element-wise inequalities, whether applied to scalars, vectors or matrices. Symmetric NMF is written as $\mathbf{S} = \mathbf{W}\mathbf{W}^T$, where $\mathbf{S}$ is $I \times I$ symmetric positive semi-definite, $\mathbf{W}$ is $I \times K$. We focus on the low-rank case, so $K < \min(I, J)$. Without loss of generality, we assume there is no all-zero column or row in any matrix. If this happens, we can simply delete it (them) first.

II. Convex Analysis Preliminaries

Before we analyze the properties of NMF, we briefly review some prerequisites from convex analysis; see [32], [33] for further background.

Definition 1 (Polyhedral Cone): A polyhedral cone $\mathcal{K}$ is a set that is both a polyhedron and a cone.

There are two ways to describe a polyhedral cone. The first is by taking the intersection of a number of half-spaces, which takes the form

$$\mathcal{K} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{A}^T\mathbf{x} \geq \mathbf{0}\}$$

where each column of $\mathbf{A}$ defines a half-space which contains the origin at the boundary. Notice that the right hand side is $\mathbf{0}$ in order to make this set a cone. Assuming there are no redundant constraints, the i-th constraint satisfied as equality is called a *facet* of $\mathcal{K}$.

The other way to describe a polyhedral cone is by taking the conic hull of a number of vectors, i.e.,

$$\mathcal{K} = \{\mathbf{x} = \mathbf{B}\boldsymbol{\lambda} | \boldsymbol{\lambda} \geq \mathbf{0}\} = \mathrm{cone}(\mathbf{B})$$

where the columns of $\mathbf{B}$ are the vectors we are taking. If a column of $\mathbf{B}$ cannot be represented by the conic combination (non-negative linear combination) of the rest of the columns, then it is called an *extreme ray* of $\mathcal{K}$.

Given a polyhedral cone represented by the intersection of half-spaces, an extreme ray of the cone would be a vector satisfying all the inequality constraints, and furthermore, at least $n-1$ linearly independent constraints are satisfied as equalities. Similarly, given a polyhedral cone represented by the conic hull of vectors, a facet of it is a hyperplane defined by at least $n-1$ linearly independent extreme rays (and the origin), and the rest of the extreme rays must be on one side of that hyperplane.

Definition 2 (Simplicial Cone): A simplicial cone is a polyhedral cone such that all of its extreme rays are linearly independent.

If $\mathcal{K} = \{\mathbf{x} = \mathbf{B}\boldsymbol{\lambda} | \boldsymbol{\lambda} \geq \mathbf{0}\} = \operatorname{cone}(\mathbf{B})$ is a simplicial cone, then for every element $\mathbf{x} \in \mathcal{K}$, there is a *unique* corresponding $\boldsymbol{\lambda}$ that indicates how to conically combine the extreme rays to generate $\mathbf{x}$. For general polyhedral cones, this combination is in most cases not unique.

Moreover, it is easy to change the representation of a simplicial cone $\mathcal{K}$ from halfspace-intersection to conic hull. For example, if $\mathcal{K} = \operatorname{cone}(\mathbf{B})$ where $\mathbf{B}$ is invertible, then $\mathcal{K}$ is a simplicial cone and $\mathcal{K} = \{\mathbf{x} | \mathbf{B}^{-1}\mathbf{x} \geq \mathbf{0}\}$. However, for general polyhedral cones this switching between representations is a hard problem [34].

Definition 3 (Dual Cone): The dual cone of a set $\mathcal{K}$, denoted by $\mathcal{K}^*$, is defined as $\mathcal{K}^* = \{\mathbf{y} | \mathbf{x}^T\mathbf{y} \geq 0, \forall \mathbf{x} \in \mathcal{K}\}$.

Some important properties of dual cones are as follows (cf. Laurberg *et al.* [25]):

Property 1: $\operatorname{cone}(\mathbf{A})^* = \{\mathbf{x} | \mathbf{A}^T\mathbf{x} \geq 0\}$.

For one column of the matrix $\mathbf{A}$, in the primal cone it defines an extreme ray, while in the dual cone it defines a facet. Therefore, there is a one-to-one correspondence between the extreme rays of the primal polyhedral cone and facets of the dual polyhedral cone, and vice versa.

Property 2: If $\mathbf{A}$ is invertible, then $\operatorname{cone}(\mathbf{A})^* = \operatorname{cone}(\mathbf{A}^{-T})$.

From this property, it is easy to see that if $\mathbf{A}$ is unitary, i.e., $\mathbf{A}^{-1} = \mathbf{A}^T$, then $\operatorname{cone}(\mathbf{A})^* = \operatorname{cone}(\mathbf{A}^{-T}) = \operatorname{cone}(\mathbf{A})$. In other words, the conic hull of a unitary matrix is *self dual*.

Property 3: If $\mathcal{A}$ and $\mathcal{B}$ are convex cones, and $\mathcal{A} \subseteq \mathcal{B}$, then $\mathcal{B}^* \subseteq \mathcal{A}^*$.

Here is an example of a cone and its dual cone, which will be useful later. Donoho and Stodden also studied the following cones in [24].

Example 1: Define the second-order cone in $\mathbb{R}^n$

$$\mathcal{C} = \{\mathbf{x} | \mathbf{x}^T\mathbf{1} \geq \sqrt{n-1}\|\mathbf{x}\|_2\} \tag{1}$$

Its dual cone is another second-order cone

$$\mathcal{C}^* = \{\mathbf{x} | \mathbf{x}^T\mathbf{1} \geq \|\mathbf{x}\|_2\} \tag{2}$$

The reason we are interested in $\mathcal{C}$ and its dual cone is because they have a very special relationship with the non-negative orthant $\mathbb{R}^n_+$: $\mathcal{C} \subseteq \mathbb{R}^n_+ \subseteq \mathcal{C}^*$. Fig. 1 gives a graphical view of $\mathcal{C}$, $\mathcal{C}^*$ and $\mathbb{R}^n_+$ in $\mathbb{R}^3$. In fact, $\mathbb{R}^n_+$ and its rotated versions are the only simplicial cones that satisfy this, as stated in the following lemma.

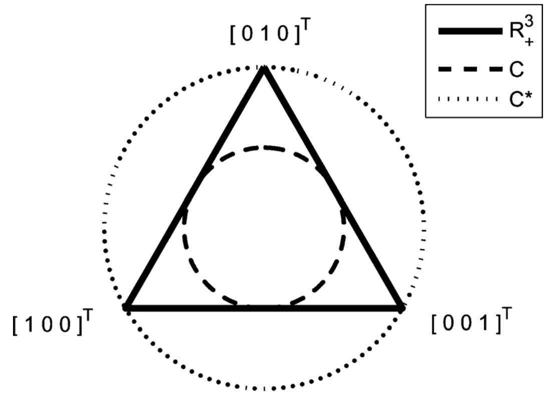

Fig. 1. A 2-D slice view of the relationship between $\mathbb{R}^3_+, \mathcal{C}, \mathcal{C}^*$ in $\mathbb{R}^3$, looking at the plane $\mathbf{1}^T\mathbf{x} = 1$.

Lemma 1: If $\operatorname{cone}(\mathbf{A})$ satisfies that $\mathcal{C} \subseteq \operatorname{cone}(\mathbf{A}) \subseteq \mathcal{C}^*$, and the columns of $\mathbf{A}$ are scaled to have unit l_2 norm, then

1) $\mathbf{A}^T\mathbf{1} = \mathbf{1}$
2) $\mathbf{A}^T\mathbf{A} = \mathbf{I}$

Proof: See Appendix A. ■

III. Uniqueness of NMF: Known Results

Recall that we are interested in the factorization $\mathbf{S} = \mathbf{W}\mathbf{H}$ where the $I \times K$ matrix $\mathbf{W}$ and the $K \times J$ matrix $\mathbf{H}$ have non-negative elements. For uniqueness analysis, we shall assume that $K = \operatorname{rank}(\mathbf{S})$, and thus both $\mathbf{W}$ and $\mathbf{H}$ are full rank. (Notice that if $\mathbf{W}$ and $\mathbf{H}$ are drawn from a jointly continuous distribution and $\mathbf{S}$ is in fact constructed by taking their product, then $\operatorname{rank}(\mathbf{S}) = K$ almost surely.) Therefore, if $\mathbf{S} = \mathbf{W}\mathbf{H} = \tilde{\mathbf{W}}\tilde{\mathbf{H}}$, then there exists a $K \times K$ full rank matrix $\mathbf{A}$ such that $\tilde{\mathbf{W}} = \mathbf{W}\mathbf{A}$ and $\tilde{\mathbf{H}} = \mathbf{A}^{-1}\mathbf{H}$. A trivial choice of $\mathbf{A}$ would be a positively scaled permutation matrix; such ambiguity is unavoidable without side information.

Definition 4 (Uniqueness of Asymmetric NMF): The NMF of $\mathbf{S} = \mathbf{W}\mathbf{H}$ is said to be (essentially) unique if $\mathbf{S} = \tilde{\mathbf{W}}\tilde{\mathbf{H}}$ implies $\tilde{\mathbf{W}} = \mathbf{W}\mathbf{P}\mathbf{D}$ and $\tilde{\mathbf{H}} = (\mathbf{P}\mathbf{D})^{-1}\mathbf{H}$, where $\mathbf{D}$ is a diagonal matrix with its diagonal entries positive, and $\mathbf{P}$ is a permutation matrix.

In the case of symmetric NMF, if $\mathbf{S} = \mathbf{W}\mathbf{W}^T = \tilde{\mathbf{W}}\tilde{\mathbf{W}}^T$, there exists a real $K \times K$ orthonormal matrix $\mathbf{A}$ such that $\tilde{\mathbf{W}} = \mathbf{W}\mathbf{A}^T$ (to maintain symmetry). This removes the scaling ambiguity.

Definition 5 (Uniqueness of Symmetric NMF): The symmetric NMF of $\mathbf{S} = \mathbf{W}\mathbf{W}^T$ is said to be (essentially) unique if $\mathbf{S} = \tilde{\mathbf{W}}\tilde{\mathbf{W}}^T$ implies $\tilde{\mathbf{W}} = \mathbf{W}\mathbf{P}$, where $\mathbf{P}$ is a permutation matrix.

Using the convex analysis fundamentals in Section II, the uniqueness of NMF has been approached from two different points of view, which are summarized next.

A. Donoho's Analysis

The stepping stone in Donoho's analysis is the following lemma, which can be traced back to [3].

Lemma 2: If $\mathrm{rank}(\mathbf{S}) = K$, the NMF of the non-negative matrix $\mathbf{S}$ is unique if and only if there is a unique simplicial cone $\mathcal{W}$ with K extreme rays satisfying

$$\mathrm{cone}(\mathbf{S}) \subseteq \mathcal{W} \subseteq \mathcal{P}_{\mathcal{V}} \tag{3}$$

where $\mathcal{P}_{\mathcal{V}}$ is defined as

$$\mathcal{P}_{\mathcal{V}} = \{\mathbf{w} | \mathbf{U}^T\mathbf{w} = 0, \mathbf{w} \geq 0\} \tag{4}$$

and $\mathbf{U}$ is a matrix whose columns span $\mathrm{Null}(\mathbf{S}^T)$.

Once we find a simplicial cone $\mathcal{W}$ satisfying (3), the matrix $\mathbf{W}$ can be obtained by taking all the extreme rays of $\mathcal{W}$ as its columns. Such a cone search procedure is in essence equivalent to the asymmetric NMF problem itself, therefore it does not actually tell us how to check uniqueness. Notice that ordering and scaling information cannot be recovered, consistent with Definition 4. Donoho and Stodden's sufficient condition for unique NMF is the following.

Theorem 1 (Donoho et al. [24]): The NMF $\mathbf{S} = \mathbf{WH}$ is unique if the following conditions are satisfied.

- *Separability:* For each $k = 1, \ldots, K$ there exists $j \in \{1, \ldots, J\}$ such that

$$\begin{aligned} &\mathbf{H}_{k,j} \neq 0 \\ &\mathbf{H}_{k,l} = 0, \quad \forall\, l \neq j \end{aligned}$$

- *Generative Model:* The set $\{1, 2, \ldots, K\}$ is partitioned into A groups $\mathcal{P}_1, \ldots, \mathcal{P}_A$, each containing exactly P elements (therefore $K = AP$). For every $i = 1, \ldots, I$ and $a = 1, \ldots, A$, there exists an element $\mathbf{W}_{i,k}$ such that

$$\begin{aligned} &\mathbf{W}_{i,k} \neq 0, \quad k \in \mathcal{P}_a, \\ &\mathbf{W}_{i,l} = 0, \quad \forall l \in \mathcal{P}_a, l \neq k \end{aligned}$$

- *Complete Factorial Sampling:* For any $k_1 \in \mathcal{P}_1, k_2 \in \mathcal{P}_2, \cdots, k_A \in \mathcal{P}_A$, there exists $i \in \{1, \ldots, I\}$ such that

$$\mathbf{W}_{i,k_1} \neq 0, \mathbf{W}_{i,k_2} \neq 0, \cdots, \mathbf{W}_{i,k_A} \neq 0$$

B. Laurberg's Analysis

Laurberg *et al.* [25] offered a different viewpoint on the uniqueness of NMF, summarized in Lemma 3.

Lemma 3: If $\mathrm{rank}(\mathbf{S}) = K$, the NMF $\mathbf{S} = \mathbf{WH}$ is unique if and only if the non-negative orthant is the only simplicial cone $\mathcal{A}$ with K extreme rays that satisfies $\mathrm{cone}(\mathbf{W}^T) \subseteq \mathcal{A} \subseteq \mathrm{cone}(\mathbf{H})^*$.

Unlike Lemma 2, Lemma 3 is stated in terms of an existing solution of NMF. Furthermore, it works in the inner dimension $\mathbb{R}^K$ rather than the column space of $\mathbf{S}$ which is typically of much higher dimension. In this paper, most of the new uniqueness conditions will be proved using Lemma 3. This type if geometric interpretation can be traced back to [23], but Laurberg *et al.*'s statement is clearer to understand in this context. Laurberg *et al.* also gave a sufficient condition for uniqueness of NMF. We use the version Laurberg used in [35], which is easier to interpret.

Theorem 2 (Laurberg et al. [35]): The NMF $\mathbf{S} = \mathbf{WH}$ is unique if the following assumptions are satisfied.

- Sufficiently Spread: For each $k = 1, \ldots, K$ there is a $j \in \{1, \ldots, J\}$ such that

$$\begin{aligned} &\mathbf{H}_{k,j} \neq 0 \\ &\mathbf{H}_{k,i} = 0, \quad \forall\, i \neq j \end{aligned}$$

- Strongly Boundary Close: Matrix $\mathbf{W}$ satisfies the following conditions
  1) For each $k = 1, \ldots, K$ there is an $i \in \{1, \ldots, I\}$ such that

$$\begin{aligned} &\mathbf{W}_{i,k} = 0 \\ &\mathbf{W}_{i,j} \neq 0, \quad \forall\, j \neq i \end{aligned}$$

  2) There exists a permutation matrix $\mathbf{P}$ such that for all $i < K$ there exists a set $\{k_1, \ldots, k_{K-i}\}$ fulfilling: $(\mathbf{WP})_{i,k_j} = 0$ for all $j \leq K - i$; and the matrix

$$\begin{bmatrix} (\mathbf{WP})_{i+1,k_1} & \cdots & (\mathbf{WP})_{i+1,k_{K-i}} \\ \vdots & \ddots & \vdots \\ (\mathbf{WP})_{K,k_1} & \cdots & (\mathbf{WP})_{K,k_{K-i}} \end{bmatrix}$$

  is invertible.

Remark 1: Apparently, the reason Laurberg *et al.* started with a compact and elegant necessary and sufficient condition (Lemma 3), but then worked out a far stricter and seemingly awkward sufficient condition is that Lemma 3 is very hard to check, even though it assumes that a particular NMF is given. In essence, it asks the following question: given a polyhedral cone described as the intersection of half-spaces, and a finite number of points contained in that polyhedral cone, can we find a simplicial cone that is both a subset of the polyhedral cone and also a superset of those points? As Vavasis [8] has shown, this problem is NP-hard. The same applies to the necessary and sufficient condition in Lemma 2.

There are no results on the uniqueness of symmetric NMF so far, to the best of our knowledge. It is not obvious how to apply Donoho's approach to the symmetric case, since only the columns of the data matrix are considered, thus symmetry is completely ignored. Mimicking Laurberg's analysis, however, we can get a similar result to be used as a geometric criterion for the uniqueness of symmetric NMF.

Lemma 4: If $\mathrm{rank}(\mathbf{S}) = K$, the symmetric NMF $\mathbf{S} = \mathbf{WW}^T$ is unique if and only if the non-negative orthant is the only self-dual simplicial cone $\mathcal{A}$ with K extreme rays that satisfies $\mathrm{cone}(\mathbf{W}^T) \subseteq \mathcal{A} = \mathcal{A}^*$.

Proof: By Definition 5, if the symmetric NMF $\mathbf{S} = \mathbf{WW}^T$ is essentially unique, then for any unitary matrix $\mathbf{A}$, $\mathbf{WA} \geq 0$ implies that $\mathbf{A}$ is a permutation matrix. Now $\mathbf{WA} \geq 0$ implies $\mathrm{cone}(\mathbf{W}^T) \subseteq \mathrm{cone}(\mathbf{A})^* = \mathrm{cone}(\mathbf{A})$, and $\mathbf{A}$ being a permutation matrix means $\mathrm{cone}(\mathbf{A}) = \mathbb{R}_+^K$. Thus, this is simply a geometric way to describe Definition 5. ■

IV. New Results: Uniqueness of Symmetric and Asymmetric NMF

We are now ready to present our new conditions on the uniqueness of symmetric and asymmetric NMF. We start with a necessary condition.

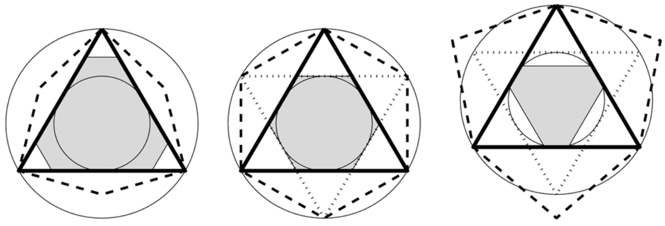

Fig. 2. A graphical view of Example 2 plotted in the same manner as Fig. 1. The triangle drawn with solid line is $\mathbb{R}^3_+$; the inner and outer circles are $\mathcal{C}$ and $\mathcal{C}^*$ respectively; the shaded area is $\mathrm{cone}(\mathbf{W}^T)$, and the polygon drawn with the dashed line is $\mathrm{cone}(\mathbf{W}^T)^*$. When $\omega = 0.25$, the condition stated in Theorem 4 is satisfied, so the symmetric NMF is unique; when $\omega = 0.75$, $\mathrm{cone}(\mathbf{W}^T) \not\supseteq \mathcal{C}$, so the symmetric NMF is not unique; when $\omega = 0.5$, although $\mathrm{cone}(\mathbf{W}^T) \supseteq \mathcal{C}$, all the extreme rays of $\mathrm{cone}(\mathbf{W}^T)^*$ lie on the boundary of $\mathcal{C}^*$. The dotted triangle in Fig. 2(b) and (c) shows another self-dual simplicial cone $\mathcal{A}$ satisfying $\mathrm{cone}(\mathbf{W}^T) \subseteq \mathcal{A}$. (a) $\omega = 0.25$. (b) $\omega = 0.5$. (c) $\omega = 0.75$.

Theorem 3 (Necessary Condition): Define

$$\mathcal{I}_k = \{i \in \{1,\ldots,I\} \,|\, \mathbf{W}_{i,k} \neq 0\}$$
$$\mathcal{J}_k = \{j \in \{1,\ldots,J\} \,|\, \mathbf{H}_{k,j} \neq 0\}$$

If the asymmetric NMF $\mathbf{S} = \mathbf{W}\mathbf{H}$ is unique, then there do not exist $k_1, k_2 \in \{1,\ldots,K\}, k_1 \neq k_2$ such that $\mathcal{I}_{k_1} \subseteq \mathcal{I}_{k_2}$, or $\mathcal{J}_{k_1} \subseteq \mathcal{J}_{k_2}$. The condition must also hold in the symmetric case, i.e., when $\mathbf{H} = \mathbf{W}^T$.

Proof: See Appendix B. In the case of asymmetric NMF the result can be found in [26] (the statement is not completely clear, but the essence is there). Gillis ([36], Remark 2) later presented the same result clearly. What is necessary for uniqueness of asymmetric NMF, however, is not automatically necessary for uniqueness of symmetric NMF. In Appendix B we prove that the result holds for symmetric NMF as well. ■

Corollary 1: If the asymmetric NMF $\mathbf{S} = \mathbf{W}\mathbf{H}$ or symmetric NMF $\mathbf{S} = \mathbf{W}\mathbf{W}^T$ is unique, then each column of $\mathbf{W}$ (and row of $\mathbf{H}$) contains at least one element that is equal to 0.

Proof: If the l-th column of $\mathbf{W}$ does not have 0 element, then clearly $\mathcal{I}_l = \{1,\ldots,I\}$, and $\mathcal{I}_k \subseteq \mathcal{I}_l, \forall\, k \neq l$, which violates the condition given in Theorem 3; and likewise for the rows of $\mathbf{H}$ in the asymmetric case. ■

Using Donoho and Stodden's analysis, the requirement that every column of $\mathbf{W}$ has a zero entry means that every extreme ray of $\mathrm{cone}(\mathbf{W})$ is on the boundary of $\mathcal{P}_\mathcal{V}$, and every row of $\mathbf{H}$ having a zero entry means there are columns of $\mathbf{S}$ on every facet of $\mathrm{cone}(\mathbf{W})$. This condition is intuitive, because otherwise we can always perturb $\mathrm{cone}(\mathbf{W})$ into a slightly bigger or smaller cone that still satisfies $\mathrm{cone}(\mathbf{S}) \subseteq \mathrm{cone}(\mathbf{W}) \subseteq \mathcal{P}_\mathcal{V}$, so that NMF will not be unique according to Lemma 2.

We need the second-order cone $\mathcal{C}$ defined in (1) from Example 1 to help derive our sufficient condition.

Theorem 4 (Sufficient Condition): Define the second-order cone

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^K | \mathbf{x}^T\mathbf{1} \geq \sqrt{K-1}\|\mathbf{x}\|_2\}$$

and let $\boldsymbol{bd}\mathcal{C}^* = \{\mathbf{x} \in \mathbb{R}^K | \mathbf{x}^T\mathbf{1} = \|\mathbf{x}\|_2\}$. If $\mathbf{W}$ satisfies that

1) $\mathrm{cone}(\mathbf{W}^T) \supseteq \mathcal{C}$
2) $\boldsymbol{bd}\mathcal{C}^* \cap \mathrm{cone}(\mathbf{W}^T)^* = \{\lambda\mathbf{e}_k | \lambda \geq 0, k = 1,\ldots,K\}$

and the same for $\mathbf{H}^T$, then the asymmetric NMF $\mathbf{S} = \mathbf{W}\mathbf{H}$ is unique. This condition is also sufficient in the symmetric case, i.e., when $\mathbf{H} = \mathbf{W}^T$.

Proof: See Appendix C. ■

The interpretation of the second part of the sufficient condition is the following. First of all, according to Property 3 of the dual cones, $\mathrm{cone}(\mathbf{W}^T)^* \subseteq \mathcal{C}^*$ is equivalent to $\mathrm{cone}(\mathbf{W}^T) \supseteq \mathcal{C}$, which means all the extreme rays of $\mathrm{cone}(\mathbf{W}^T)^*$ are contained in $\mathcal{C}^*$. Per the proof of Theorem 4, $\mathbf{e}_k$'s are the extreme rays of $\mathrm{cone}(\mathbf{W}^T)^*$, and they lie on the boundary of $\mathcal{C}^*$ too. This statement requires that all the other extreme rays of $\mathrm{cone}(\mathbf{W}^T)^*$ lie in the interior of $\mathcal{C}^*$. An example shows how to use the sufficient condition, and the importance of the second requirement in order to achieve sufficiency.

Example 2 [25]: Consider the symmetric NMF $\mathbf{S} = \mathbf{W}\mathbf{W}^T$ where

$$\mathbf{W} = \begin{bmatrix} \omega & 1 & 1 & \omega & 0 & 0 \\ 1 & \omega & 0 & 0 & \omega & 1 \\ 0 & 0 & \omega & 1 & 1 & \omega \end{bmatrix}^T$$

For $0 \leq \omega \leq 1$, symmetric NMF is unique if and only if $\omega < 0.5$. This is because if $\omega < 0.5$, the proposed sufficient condition is satisfied, therefore the symmetric NMF is unique. If $\omega \geq 0.5$, define

$$\mathbf{A} = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}.$$

One can check that $\mathbf{A}$ is unitary and $\mathbf{W}\mathbf{A} \geq 0$ in this case, hence symmetric NMF is not unique. Fig. 2 illustrates the relationship between $\mathrm{cone}(\mathbf{W}^T)$ and $\mathcal{C}$ for selected values of ω. Notice that when $\omega = 0.5$, as is shown in Fig. 2(b), although $\mathrm{cone}(\mathbf{W}^T) \supseteq \mathcal{C}$ is satisfied, there are other extreme rays of $\mathrm{cone}(\mathbf{W}^T)^*$ that lie on the boundary of $\mathcal{C}^*$ and are orthogonal to each other. Thus, it is still possible in this case to form a suitable simplicial cone which satisfies the requirement given in Lemma 3.

Laurberg *et al.* [25] first gave this example and pointed out that uniqueness depends on the value of ω in this case. However, the sufficient condition for uniqueness given in [25] fails to demonstrate when uniqueness holds in this case, except for $\omega = 0$; whereas our new sufficient condition in Theorem 4 is able to identify the full interval where uniqueness holds.

Remark 2: Example 2 illustrates the usage of Theorem 4 to check the uniqueness of NMF in low-dimensional cases. Unfortunately, as the dimension increases, it becomes very hard to check the sufficient condition, since checking whether $\mathrm{cone}(\mathbf{W}^T) \supseteq \mathcal{C}$ is true is NP-complete. To see the NP-completeness of this problem, we can first intersect both $\mathrm{cone}(\mathbf{W}^T)$ and $\mathcal{C}$ with the hyperplane $\mathbf{1}^T\mathbf{x} = 1$. Then it becomes checking whether a ball is a subset of a polytope described as the convex hull of points. Freund and Orlin [37] considered several set containment problems and proved that the aforementioned one is NP-complete.

From a computational complexity point of view, we started from a criterion (Lemma 3) that is NP-hard to check, and reached a sufficient condition that is NP-complete to check, which does not seem like much progress. However, Theorem 4 treats $\mathbf{W}$ and $\mathbf{H}$ in a balanced fashion, unlike other sufficient conditions which are strict on one matrix factor, lenient on the other. Furthermore, Theorem 4 gives implications about how the latent factors look like if they indeed satisfy its sufficient conditions, as shown next.

Corollary 2: If the condition given in Theorem 4 is satisfied for the asymmetric NMF $\mathbf{S} = \mathbf{W}\mathbf{H}$ or the symmetric NMF $\mathbf{S} = \mathbf{W}\mathbf{W}^T$, then for all $k = 1, 2, \ldots, K$, $\mathbf{e}_k$ is an extreme ray of $\operatorname{cone}(\mathbf{W}^T)^*(\operatorname{cone}(\mathbf{H})^*)$. Thus, each column of $\mathbf{W}$ (row of $\mathbf{H}$) contains at least $K-1$ zero entries.

Proof: We have argued in the proof of Theorem 4 (cf. Appendix C) that if the sufficient condition is satisfied, the $\mathbf{e}_k$'s are extreme rays of $\operatorname{cone}(\mathbf{W}^T)^*$ (and $\operatorname{cone}(\mathbf{H})^*$ in the asymmetric case). An extreme ray is the intersection of at least $K-1$ independent facets of a polyhedral cone with dimension K. Therefore for $\operatorname{cone}(\mathbf{W}^T)^*$, we have

$$\mathbf{W}\mathbf{e}_k \geq 0$$

and that at least $K-1$ of them are satisfied as equalities. Therefore, each column of $\mathbf{W}$ (row of $\mathbf{H}$) contains at least $K-1$ zero entries. ■

Corollary 2 builds a link between uniqueness and latent sparsity of NMF. It is observed in practice that if the true latent factors are sparse, NMF usually tends to recover the correct solution, up to scaling and permutation. However, counter-examples do exist, e.g., Example 2 when $0.5 \leq \omega < 1$. Now we understand that the reason sparse latent factors usually lead to unique NMF is because if the latent factors are sparse, it is more *likely* that the sufficient condition given in Theorem 4 is satisfied. Using Donoho's analysis, we also have the following result if only the latent sparsity condition is met.

Proposition 1: If for any $k = 1, 2, \ldots, K$, $\mathbf{e}_k$ is an extreme ray of both $\operatorname{cone}(\mathbf{W}^T)^*$ and $\operatorname{cone}(\mathbf{H})^*$, then there does not exist another simplicial cone $\mathcal{W}$ with K extreme rays that satisfies either $\operatorname{cone}(\mathbf{W}) \subset \mathcal{W} \subseteq \mathcal{P}_{\mathcal{V}}$ or $\operatorname{cone}(\mathbf{S}) \subseteq \mathcal{W} \subset \operatorname{cone}(\mathbf{W})$.

Proof: See Appendix D. ■

The result in Proposition 1 is not a strong statement, since set containment is not a linear ordering. However, it does rule out a lot of possibilities of finding another NMF, especially within a local neighborhood of the current solution. We examine the ability to reconstruct the sparse latent factors by NMF, under the constraint that each column of $\mathbf{W}$ (row of $\mathbf{H}$) contains at least $K-1$ zeros, both in the symmetric and asymmetric case, in the following example.

Example 3: In this example we randomly generate a 200×30 non-negative matrix $\hat{\mathbf{W}}$, with a certain proportion of randomly selected entries set to zero, and the non-zero entries drawn from an i.i.d. exponential distribution. The columns of $\hat{\mathbf{W}}$ are ordered so that

$$\sum_{i=1}^{I} \hat{\mathbf{W}}_{i,1} > \sum_{i=1}^{I} \hat{\mathbf{W}}_{i,2} > \cdots > \sum_{i=1}^{I} \hat{\mathbf{W}}_{i,K}$$

Then we form the low rank complete positive matrix $\mathbf{S} = \hat{\mathbf{W}}\hat{\mathbf{W}}^T$. Symmetric NMF is applied to $\mathbf{S} = \mathbf{W}\mathbf{W}^T$. The columns of $\mathbf{W}$ are then ordered analogously, to fix the permutation ambiguity. For density (the proportion of non-zero entries in $\hat{\mathbf{W}}$) varying from 0.5 to 0.8, in which case the matrix $\hat{\mathbf{W}}$ we randomly generated satisfies that $\mathbf{e}_k$'s are extreme rays of $\operatorname{cone}(\hat{\mathbf{W}}^T)^*$ with high probability, this procedure is repeated 100 times, and the maximum reconstruction error $\|\hat{\mathbf{W}} - \mathbf{W}\|_F$ is given in Table I.

TABLE I
MAXIMUM RECONSTRUCTION ERROR FOR SYMMETRIC NMF

density	max $\|\hat{\mathbf{W}} - \mathbf{W}\|_F$
0.5	2.57×10^{-13}
0.6	4.20×10^{-13}
0.7	6.42×10^{-13}
0.8	3.29×10^{-12}

TABLE II
MAXIMUM RECONSTRUCTION ERROR FOR ASYMMETRIC NMF

density	max $\|\hat{\mathbf{W}} - \mathbf{W}\|_F$	max $\|\hat{\mathbf{H}} - \mathbf{H}\|_F$
0.5	0.0070×10^{-7}	0.0083×10^{-5}
0.6	0.0030×10^{-7}	0.0067×10^{-5}
0.7	0.0184×10^{-7}	0.0302×10^{-5}
0.8	0.1154×10^{-7}	0.1991×10^{-5}

This indicates that over the 400 Monte-Carlo tests we tried, symmetric NMF successfully recovered the latent factors in each and every case. The algorithm we used for symmetric NMF is the one proposed later in this paper, see Section V.

Example 4: In this example we randomly generate a 200×30 non-negative matrix $\hat{\mathbf{W}}$ and a 30×250 non-negative matrix $\hat{\mathbf{H}}$, with a certain proportion of randomly selected entries set to zero, and the non-zero entries drawn from an i.i.d. exponential distribution. The columns of $\hat{\mathbf{W}}$ are scaled to sum up to 1

$$\sum_{i=1}^{I} \hat{\mathbf{W}}_{i,1} = \sum_{i=1}^{I} \hat{\mathbf{W}}_{i,2} = \cdots = \sum_{i=1}^{I} \hat{\mathbf{W}}_{i,K} = 1$$

and the rows of $\hat{\mathbf{H}}$ are ordered such that

$$\sum_{j=1}^{J} \hat{\mathbf{H}}_{1,j} > \sum_{j=1}^{J} \hat{\mathbf{H}}_{2,j} > \cdots > \sum_{j=1}^{J} \hat{\mathbf{H}}_{K,j}$$

Then we form the low rank non-negative matrix $\mathbf{S} = \hat{\mathbf{W}}\hat{\mathbf{H}}$. Asymmetric NMF is applied to $\mathbf{S} = \mathbf{W}\mathbf{H}$. The columns of $\mathbf{W}$ are scaled to sum up to 1, with the scaling absorbed into the rows of $\mathbf{H}$, in order to avoid the scaling ambiguity; and the rows of $\mathbf{H}$ are then re-ordered, with the same re-ordering applied to the columns of $\mathbf{W}$, to fix the permutation ambiguity. For density varying from 0.5 to 0.8, in which case the matrices $\hat{\mathbf{W}}$ and $\hat{\mathbf{H}}$ we randomly generated satisfy that $\mathbf{e}_k$'s are extreme rays of $\operatorname{cone}(\hat{\mathbf{W}}^T)^*$ and $\operatorname{cone}(\hat{\mathbf{H}})^*$ with high probability, this procedure is repeated 100 times, and the maximum reconstruction errors for $\hat{\mathbf{W}}$ and $\hat{\mathbf{H}}$ are given in Table II.

Notice that the zero elements are randomly located, therefore neither Donoho's separability assumption [24] nor Laurberg's sufficiently spread assumption [25] are satisfied. However, as can be observed, in all cases the asymmetric NMF is able to reconstruct the true latent factors. The algorithm used here was Fast-HALS cf. [13, Algorithm 2].

The results given in the above examples are reassuring to a certain degree, in light of the fact that the true sufficient condition is NP-complete to check. They also showcase how strict the previously suggested sufficient conditions [24], [25] are, since

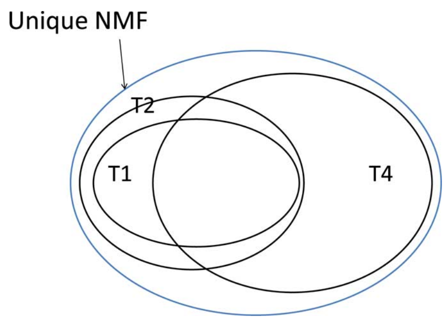

Fig. 3. The set of non-negative matrix pairs $\{\mathbf{W}, \mathbf{H}\}$ that: **Unique NMF**: gives an essentially unique NMF **T1**: satisfies the sufficient condition given in Theorem 1. **T2**: satisfies the sufficient condition given in Theorem 2. **T4**: satisfies the sufficient condition given in Theorem 4.

in those examples the columns of $\mathbf{H}$ are highly unlikely to contain all scaled versions of $\mathbf{e}_k$'s.

Fig. 3 shows graphically the relationship between some of the uniqueness results. One thing to notice is that both Donoho *et al.* and Laurberg *et al.* gave asymmetric conditions—their condition on $\mathbf{H}$ is much stricter than their condition on $\mathbf{W}$, while our new result poses the same conditions on both $\mathbf{W}$ and $\mathbf{H}$. As a result, their weaker conditions on $\mathbf{W}$ are loose enough to preclude containment of Theorem 1, Theorem 2 in Theorem 4.

V. Symmetric NMF: Algorithm

A. Formulation

Suppose there exists an exact symmetric NMF of $\mathbf{S}$ with $K = \mathrm{rank}(\mathbf{S})$ components. Then $\mathbf{S}$ is symmetric positive semi-definite; consider its reduced eigen-decomposition

$$\mathbf{S} = \mathbf{U}_s \mathbf{\Lambda}_s \mathbf{U}_s^T$$

where $\mathbf{U}_s$ is $I \times K$ orthogonal and $\mathbf{\Lambda}_s$ is $K \times K$ diagonal. Define

$$\mathbf{B} = \mathbf{U}_s \mathbf{\Lambda}_s^{1/2}$$

Since

$$\mathbf{S} = \mathbf{B}\mathbf{B}^T = \mathbf{W}\mathbf{W}^T$$

where both $\mathbf{B}$ and $\mathbf{W}$ are $I \times K$, there exists a unitary matrix $\mathbf{Q}$ such that

$$\mathbf{B}\mathbf{Q} = \mathbf{W}$$

Therefore, after obtaining $\mathbf{B}$ via eigen-analysis, we can formulate the recovery of $\mathbf{W}$ as follows:

$$\min_{\mathbf{W},\mathbf{Q}} \quad \|\mathbf{W} - \mathbf{B}\mathbf{Q}\|_F^2 \tag{5a}$$

$$\text{subject to} \quad \mathbf{W} \geq 0, \mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I} \tag{5b}$$

B. Method

The constraint $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ is not convex with respect to $\mathbf{Q}$, suggesting that (5) is a hard problem. We propose updating

$\mathbf{S} = \mathbf{U}_s \mathbf{\Lambda}_s \mathbf{U}_s^T$ ▷ Reduced EVD
$\mathbf{B} \leftarrow \mathbf{U}_s \mathbf{\Lambda}_s^{1/2}$, $\mathbf{Q} \leftarrow \mathbf{I}$
repeat
  $\mathbf{W} \leftarrow \max(\mathbf{0}, \mathbf{B}\mathbf{Q})$
  $\mathbf{W}^T\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ ▷ SVD
  $\mathbf{Q} \leftarrow \mathbf{V}\mathbf{U}^T$
until $|\mathrm{trace}(\mathbf{W}^T(\mathbf{W} - \mathbf{B}\mathbf{Q}))| <$ tolerance

Fig. 4. Proposed algorithm for symmetric NMF.

$\mathbf{W}$ and $\mathbf{Q}$ in an alternating fashion. The updating rule for $\mathbf{W}$ is extremely simple: since $\mathbf{W}$ is non-negative, we simply set

$$\mathbf{W} \leftarrow \max(0, \mathbf{B}\mathbf{Q}) \tag{6}$$

When updating $\mathbf{Q}$, the solution is given by the Procrustes projection [38], i.e.,

$$\mathbf{Q} \leftarrow \mathbf{V}\mathbf{U}^T \tag{7}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices given by the singular value decomposition of $\mathbf{W}^T\mathbf{B}$

$$\mathbf{W}^T\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{8}$$

This simple algorithm is summarized in Fig. 4.

Proposition 2: The value of the objective function in (5a) is monotonically non-increasing during iterations of the algorithm given in Fig. 4, since each update step is conditionally optimal for $\mathbf{W}$ given $\mathbf{Q}$ or vice-versa. Furthermore, with a (possibly very loose) upper bound on the elements of $\mathbf{W}$, every limit point generated by the algorithm shown in Fig. 4 is a stationary point of problem (5).

Proof: The set of unitary matrices is not convex, hence well-known results on the limit point behavior of block coordinate descent such as [39], [40] do not apply. In [41, Theorem 3.1], however, it is shown that every limit point of block coordinate descent is a stationary point if

1) The constraint set of each block is compact; and
2) Conditional updates are optimal, *and* the block that yields the best improvement is chosen to be updated in each iteration.

In our case, the constraint set for $\mathbf{Q}$ is the set of real unitary matrices (i.e., the orthogonal group), which is compact [42, pp. 4]. The constraint set of $\mathbf{W}$ is the non-negative orthant, which is not bounded; but we can add a (very loose) bound on the elements of $\mathbf{W}$ to make it compact. Moreover, since we only use two blocks, after updating one block, in the next step the other block will give us a better (or at least no worse) improvement. In other words, the update strategy proposed in [41] always yields an alternating update in the case of only two blocks. ■

In terms of per-iteration complexity, the matrix multiplications $\mathbf{B}\mathbf{Q}$ and $\mathbf{W}^T\mathbf{B}$ both require $O(IK^2)$ flops, whereas the SVD performed on the relatively small-sized $K \times K$ matrix $\mathbf{W}^T\mathbf{B}$ requires $O(K^3)$ flops [43, pp. 254]. If we assume $I \gg K$, which is typically the case in practice, then the $O(IK^2)$ term dominates, which results in a $O(IK^2)$ per-iteration complexity.

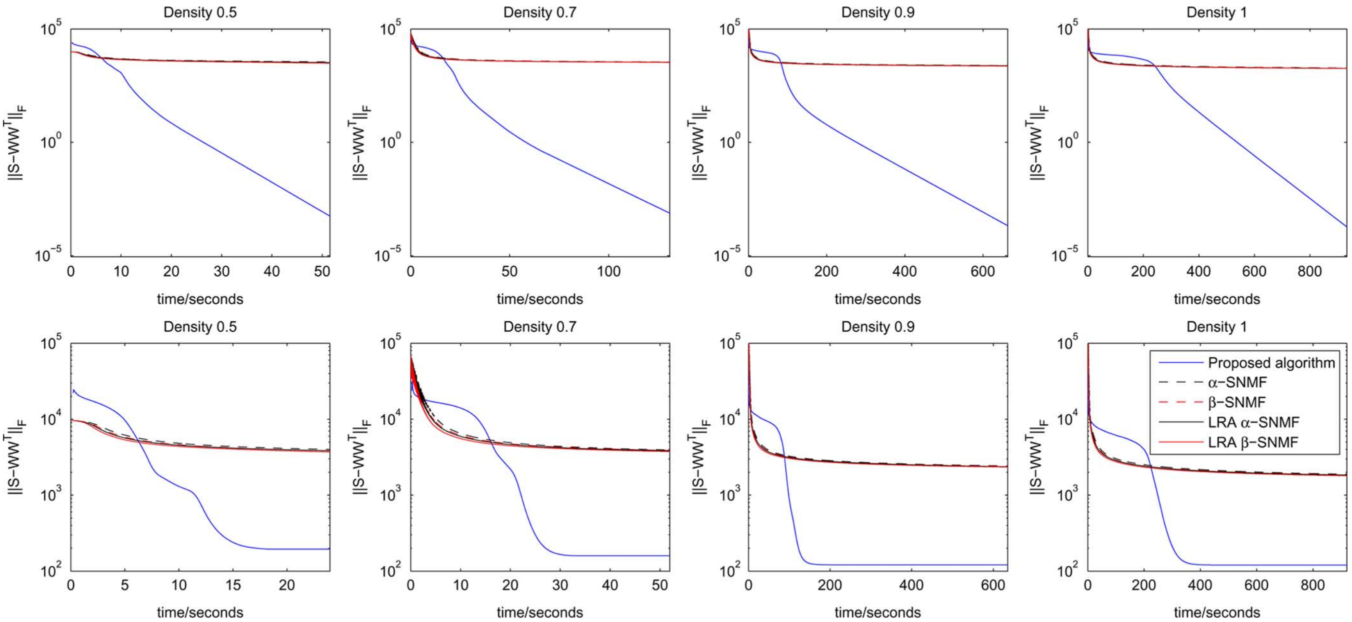

Fig. 5. Convergence of the proposed algorithm vs. α-SNMF and β-SNMF [29] with $\alpha = \beta = 0.99$ and their modified versions [45] employing low rank approximation (LRA): noiseless (top row) and noisy (bottom row). x-axis counts elapsed time.

The computation of K dominant eigenvalues and eigenvectors in the very first step entails complexity $O(I^2K)$, but considering the fact that this is done only once, its $O(I^2K)$ cost is amortized over many iterations of symmetric NMF.

The Karush-Kuhn-Tucker conditions [44] for problem (5) are

$$2\mathbf{W}_* - 2\mathbf{B}\mathbf{Q}_* - \mathbf{\Omega}_* = 0 \tag{9a}$$
$$-2\mathbf{W}_*^T\mathbf{B} + 2\mathbf{Q}_*\mathbf{\Theta}_*^T = 0 \tag{9b}$$
$$\mathbf{W}_* \geq 0 \tag{9c}$$
$$\mathbf{\Omega}_* \geq 0 \tag{9d}$$
$$\mathbf{W}_* \circ \mathbf{\Omega}_* = 0 \tag{9e}$$
$$\mathbf{Q}_*^T\mathbf{Q}_* = \mathbf{Q}_*\mathbf{Q}_*^T = \mathbf{I} \tag{9f}$$

where $\mathbf{\Omega}_*$ and $\mathbf{\Theta}_*$ are Lagrange multipliers for the constraints $\mathbf{W}$ being element-wise non-negative and $\mathbf{Q}$ unitary, respectively, and "$\circ$" denotes the Hadamard product, i.e., element-wise matrix multiplication. The iterates in Fig. 4 always satisfy (9c) and (9f) by construction. Since (9b) is the only condition that involves $\mathbf{\Theta}_*$, we may simply solve (9b) to obtain a suitable $\mathbf{\Theta}_*^T = \mathbf{Q}_*^T\mathbf{W}_*^T\mathbf{B}$ for any given $\mathbf{Q}_*$, $\mathbf{W}_*$. For the remaining KKT conditions, we first eliminate $\mathbf{\Omega}_*$ and rewrite them as

$$\mathbf{W}_* - \mathbf{B}\mathbf{Q}_* \geq 0 \tag{10a}$$
$$\mathbf{W}_* \circ (\mathbf{W}_* - \mathbf{B}\mathbf{Q}_*) = 0 \tag{10b}$$

The left hand side of (10b) can be used as the termination criterion. One must notice that this quantity must be checked after the $\mathbf{Q}$-update, since it is automatically satisfied after the $\mathbf{W}$-update. Moreover, since $\mathrm{tr}(\mathbf{W}^T\mathbf{\Omega}) = \sum_{i,j}\mathbf{W}_{i,j}\mathbf{\Omega}_{i,j}$, we can also check $\mathrm{tr}(\mathbf{W}^T(\mathbf{W} - \mathbf{B}\mathbf{Q}))$ instead. Notice that $\mathrm{tr}(\mathbf{W}_*^T(\mathbf{W}_* - \mathbf{B}\mathbf{Q}_*)) = 0$ happens to be the orthogonality condition of the least squares problem, cf. the cost function in (5). Checking $\mathrm{tr}(\mathbf{W}^T(\mathbf{W} - \mathbf{B}\mathbf{Q}))$ is preferable to checking successive differences of the cost in (5), because it avoids early termination during *swamps*—intervals during which the progress in terms of the cost function is slow.

In practice, we may encounter cases where $K \neq \mathrm{rank}(\mathbf{S})$. For $K < \mathrm{rank}(\mathbf{S})$, we are trying to find a good non-negative low rank approximation of $\mathbf{S}$, and we can simply take the first K dominant eigen-components, then apply the same updating rules afterwards. For $K > \mathrm{rank}(\mathbf{S})$, we need to modify the basic algorithm; the modified version can be found in Appendix E.

C. Simulation

1) Synthetic Data: The matrix $\mathbf{S}$ is generated by taking $\mathbf{S} = \hat{\mathbf{W}}\hat{\mathbf{W}}^T + \mathbf{N}$, where $\hat{\mathbf{W}}$ is a non-negative matrix with certain amount of zeros and the non-zero entries drawn from an i.i.d. exponential distribution, and the elements of $\mathbf{N}$ first drawn from an i.i.d. Gaussian distribution, and then symmetrized by taking $\mathbf{N} \leftarrow \mathbf{N} + \mathbf{N}^T$. We take the size of $\hat{\mathbf{W}}$ to be 1000×150. The tolerance we set to terminate the algorithm in Fig. 4 is 10^{-5}, and we let α-SNMF and β-SNMF to run the same amount of time to compare their performances.

The convergence of a single run of our proposed algorithm under various conditions is illustrated in Fig. 5 (in terms of time used) and Fig. 6 (in terms of number of iterations used), comparing to α-SNMF and β-SNMF provided in [29] with $\alpha = \beta = 0.99$, since their experiments showed (and we verified) that this value gives faster convergence, and the low-rank approximation (LRA) version of these algorithms using the strategy provided in [45], on the same $\mathbf{S}$. The cost employed in both α-SNMF and β-SNMF is $\|\mathbf{S} - \mathbf{W}\mathbf{W}^T\|_F^2$, which is different from (5a), but we compare all of them using $\|\mathbf{S} - \mathbf{W}\mathbf{W}^T\|_F$ on the y-axis as common basis. Since our proposed algorithm uses eigen-decomposition of the data matrix as a pre-processing step, we include the time it takes to compute this eigen-decomposition in the timing reported on the x-axis in Fig. 5, for fair comparison.

In Fig. 5, we show the convergence when $\mathbf{S}$ is noiseless ($\mathbf{N} = 0$) in the top row, and with small noise (the entries of $\mathbf{N}$ are first

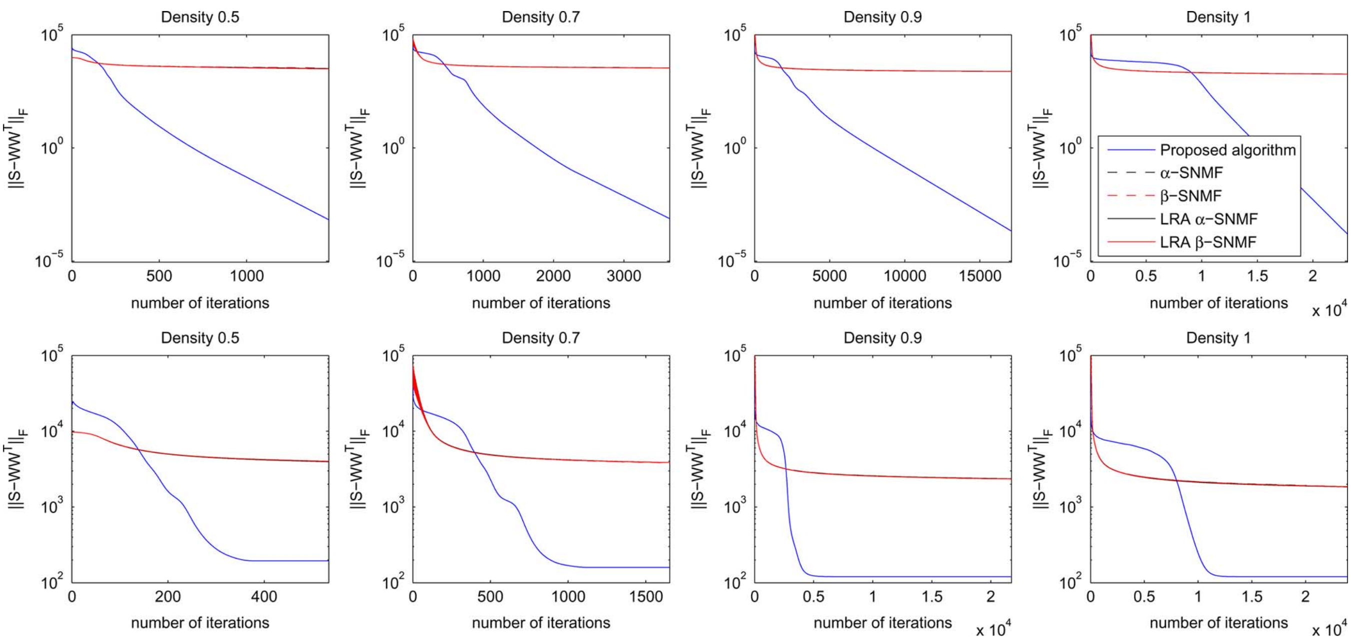

Fig. 6. Convergence of the proposed algorithm vs. α-SNMF and β-SNMF [29] with $\alpha = \beta = 0.99$ and their modified versions [45] employing low rank approximation (LRA): noiseless (top row) and noisy (bottom row). x-axis counts number of iterations.

drawn from an i.i.d. Gaussian distribution with standard deviation $\sigma = 10^{-1}$ and then symmetrized) in the bottom row; and the densities (proportion of non-zero entries) of the true latent factor $\hat{\mathbf{W}}$ are (from left to right) 0.5, 0.7, 0.9, 1. In the noiseless case, our proposed algorithm tends to converge to an exact decomposition, whereas none of the SNMF variants is able to give a good approximation within that amount of time, although at the beginning they reduce the cost function faster. When small noise is added, the proposed algorithm shows good robustness, and again out-performs the two SNMF algorithms after some point. Notice that given the noise power, the symmetrization strategy, and the size of the matrix, the value of $\|\mathbf{N}\|_F$ is approximately 150, and our proposed algorithm is able to reach that error bound. An interesting observation is that the rate of convergence is somehow related to the sparsity of the true latent factor—the smaller the density, the faster the algorithm converges. Furthermore, overall the convergence rate looks linear, but swamps are sometimes encountered, which is why our proposed termination criterion is preferable than checking successive differences of the cost function. Our algorithm clearly outperforms all of the SNMF variants in this case.

It is important to note that after the computation of the K dominant eigen-components in the first step of our algorithm, each iteration (update cycle) entails complexity $O(IK^2)$, whereas one iteration of either α-SNMF or β-SNMF entails complexity $O(I^2K)$ (note that $K \leq I$). Therefore our algorithm also has an edge in terms of scaling up for big data, provided we have a scalable way to compute dominant eigenvalues and eigenvectors. Reduced-complexity variants of α-SNMF and β-SNMF have been very recently proposed in [45], employing low-rank approximation (LRA) preprocessing to reduce the per-iteration complexity to $O(IK^2)$. Such a comparison of per-iteration counts is incomplete, however, as it does not take into account the number of iterations till convergence. Fig. 6 shows that the number of iterations is much smaller and the average convergence rate much faster for the proposed algorithm relative to the SNMF variants, in all cases considered. Note that in Fig. 6 the x-axis counts the number of iterations instead of total elapsed time as in Fig. 5.

2) Real Data: We applied the new algorithm given in Fig. 4 to a real-life dataset containing co-authorship data from the U.S. Army Research Laboratory Collaborative Technology Alliance (ARL-CTA) on Communications and Networks (C&N), a large-scale research project that involved multiple academic and industry research groups, led by Telcordia. The ARL C&N CTA was an 8-year program, and produced numerous publications, involving over 500 individuals. A. Swami and N. Sidiropoulos were both involved as researchers and authors in this project, and A. Swami had significant oversight on much of the research—they know the 'social dynamics' and history of the consortium, and can interpret/sanity check the results of automated social network analysis of this dataset. The particular data analyzed here is a 518×518 symmetric non-negative matrix $\mathbf{A}$, where $\mathbf{A}_{i,j}$ is the number of papers co-authored by author-i and author-j ($\mathbf{A}_{i,i}$ is the number of papers written by author-i). The task is to cluster the authors, based only on $\mathbf{A}$. Ding *et al.* [31] have shown that k-means clustering can be approximated by symmetric NMF of the pair-wise similarity matrix $\mathbf{S} = \mathbf{X}^T\mathbf{X} = \mathbf{W}\mathbf{W}^T$, where the columns of $\mathbf{X}$ represent the data points that we want to cluster, and the number of columns of $\mathbf{W}$, K, is the number of clusters. The cluster that $\mathbf{X}_{:,i}$ belongs to is determined by taking $\arg\max_k \mathbf{W}_{i,k}$. In our case, we do not have access to $\mathbf{X}$, but we may interpret $\mathbf{A}$ as the pair-wise similarity matrix $\mathbf{S} = \mathbf{X}^T\mathbf{X}$, to be decomposed as $\mathbf{S} = \mathbf{W}\mathbf{W}^T$, with $\mathbf{W} \geq 0$.

We run symmetric NMF on $\mathbf{A}$ for $K = 3, 10$. The weight of cluster k is measured by $\|\mathbf{W}_{:,k}\|_2$, and the weight of author i in the cluster k is measured by $\mathbf{W}_{i,k}$. Table III lists the top-10

TABLE III
TOP-10 CONTRIBUTORS OF THE TOP-3 CLUSTERS

$K=3$	cluster 1	cluster 2	cluster 3
	Georgios B. Giannakis	Lang Tong	Mariusz A. Fecko
	Shengli Zhou	Ananthram Swami	Sunil Samtani
	Xiaoli Ma	Qing Zhao	M. Umit Uyar
	Pengfei Xia	Brian M. Sadler	Ibrahim Hokelek
	Xiaodong Cai	Yunxia Chen	Jianping Zou
	Tairan Wang	Min Dong	Jianliang Zheng
	Qingwen Liu	Youngchul Sung	Myung Jong Lee
	Xing Wang	Ting He	Tarek N. Saadawi
	Zhengdao Wang	P. Venkitasubramaniam	Ulas C. Kozat
	Alfonso Cano	Zhengyuan Xu	Phillip T. Conrad
$K=10$	cluster 1	cluster 2	cluster 3
	Georgios B. Giannakis	Lang Tong	Mariusz A. Fecko
	Shengli Zhou	Ananthram Swami	Sunil Samtani
	Pengfei Xia	Brian M. Sadler	M. Umit Uyar
	Xiaodong Cai	Min Dong	Ibrahim Hokelek
	Qingwen Liu	Ting He	Jianping Zou
	Tairan Wang	Youngchul Sung	Jianliang Zheng
	Xing Wang	P. Venkitasubramaniam	Ulas C. Kozat
	Zhengdao Wang	Srihari Adireddy	Phillip T. Conrad
	Yingqun Yu	Gokhan Mergen	Ahmed Abdelal
	Alfonso Cano	Animashree Anandkumar	J. Sucec

contributors of the top-3 clusters, for $K = 3$ (top) and $K = 10$ (bottom). The results are very reasonable. The first cluster is Georgios Giannakis' group at the University of Minnesota, the participant who contributed most publications to the project. The second cluster is more interesting: it comprises Lang Tong's group at Cornell, but also close collaborators from ARL (Brian Sadler, Ananthram Swami) who co-authored many papers with Cornell researchers and alumni over the years. The third cluster is even more interesting, and would have been harder to decipher for someone without direct knowledge of the project. It consists of Telcordia researchers (Telcordia was the lead of the project), but it also contains researchers from the City University of New York (CUNY), and, to a lesser extent, the University of Delaware (UDEL), suggesting that geographic proximity may have a role. Interestingly, the network of collaborations between Telcordia, CUNY, and UDEL dates back to the FEDLAB project (which was in a sense the predecessor of the CTA), and continued through much of the CTA as well. Notice that the three clusters remain stable even when $K = 10 > 3$ is used, although NMF is not guaranteed to be nested (for even higher K, e.g., $K = 30$, this stability breaks down, as larger clusters are broken down into more tightly woven pieces).

VI. CONCLUSION

We have revisited NMF from a geometric point of view, paying particular attention to uniqueness and algorithmic issues. NMF has found numerous applications in diverse areas, and its success stems in good measure by its ability to unravel the true latent factors in certain cases—which makes our limited understanding of when uniqueness holds particularly annoying. Symmetric NMF is element-wise non-negative square-root factorization of positive semidefinite matrices, and it too has many applications—not least as an approximation to the NP-hard k-means problem. We provided new uniqueness conditions that help shed light into the matter, and explained why sparse latent factors usually lead to unique NMF in practice, although checking a key condition that we derived was also shown to be NP-complete. Beyond uniqueness, a new algorithm for symmetric NMF was proposed, using Procrustes rotations. This was shown to be a useful addition to the existing NMF toolbox. We also applied our new symmetric NMF algorithm to a clustering problem for co-authorship data from the ARL C&N CTA, and we obtained meaningful and nicely interpretable results.

APPENDIX A
PROOF OF LEMMA 1

Assume $\mathrm{cone}(\mathbf{A})$ satisfies that

$$\mathcal{C} \subseteq \mathrm{cone}(\mathbf{A}) \subseteq \mathcal{C}^*$$

According to Property 2 and 3 of dual cone, this is equivalent to

$$\mathrm{cone}(\mathbf{A}) \subseteq \mathcal{C}^*, \mathrm{cone}(\mathbf{A}^{-T}) \subseteq \mathcal{C}^*$$

$\mathrm{cone}(\mathbf{A}) \subseteq \mathcal{C}^*$ means that every column of $\mathbf{A}$ is in $\mathcal{C}^*$, therefore

$$\mathbf{1}^T\mathbf{A}\mathbf{e}_i \geq \|\mathbf{A}\mathbf{e}_i\|_2, \quad i = 1, \ldots, n \tag{11}$$

Let $\mathbf{B} = \mathbf{A}^{-T}$, i.e., $\mathbf{A}^T\mathbf{B} = \mathbf{I}$. Similarly, for $\mathrm{cone}(\mathbf{B}) \subseteq \mathcal{C}^*$, we have

$$\mathbf{1}^T\mathbf{B}\mathbf{e}_i \geq \|\mathbf{B}\mathbf{e}_i\|_2, \quad i = 1, \ldots, n \tag{12}$$

Both (11) and (12) involve only non-negative numbers, so we can take their product and sum over all i's to get

$$\mathbf{1}^T\mathbf{A}\mathbf{B}^T\mathbf{1} \geq \sum_{i=1}^{n} \|\mathbf{A}\mathbf{e}_i\|_2\|\mathbf{B}\mathbf{e}_i\|_2 \tag{13}$$

The left hand side of (13) equals n, since $\mathbf{A}\mathbf{B}^{\mathbf{T}} = \mathbf{I}$. Using the Cauchy-Schwarz inequality, the right hand side of (13) is

$$\sum_{i=1}^{n} \|\mathbf{A}\mathbf{e}_i\|_2\|\mathbf{B}\mathbf{e}_i\|_2 \geq \sum_{i=1}^{n} (\mathbf{B}\mathbf{e}_i)^T\mathbf{A}\mathbf{e}_i \tag{14}$$

The right hand side equals to n too, again thanks to $\mathbf{A}^T\mathbf{B} = \mathbf{I}$. Therefore, all the inequalities (11)–(14) are equalities. Notice that (14) is satisfied as an equality if and only if $\mathbf{B}\mathbf{e}_i$ is a positively scaled version of $\mathbf{A}\mathbf{e}_i$, for all $i = 1, \ldots, n$. Since we assume $\|\mathbf{A}\mathbf{e}_i\|_2 = 1$ for all i's, then $\mathbf{A} = \mathbf{B}$, i.e., $\mathbf{A} = \mathbf{A}^{-T}$. Therefore, the columns of $\mathbf{A}$ are orthogonal to each other. Furthermore, if (13) is satisfied as equality, it implies that (11) and (12) are also equalities. In other words, the extreme rays of $\mathbf{A}$ lie on the boundary of $\mathcal{C}^*$, i.e., $\mathbf{A}^T\mathbf{1} = \mathbf{1}$.

APPENDIX B
PROOF OF THEOREM 3

A. Asymmetric Case

For the case of asymmetric NMF the essence of the result can be found in [26]; see also Gillis [36, Remark 2]. We provide a short proof because it is instructive for the new leg of the proof for the symmetric case. Suppose $\mathcal{I}_{k_1} \subseteq \mathcal{I}_{k_2}$, then there exist a positive scalar α such that

$$\mathbf{W}_{:,k_2} - \alpha\mathbf{W}_{:,k_1} \geq 0$$

Define a $K \times K$ upper triangular matrix $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{I} - \alpha\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T$$

then

$$\mathbf{A}^{-1} = \mathbf{I} + \alpha \mathbf{e}_{k_1} \mathbf{e}_{k_2}^T$$

Let $\tilde{\mathbf{W}} = \mathbf{WA}$ and $\tilde{\mathbf{H}} = \mathbf{A}^{-1}\mathbf{H}$. Since $\mathbf{A}^{-1} \geq 0$, $\tilde{\mathbf{H}} \geq 0$. $\tilde{\mathbf{W}}$ satisfies that

$$\tilde{\mathbf{W}}_{:,k} = \begin{cases} \mathbf{W}_{:,k_2} - \alpha \mathbf{W}_{:,k_1}, & k = k_2 \\ \mathbf{W}_{:,k}, & \text{else} \end{cases}$$

Therefore $\tilde{\mathbf{W}} \geq 0$, which means $\mathbf{S} = \tilde{\mathbf{W}}\tilde{\mathbf{H}}$ is an alternative NMF of $\mathbf{S}$. A similar argument can be applied when $\mathcal{J}_{k_1} \subseteq \mathcal{J}_{k_2}$.

B. Symmetric Case

Suppose $\mathcal{I}_{k_1} \subseteq \mathcal{I}_{k_2}$, then there exists a positive scalar α such that

$$\mathbf{W}_{:,k_2} - \alpha \mathbf{W}_{:,k_1} \geq 0$$

Let $\mathbf{A}$ be a Givens rotation matrix [43] defined as

$$\mathbf{A} = [\mathbf{e}_{k_1} \quad \mathbf{e}_{k_2}] \begin{bmatrix} c & -s \\ s & c \end{bmatrix} [\mathbf{e}_{k_1} \quad \mathbf{e}_{k_2}]^T + \sum_{\substack{k=1 \\ k \neq k_1, k_2}}^{K} \mathbf{e}_k \mathbf{e}_k^T$$

where

$$c = \frac{1}{\sqrt{1+\alpha^2}}, \quad s = \frac{\alpha}{\sqrt{1+\alpha^2}}$$

Since $\mathbf{A}$ is unitary but not a permutation matrix, and $\mathbf{WA} \geq 0$, according to Definition 5, the symmetric NMF of $\mathbf{S}$ is not unique.

Appendix C
Proof of Theorem 4

We first prove that if $\operatorname{cone}(\mathbf{W}^T) \supseteq \mathcal{C}$, then $\forall k = 1, \ldots, K$, $\mathbf{e}_k$ is an extreme ray of $\operatorname{cone}(\mathbf{W}^T)^*$. This is because

1) $\mathbf{e}_k$ is an extreme ray of $\mathcal{C}^*$ (the constraint (2) is satisfied as equality at $\mathbf{e}_k$, hence $\mathbf{e}_k$'s are on the boundary of $\mathcal{C}^*$, and every ray that lies on the boundary of a second order cone is an extreme ray of this cone); and
2) $\mathbf{e}_k$ is contained in $\operatorname{cone}(\mathbf{W}^T)^*$ (obviously $\mathbf{W}\mathbf{e}_k \geq 0$).

The $\mathbf{e}_k$'s being *extreme* rays of $\mathcal{C}^*$ means that there do not exist two other elements in $\mathcal{C}^*$ that can conically generate $\mathbf{e}_k$, and since $\operatorname{cone}(\mathbf{W}^T)^* \subseteq \mathcal{C}^*$, there certainly do not exist two other elements in $\operatorname{cone}(\mathbf{W}^T)^*$ that can conically generate $\mathbf{e}_k$; therefore, the $\mathbf{e}_k$'s are the extreme rays of $\operatorname{cone}(\mathbf{W}^T)^*$. Since $\operatorname{cone}(\mathbf{W}^T)^* \subseteq \mathcal{C}^*$, the condition given in Theorem 4 implies that all other extreme rays of $\operatorname{cone}(\mathbf{W}^T)^*$ lie strictly inside $\mathcal{C}^*$.

A. Asymmetric Case

According to Lemma 3, we need to show that under this condition, if a simplicial cone $\mathcal{A}$ satisfies that $\operatorname{cone}(\mathbf{W}^T) \subseteq \mathcal{A} \subseteq \operatorname{cone}(\mathbf{H})^*$, then $\mathcal{A} = \mathbb{R}_+^K$. Since $\operatorname{cone}(\mathbf{W}^T) \supseteq \mathcal{C}$ and $\operatorname{cone}(\mathbf{H}) \supseteq \mathcal{C}$, obviously $\mathcal{C} \subseteq \mathcal{A} \subseteq \mathcal{C}^*$. According to Lemma 3, we know that $\mathcal{A}$ can only be a rotated version of $\mathbb{R}_+^K$, and that all of its extreme rays lie on the boundary of $\mathcal{C}^*$. However, since none of the extreme rays of $\operatorname{cone}(\mathbf{H})^*$ except $\mathbf{e}_k$'s lie on the boundary of $\mathcal{C}^*$, $\mathcal{A}$ can only be the non-negative orthant $\mathbb{R}_+^K$ itself. Therefore under this condition the asymmetric NMF $\mathbf{S} = \mathbf{WH}$ is unique.

B. Symmetric Case

According to Lemma 4, if the symmetric NMF is unique, then we cannot rotate $\mathbb{R}_+^K$ to $\mathcal{A}$ such that $\operatorname{cone}(\mathbf{W}^T) \subseteq \mathcal{A}$. Since $\operatorname{cone}(\mathbf{W}^T) \supseteq \mathcal{C}$, and $\mathcal{A}$ is self-dual, then we have $\mathcal{A} \subseteq \operatorname{cone}(\mathbf{W}^T)^* \subseteq \mathcal{C}^*$. Again, any rotated version of $\mathbb{R}_+^K$ that is a subset of $\mathcal{C}^*$ satisfies that its extreme rays lie on the boundary of $\mathcal{C}^*$. However, none of the extreme rays of $\operatorname{cone}(\mathbf{W}^T)$ except $\mathbf{e}_k$'s lie on the boundary of $\mathcal{C}^*$, therefore $\mathcal{A}$ can only be the non-negative orthant $\mathbb{R}_+^K$ itself. As a result, under this condition the symmetric NMF $\mathbf{S} = \mathbf{W}\mathbf{W}^T$ is unique.

Appendix D
Proof of Proposition 1

A. $\nexists \mathcal{W}$ Such That $\operatorname{cone}(\mathbf{W}) \subset \mathcal{W} \subseteq \mathcal{P}_{\mathcal{V}}$

Recall that the columns of $\mathbf{W}$ are in the polyhedral cone $\mathcal{P}_{\mathcal{V}} = \{\mathbf{w} | \mathbf{U}^T\mathbf{w} = 0, \mathbf{w} \geq 0\}$, which is a subset of $\mathbb{R}^I$. If a direction is an extreme ray of a polyhedral cone in $\mathbb{R}^I$ described by inequalities, then at least $I - 1$ linearly independent constraints are satisfied as equalities. Suppose $\mathbf{w}$ is one column of $\mathbf{W}$. $\mathbf{U}$ is $I \times (I - K)$, so $\mathbf{U}^T\mathbf{w} = \mathbf{0}$ gives us $I - K$ equalities; if $\mathbf{w}$ is a column of $\mathbf{W}$, it has at least $K - 1$ zeros since $\mathbf{e}_k$ is an extreme ray of $\operatorname{cone}(\mathbf{W}^T)^*$, so at least $K - 1$ of the inequality constraints $\mathbf{w} \geq 0$ are satisfied as equalities. Therefore we have overall at least $I - 1$ constraints satisfied as equalities, and so all columns of $\mathbf{W}$ lie on different extreme rays of $\mathcal{P}_{\mathcal{V}}$. Donoho and Stodden [24, Lemma 4] proved that under this case, there does not exist another simplicial cone $\mathcal{W}$ with K extreme rays such that $\operatorname{cone}(\mathbf{W}) \subset \mathcal{W} \subseteq \mathcal{P}_{\mathcal{V}}$.

B. $\nexists \mathcal{W}$ Such That $\operatorname{cone}(\mathbf{S}) \subseteq \mathcal{W} \subset \operatorname{cone}(\mathbf{W})$

$\operatorname{cone}(\mathbf{W})$ is a simplicial cone with K extreme rays, so every $K - 1$ of its extreme rays, i.e., $K - 1$ columns of $\mathbf{W}$, define a facet of $\operatorname{cone}(\mathbf{W})$. If $\mathbf{H}_{k,j} = 0$, then $\mathbf{S}_{:,j} = \mathbf{W}\mathbf{H}_{:,j}$ is a linear combination of all the columns of $\mathbf{W}$ except the k-th one. Therefore $\mathbf{S}_{:,j}$ is in the facet defined by all the columns of $\mathbf{W}$ except the k-th one. We have assumed that every row of $\mathbf{H}$ has at least $K-1$ zeros. Therefore, there are at least $K-1$ columns of $\mathbf{S}$ on each facet of $\operatorname{cone}(\mathbf{W})$. If they are linearly independent, which is guaranteed by the fact that $\mathbf{e}_k$ is an extreme ray of $\operatorname{cone}(\mathbf{H})^*$, then every facet of $\operatorname{cone}(\mathbf{W})$ contains $K - 1$ linearly independent columns of $\mathbf{S}$. Then there does not exist another order-K simplicial cone $\mathcal{W}$ such that $\operatorname{cone}(\mathbf{S}) \subseteq \mathcal{W} \subset \operatorname{cone}(\mathbf{W})$, as has been argued in the proof of [24, Theorem 1].

Appendix E
Modified Symmetric NMF Algorithm for the Case Rank < CP-Rank

The completely positive rank (cp-rank) of a completely positive matrix $\mathbf{S}$ is the minimum K that allows exact symmetric NMF of $\mathbf{S}$ [27]. It is well-known that the cp-rank need not be equal to the rank of $\mathbf{S}$. If we indeed encounter a completely positive matrix with cp-rank strictly larger than its rank, this Appendix shows how to modify the algorithm given in Fig. 4 to seek an exact symmetric NMF, assuming we know the cp-rank.

Let the rank of the complete positive matrix $\mathbf{S}$ be r, then we can perform the thin eigenvalue decomposition

$$\mathbf{S} = \mathbf{U}_s \mathbf{\Lambda}_s \mathbf{U}_s^T$$

where $\mathbf{U}_s$ is $I \times r$ with orthonormal columns, and $\mathbf{\Lambda}_s$ is $r \times r$ diagonal. Assume the cp-rank is K, then there exists a symmetric NMF of $\mathbf{S}$

$$\mathbf{S} = \mathbf{W}\mathbf{W}^T$$

where $\mathbf{W}$ is $I \times K$ with non-negative elements. The rank of $\mathbf{W}$ is r, otherwise the rank of $\mathbf{S}$ would not be r. Therefore, we can take the thin SVD of $\mathbf{W}$

$$\mathbf{W} = \mathbf{U}_w \mathbf{\Sigma}_w \mathbf{V}_w^T$$

where $\mathbf{U}_w$ is $I \times r$ with orthonormal columns, $\mathbf{\Sigma}_w$ is $r \times r$ diagonal, and $\mathbf{V}_w$ is $r \times K$ with orthonormal columns. Then

$$\mathbf{S} = \mathbf{W}\mathbf{W}^T = \mathbf{U}_w \mathbf{\Sigma}_w^2 \mathbf{U}_w^T$$

As we can see, the right-hand-side is also an eigenvalue decomposition. Since the eigenvalue decomposition is unique, this implies that $\mathbf{U}_s = \mathbf{U}_w$ and $\mathbf{\Lambda}_s^{1/2} = \mathbf{\Sigma}_w$. In other words, let

$$\mathbf{S} = \mathbf{B}\mathbf{B}^T$$

where $\mathbf{B} = \mathbf{U}_s \mathbf{\Lambda}_s^{1/2}$, then there exists a $K \times r$ orthonormal matrix $\mathbf{Q}$ such that $\mathbf{B}\mathbf{Q}^T \geq 0$. Thus, finding $\mathbf{W}$ can be posed as the following optimization problem

$$\min_{\mathbf{W},\mathbf{Q}} \quad \|\mathbf{W} - \mathbf{B}\mathbf{Q}^T\|_F^2 \tag{15a}$$

$$\text{subject to} \quad \mathbf{W} \geq 0, \mathbf{Q}^T\mathbf{Q} = \mathbf{I} \tag{15b}$$

Notice that compared to problem (5), we only have $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, since $\mathbf{Q}$ is not square, and $\mathbf{Q}\mathbf{Q}^T$ is now a projection matrix.

Similar to algorithm given in Fig. 4, we propose to solve this problem by alternatingly updating $\mathbf{W}$ and $\mathbf{Q}$. For $\mathbf{W}$, the update rule is simply

$$\mathbf{W} \leftarrow \max(0, \mathbf{B}\mathbf{Q}^T) \tag{16}$$

For the case of $\mathbf{Q}$, the answer is not directly given by the Procrustes rotation. However, it is easy to show that the solution is similar to what the Procrustes rotation provided us in the unitary $\mathbf{Q}$ case. Since

$$\begin{aligned}\|\mathbf{W} - \mathbf{B}\mathbf{Q}^T\|_F^2 &= \mathrm{tr}\{(\mathbf{W} - \mathbf{B}\mathbf{Q}^T)(\mathbf{W} - \mathbf{B}\mathbf{Q}^T)^T\} \\ &= \mathrm{tr}\{\mathbf{W}\mathbf{W}^T\} + \mathrm{tr}\{\mathbf{B}\mathbf{B}^T\} - 2\mathrm{tr}\{\mathbf{W}^T\mathbf{B}\mathbf{Q}^T\}\end{aligned}$$

minimizing $\|\mathbf{W} - \mathbf{B}\mathbf{Q}^T\|_F^2$ is equivalent to maximizing $\mathrm{tr}\{\mathbf{W}^T\mathbf{B}\mathbf{Q}^T\}$.

Proposition 3: The solution of the following optimization problem

$$\begin{aligned}\max_{\mathbf{Q}} \quad & \mathrm{tr}\{\mathbf{W}^T\mathbf{B}\mathbf{Q}^T\} \\ \text{subject to} \quad & \mathbf{Q}^T\mathbf{Q} = \mathbf{I}\end{aligned}$$

is $\mathbf{Q} = \mathbf{U}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ come from the singular value decomposition of $\mathbf{W}^T\mathbf{B}$.

$$\mathbf{W}^T\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

Proof: Let the singular value decomposition of $\mathbf{W}^T\mathbf{B}$ be

$$\mathbf{W}^T\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

where $\mathbf{U}$ is $K \times r$, $\mathbf{\Sigma}$ is $r \times r$ and $\mathbf{V}$ is $r \times r$. Then we have

$$\begin{aligned}\mathrm{tr}\{\mathbf{W}^T\mathbf{B}\mathbf{Q}^T\} &= \mathrm{tr}\{\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{Q}^T\} \\ &= \mathrm{tr}\{\mathbf{\Sigma}\mathbf{V}^T\mathbf{Q}^T\mathbf{U}\} \\ &= \sum_{k=1}^{r} \sigma_k \mathbf{v}_k^T\mathbf{Q}^T\mathbf{u}_k\end{aligned}$$

where σ_k is the k-th diagonal entry of $\mathbf{\Sigma}$, and $\mathbf{v}_k$, $\mathbf{u}_k$ are the k-th column of $\mathbf{V}$, $\mathbf{U}$, respectively. Therefore,

$$\mathrm{tr}\{\mathbf{W}^T\mathbf{B}\mathbf{Q}^T\} \leq \sum_{k=1}^{r} \sigma_k$$

because

$$\begin{aligned}\mathbf{v}_k^T\mathbf{Q}^T\mathbf{u}_k &\leq \|\mathbf{v}_k\|_2\|\mathbf{Q}^T\mathbf{u}_k\|_2 \\ &\leq \|\mathbf{v}_k\|_2\|\mathbf{u}_k\|_2 \\ &= 1\end{aligned}$$

Furthermore, let $\mathbf{Q} = \mathbf{U}\mathbf{V}^T$, then

$$\begin{aligned}\mathrm{tr}\{\mathbf{W}^T\mathbf{B}\mathbf{Q}^T\} &= \mathrm{tr}\{\mathbf{\Sigma}\mathbf{V}^T\mathbf{Q}^T\mathbf{U}\} \\ &= \mathrm{tr}\{\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{U}^T\mathbf{U}\} \\ &= \mathrm{tr}\{\mathbf{\Sigma}\}\end{aligned}$$

which attains the upper bound we just obtained. Therefore, the solution for this optimization problem is $\mathbf{Q} = \mathbf{U}\mathbf{V}^T$. ∎

Example 5: Let

$$\mathbf{S} = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}$$

Then $\mathrm{rank}(\mathbf{S}) = 3$ while $\mathrm{cp-rank}(\mathbf{S}) = 4$. If we apply the algorithm provided in Fig. 4, in which case we set $K = 3$, the result is

$$\mathbf{W} = \begin{bmatrix} 0 & 1.3635 & 0.3630 \\ 0 & 0.3637 & 1.3633 \\ 1.1193 & 0.8516 & 0 \\ 1.1190 & 0 & 0.8520 \end{bmatrix}$$

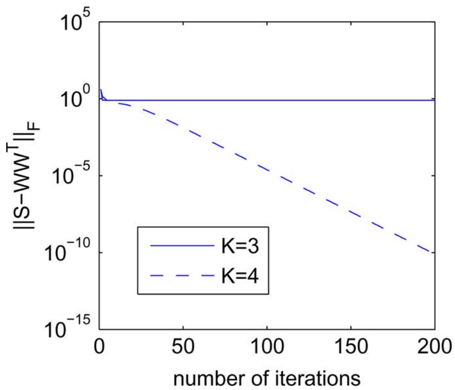

Fig. 7. Convergence of the modified proposed algorithm vs. original algorithm given in Fig. 4 for a matrix whose rank is less than its cp-rank.

and the factorization is not exact, whereas if we set $K = 4$ and apply the modified algorithm described in this Appendix, the result is

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

and the factorization is exact. The convergence of each case is shown in Fig. 7.

ACKNOWLEDGMENT

The authors would like to thank David Brie, Said Moussaoui, and Lek-Heng Lim for providing useful references, and Xiao Fu for pointing out that [41, Theorem 3.1] can be used to prove convergence of our symmetric NMF algorithm to a stationary point.

REFERENCES

[1] K. Huang, N. D. Sidiropoulos, and A. Swami, “NMF revisited: New uniqueness results and algorithms,” in *Proc. IEEE ICASSP 2013*, Vancouver, Canada, May 26–31, 2013.

[2] S. L. Campbell and G. D. Poole, “Computing nonnegative rank factorizations,” *Linear Algebra Appl.*, vol. 35, pp. 175–182, 1981.

[3] J.-C. Chen, “The nonnegative rank factorizations of nonnegative matrices,” *Linear Algebra Appl.*, vol. 62, pp. 207–217, 1984.

[4] P. Paatero and U. Tapper, “Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values,” *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.

[5] A. Berman and R. J. Plemmons, *Nonnegative Matrices*. Philadelphia, PA, USA: SIAM, 1979.

[6] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

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

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

[9] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” *Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 13, pp. 556–562, 2001.

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

[11] R. Zdunek and A. Cichocki, “Nonnegative matrix factorization with constrained second-order optimization,” *Signal Process.*, vol. 87, no. 8, pp. 1904–1916, 2007.

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

[13] A. Cichocki and A.-H. Phan, “Fast local algorithms for large scale nonnegative matrix and tensor factorizations,” *IEICE Trans. Fund. Electron., Commun., Comput. Sci.*, vol. 92, no. 3, pp. 708–721, 2009.

[14] J. Li and J. M. Bioucas-Dias, “Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data,” in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS) 2008*, vol. 3, pp. 250–253.

[15] B. Klingenberg, J. Curry, and A. Dougherty, “Non-negative matrix factorization: ILL-posedness and a geometric algorithm,” *Pattern Recogn.*, vol. 42, no. 5, pp. 918–928, 2009.

[16] R. Zdunek, “Initialization of nonnegative matrix factorization with vertices of convex polytope,” in *Artificial Intelligence and Soft Computing*. New York, NY, USA: Springer, 2012, pp. 448–455.

[17] W. Ouedraogo, A. Souloumiac, M. Jaidane, and C. Jutten, “Geometrical method using simplicial cones for overdetermined nonnegative blind source separation: Application to real PET images,” *Latent Variable Anal. Signal Separat.*, vol. 7191, pp. 494–501, 2012.

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

[19] F. A. Nielsen, “Clustering of scientific citations in Wikipedia,” in *Proc. WikiMania*, Alexandria, Egypt, Jul. 2008.

[20] W. Xu, X. Li, and Y. Gong, “Document clustering based on non-negative matrix factorization,” in *Proc. 26th Annu. ACM Int. Conf. Res. Develop. Inf. Retrieval (SIGIR’03)*, Aug. 2003, pp. 267–273.

[21] F. J. Theis, K. Stadlthanner, and T. Tanaka, “First results on uniqueness of sparse non-negative matrix factorization,” in *Proc. 13th Eur. Signal Process. Conf. (EUSIPCO’05)*, Antalya, Turkey, Sep. 4–8, 2005.

[22] R. Schachtner, G. Pöppel, and E. W. Lang, “Towards unique solutions of non-negative matrix factorization problems by a determinant criterion,” *Digit. Signal Process.*, vol. 21, no. 4, pp. 528–534, 2011.

[23] L. B. Thomas, “Solution to problem 73–14, rank factorization of nonnegative matrices (by A. Berman, and R.J. Plemmons),” *SIAM Rev.*, vol. 16, no. 3, pp. 393–394, 1974.

[24] D. L. Donoho and V. C. Stodden, “When does non-negative matrix factorization give a correct decomposition into parts?,” in *Advances in Neural Information Processing Systems (NIPS)*. Cambridge, MA, USA: MIT Press, 2003, vol. 16, pp. 1141–1148.

[25] H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen, “Theorems on positive data: On the uniqueness of NMF,” *Comput. Intell. Neurosci.*, vol. 2008, 2008 [Online]. Available: http://www.hindawi.com/journals/cin/2008/764206/cta/, DOI:10.1155/2008/764206

[26] S. Moussaoui, D. Brie, and J. Idier, “Non-negative source separation: Range of admissible solutions and conditions for the uniqueness of the solution,” in *IEEE ICASSP 2005*, Philadelphia, PA, USA, Mar. 18–23, 2005, vol. 5, pp. 289–292.

[27] A. Berman and N. Shaked-Monderer, *Completely Positive Matrices*. Singapore: World Scientific, 2003.

[28] P. J. C. Dickinson and L. Gijben, “On the computational complexity of membership problems for the completely positive cone and its dual,” *Comput. Optim. Appl.* 2013 [Online]. Available: http://dx.doi.org/10.1007/s10589-013-9594-z

[29] Z. He, S. Xie, R. Zdunek, G. Zhou, and A. Cichocki, “Symmetric nonnegative matrix factorization: Algorithms and applications to probabilistic clustering,” *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2117–2131, 2011.

[30] M. Catral, L. Han, M. Neumann, and R. J. Plemmons, “On reduced rank nonnegative matrix factorization for symmetric nonnegative matrices,” *Linear Algebra Appl.*, vol. 393, pp. 107–126, 2004.

[31] C. Ding, X. He, and H. D. Simon, “On the equivalence of nonnegative matrix factorization and spectral clustering,” in *Proc. SIAM Data Min. Conf.*, 2005, pp. 606–610.

[32] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1996, vol. 28.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[34] L. Khachiyan, E. Boros, K. Borys, K. Elbassioni, and V. Gurvich, “Generating all vertices of a polyhedron is hard,” *Discrete Comput. Geometry*, vol. 39, no. 1, pp. 174–190, 2008.

[35] H. Laurberg, “Non-negative matrix factorization: Theory and methods,” Ph.D. dissertation, Aalborg Univ., Aalborg, Denmark, 2008.

[36] N. Gillis, “Sparse and unique nonnegative matrix factorization through data preprocessing,” *J. Mach. Learn. Res.*, vol. 13, pp. 3349–3386, 2012.

[37] R. M. Freund and J. B. Orlin, “On the complexity of four polyhedral set containment problems,” *Math. Programm.*, vol. 33, no. 2, pp. 139–145, 1985.

[38] P. H. Schönemann, “A generalized solution of the orthogonal Procrustes problem,” *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.

[39] D. B. Bertsekas, *Nonlinear Probramming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999.

[40] L. Grippo and M. Sciandrone, “On the convergence of the block nonlinear Gauss-Seidel method under convex constraints,” *Oper. Res. Lett.*, vol. 26, no. 3, pp. 127–136, 2000.

[41] B. Chen, S. He, Z. Li, and S. Zhang, “Maximum block improvement and polynomial optimization,” *SIAM J. Optim.*, vol. 22, no. 1, pp. 87–107, 2012.

[42] M. R. Sepanski, *Compact Lie Groups*. New York, NY, USA: Springer, 2007, vol. 235.

[43] G. H. Golub and C. F. Van Loan, *Matrix Computations*. : Johns Hopkins University Press, 1996.

[44] H. W. Kuhn and A. W. Tucker, “Nonlinear programming,” in *Proc. 2nd Berkeley Symp. Math. Statist. Probab.*, 1951, vol. 1, pp. 481–492.

[45] 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, 2012.

Kejun Huang (S’13) received the B.Eng. in communication engineering from Nanjing University of Information Science and Technology, Nanjing, China in 2010. Since September 2010, he has been working towards his Ph.D. degree in the Department of Electrical and Computer Engineering, University of Minnesota. His research interests include signal processing, machine learning, and data analytics. His current research focuses on identifiability, algorithms, and performance analysis for factor analysis of big matrix and tensor data.

Nicholas D. Sidiropoulos (F’09) received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Greece, and M.S. and Ph.D. degrees in electrical engineering from the University of Maryland-College Park, in 1988, 1990 and 1992, respectively. He served as Assistant Professor at the University of Virginia (1997–1999); Associate Professor at the University of Minnesota-Minneapolis (2000–2002); Professor at the Technical University of Crete, Greece (2002–2011); and Professor at the University of Minnesota-Minneapolis (2011-). His current research focuses primarily on signal and tensor analytics, with applications in cognitive radio, big data, and preference measurement. He received the NSF/CAREER award (1998), the IEEE Signal Processing Society (SPS) Best Paper Award (2001, 2007, 2011), and the IEEE SPS Meritorious Service Award (2010). He has served as IEEE SPS Distinguished Lecturer (2008–2009), and Chair of the IEEE Signal Processing for Communications and Networking Technical Committee (2007–2008). He received the Distinguished Alumni Award of the Department of Electrical and Computer Engineering, University of Maryland, College Park (2013).

Ananthram Swami (F’08) received the B.Tech. degree from IIT-Bombay; the M.S. degree from Rice University, and the Ph.D. degree from the University of Southern California (USC), all in electrical engineering. He has held positions with Unocal Corporation, USC, CS-3 and Malgudi Systems. He was a Statistical Consultant to the California Lottery, developed a Matlab-based toolbox for non-Gaussian signal processing, and has held visiting faculty positions at INP, Toulouse. He is with the US Army Research Laboratory (ARL) where he is the ST for Network Science. His work is in the broad area of network science, with emphasis on wireless communication networks. He is an ARL Fellow and a Fellow of the IEEE. Dr. Swami serves on the Senior Editorial Board of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING. He has served as Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, SPL, SPM, C&S, TWC, and as guest editor for JSAC. He was a tutorial speaker on “Networking Cognitive Radios for Dynamic Spectrum Access” at ICASSP 2008, DySpan 2008, MILCOM 2008, and ICC 2010, co-editor of the 2007 Wiley book *Wireless Sensor Networks: Signal Processing & Communications Perspectives*; recipient of best conference paper award at IEEE Trustcom 2009, co-organizer and co-chair of three IEEE workshops related to signal processing and communications, including IEEE SPAWC’10.