NMF REVISITED: NEW UNIQUENESS RESULTS AND ALGORITHMS

K. Huang, N.D. Sidiropoulos[*]

Dept. of ECE, Univ. of Minnesota
Minneapolis, MN 55455, USA

A. Swami[†]

Army Research Lab
Adelphi, MD 20783, USA

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 and computational 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 and insightful 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 is shown to be NP-complete; yet it offers insights on latent sparsity, as is also shown in a new necessary condition, to a smaller extent. On the computational side, a new efficient algorithm for symmetric NMF is proposed which uses Procrustes rotations. Simulation results show promising performance with respect to the state-of-art. The new algorithm is also applied to a clustering problem for co-authorship data, yielding meaningful and interpretable results.

Index Terms— Non-negative Matrix Factorization, Uniqueness, Simplicial cone, Dual cone, Procrustes rotation

1. INTRODUCTION

Non-negative matrix factorization (NMF) $\mathbf{S} = \mathbf{W}\mathbf{H}$ with $\mathbf{W}$, $\mathbf{H}$ having non-negative elements, was first proposed by Paatero and Tapper [1] who called it positive matrix factorization. Lee and Seung [2] first discovered a very interesting property of NMF when applied to image processing, namely that "NMF is able to learn the parts of objects" – meaning, it tends to decompose objects in meaningful parts. Lee and Seung [2] popularized NMF, which quickly found numerous other applications in diverse disciplines – see [3] and references therein.

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 $\mathbf{W}, \mathbf{H}$. 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. Donoho and Stodden [4] were the first to study the uniqueness of NMF, and they provided a sufficient condition showing that if $\mathbf{W}$ follows the *complete factorial sampling rule* and $\mathbf{H}$ follows the *separability rule*, then the NMF of their product is unique. Laurberg *et al.* [5] loosened the sufficient condition to what they called a *strongly boundary close* $\mathbf{W}$ and *sufficiently spread* $\mathbf{H}^T$. Essentially, these two latter conditions require the columns of $\mathbf{H}$ to contain scaled versions of all columns of the identity matrix, which is a very strict condition. Laurberg *et al.* also gave a necessary condition called *boundary close*.

The symmetric version of NMF ($\mathbf{P} = \mathbf{W}\mathbf{W}^T$ with $\mathbf{W}$ having non-negative elements) is relatively less studied [6–8]. He *et al.* [8] derived three algorithms for symmetric NMF – the multiplicative update (see also [7]), α-SNMF and β-SNMF, and simulations showed that the latter two outperform other alternatives.

• **Notation:** $\mathbf{A}_{i,:}$ 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., $\mathcal{A}$. $\mathbb{R}^n_+ = \{\mathbf{x}|x_i \geq 0, i = 1, \cdots, n\}$ 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. Inequality marks represent element-wise inequalities, whether applied to scalars, vectors or matrices. Asymmetric NMF is written out as $\mathbf{S} = \mathbf{W}\mathbf{H}$, where $\mathbf{S}$ is $I \times J$, $\mathbf{W}$ is $I \times K$ and $\mathbf{H}$ is $K \times J$. 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 are no $\mathbf{0}$ columns or rows in any matrix. If this happens, we can simply delete them first.

• **Preliminaries:** We briefly review some prerequisites from convex analysis; see [9, 10] 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: 1) by taking the intersection of a finite number of halfspaces $\mathcal{K} = \{\mathbf{x} \in \mathbb{R}^n|\mathbf{A}\mathbf{x} \geq \mathbf{0}\}$. If $\mathbf{A}_{i,:}\mathbf{x} \geq 0$ is not a redundant constraint, then $\mathcal{K} \cap \{\mathbf{x}|\mathbf{A}_{i,:}\mathbf{x} = 0\}$ is called a *facet* of $\mathcal{K}$; 2) by taking the conic hull of a finite number of vectors $\mathcal{K} = \{\mathbf{x} = \mathbf{B}\lambda|\lambda \geq \mathbf{0}\} = \text{cone}(\mathbf{B})$. If $\mathbf{B}_{:,i}$ cannot be represented by the conic combinations of the other columns of $\mathbf{B}$, then it is called an *extreme ray* of $\mathcal{K}$.

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

If $\mathcal{K} = \{\mathbf{x} = \mathbf{B}\lambda|\lambda \geq \mathbf{0}\} = \text{cone}(\mathbf{B})$ is a simplicial cone, then for every element $\mathbf{x} \in \mathcal{K}$, there is a *unique* corresponding λ 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.

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.* [5]): 1) If $\mathcal{A} = \text{cone}(\mathbf{A})$, then $\mathcal{A}^* = \{\mathbf{x}|\mathbf{A}^T\mathbf{x} \geq 0\}$; 2)

[*] E-mail: (huang663,nikos)@umn.edu, Tel: +16126251242, Fax: +16126254583. Supported in part by ARO STIR W911NF-11-1-0500, and NSF IIS-1247632.

[†] E-mail: a.swami@ieee.org

If $\mathcal{A} = \text{cone}(\mathbf{A}^T)$, where $\mathbf{A}$ is invertible, then $\mathcal{A}^* = \text{cone}(\mathbf{A}^{-1})$; 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.

Example 1. [4] Define the second-order cone in $\mathbb{R}^n$: $\mathcal{C} = \{\mathbf{x} | \mathbf{x}^T\mathbf{1} \geq \sqrt{n-1}\|\mathbf{x}\|_2\}$. Its dual cone is another second-order cone $\mathcal{C}^* = \{\mathbf{x} | \mathbf{x}^T\mathbf{1} \geq \|\mathbf{x}\|_2\}$

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: $\mathcal{C} \subseteq \mathbb{R}^n_+ \subseteq \mathcal{C}^*$. Moreover, this relation is tight, as stated in the following lemma (proof skipped due to space considerations – will be provided in the journal version).

Lemma 1. *If a simplicial cone $\mathcal{K}$ satisfies $\mathcal{C} \subseteq \mathcal{K} \subseteq \mathcal{C}^*$, then $\mathcal{K}$ is a rotated version of $\mathbb{R}^n_+$.*

2. UNIQUENESS OF NMF

For uniqueness analysis, we assume that $K = \text{rank}(\mathbf{S})$, and thus both $\mathbf{W}$ and $\mathbf{H}$ are full rank.

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

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

• **Donoho and Stodden's Analysis:** Each column of $\mathbf{S}$ is a non-negative linear combination of all the columns of $\mathbf{W}$, therefore $\text{cone}(\mathbf{S}) \subseteq \text{cone}(\mathbf{W})$. Furthermore, let $\mathcal{P}_\mathcal{V} = \mathbb{R}^K_+ \cap \text{span}(\mathbf{S})$, since $\text{rank}(\mathbf{W}) = \text{rank}(\mathbf{S})$, $\mathbf{W} \geq 0$, obviously $\text{cone}(\mathbf{W}) \subseteq \mathcal{P}_\mathcal{V}$. Thus, we have the following relation

$$\text{cone}(\mathbf{S}) \subseteq \text{cone}(\mathbf{W}) \subseteq \mathcal{P}_\mathcal{V}, \tag{1}$$

and NMF can be interpreted as finding a simplicial cone with K extreme rays that satisfies (1). This naturally leads to the following lemma.

Lemma 2. [4] *The NMF of the non-negative matrix $\mathbf{S}$ is unique if and only if there is a unique order-K simplicial cone $\mathcal{W}$ such that $\text{cone}(\mathbf{S}) \subseteq \mathcal{W} \subseteq \mathcal{P}_\mathcal{V}$.*

Once we find such a simplicial cone, the matrix $\mathbf{W}$ can be obtained by taking all the extreme rays of $\mathcal{W}$ as its columns. Obviously the extreme rays are invariant under scaling, and it does not matter how we order the extreme rays, which is in agreement with Definition 4.

• **Laurberg's Analysis:** Laurberg *et al.* [5] offered a different view point on the uniqueness of NMF, summarized in Lemma 3.

Lemma 3. [5] *If $\text{rank}(\mathbf{S}) = K$, the NMF $\mathbf{S} = \mathbf{WH}$ is unique if and only if the positive orthant is the only simplicial cone $\mathcal{A}$ with K extreme rays that satisfies $\text{cone}(\mathbf{W}^T) \subseteq \mathcal{A} \subseteq \text{cone}(\mathbf{H})^*$.*

Both Donoho *et al.* and Laurberg *et al.* derived their own sufficient conditions based on the lemmas that they proposed. However, these sufficient conditions both require one of the latent factor matrices to contain a diagonal matrix as a sub-matrix (after proper permutation) – a very strict condition that is unlikely to hold in practice. Can we come up with a more general condition that satisfies Lemma 2 and Lemma 3?

• **New Results on Uniqueness of Symmetric and Asymmetric NMF:** We are now ready to present our new conditions on the uniqueness of symmetric and asymmetric NMF (all proofs skipped due to space considerations – they will be provided in the journal version). We start with a necessary condition.

Theorem 1 (Necessary Condition)**.** *Define $\mathcal{I}_k = \{i \,|\, \mathbf{W}_{i,k} \neq 0\}$ and $\mathcal{J}_k = \{j \,|\, \mathbf{H}_{k,j} \neq 0\}$. If the NMF $\mathbf{S} = \mathbf{WH}$ is unique, then there do not exist $k_1, k_2 \in \{1, \cdots, K\}, k_1 \neq k_2$ 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 $\mathbf{H} = \mathbf{W}^T$.*

For the case of asymmetric NMF, we recently found the condition in Theorem 1 in the preprint of [11, Remark 2]. What is necessary for uniqueness of asymmetric NMF, however, is not automatically necessary for uniqueness of symmetric NMF; and we claim that the same result holds for symmetric NMF as well.

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

Corollary 1 is exactly the "boundary close" condition given in [5]. Using Donoho and Stodden's analysis, the requirement that every column of $\mathbf{W}$ has a zero entry means that every extreme ray of $\text{cone}(\mathbf{W})$ 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 $\text{cone}(\mathbf{W})$. This condition is intuitive, because otherwise we can always perturb $\text{cone}(\mathbf{W})$ into a slightly bigger or smaller cone that still satisfies $\text{cone}(\mathbf{S}) \subseteq \text{cone}(\mathbf{W}) \subseteq \mathcal{P}_\mathcal{V}$, so that NMF won't be unique according to Lemma 2.

We need the second-order cone $\mathcal{C}$ defined in Example 1 for our sufficient condition.

Theorem 2 (Sufficient Condition)**.** *If $\text{rank}(\mathbf{W}) = \text{rank}(\mathbf{H}) = K$, $\text{cone}(\mathbf{W}^T) \supseteq \mathcal{C}$ and $\text{cone}(\mathbf{H}) \supseteq \mathcal{C}$, and none of the extreme rays of $\text{cone}(\mathbf{W}^T)^*$ or $\text{cone}(\mathbf{H})^*$ except $\mathbf{e}_k$'s lie on the boundary of $\mathcal{C}^*$, then the NMF $\mathbf{S} = \mathbf{WH}$ is unique. This condition is also sufficient in the symmetric case, i.e., when $\mathbf{H} = \mathbf{W}^T$.*

Example 2. [5] Consider the symmetric NMF $\mathbf{S} = \mathbf{WW}^T$ 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 $\omega < 0.5$.

If $\omega < 0.5$, $\text{cone}(\mathbf{W}^T) \supseteq \mathcal{C}$, and none of the extreme rays of it except $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ lie on the boundary of $\mathcal{C}^*$, so, according to Theorem 2, symmetric NMF is unique. If $\omega \geq 0.5$, we can explicitly construct an alternative NMF. Laurberg *et al.* [5] 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 [5] fails to demonstrate when uniqueness holds in this case, except for $\omega = 0$; whereas our new sufficient condition in Theorem 2 is able to identify the full interval where uniqueness holds in this case.

Corollary 2. *If the NMF $\mathbf{S} = \mathbf{WH}$ satisfies the condition given in Theorem 2, then each column of $\mathbf{W}$ and row of $\mathbf{H}$ contains at least $K-1$ elements that are equal to 0.*

Corollary 2 has a very interesting interpretation using Donoho and Stodden's analysis, leading to the following proposition.

Proposition 1. *If* $\mathbf{W}$ *and* $\mathbf{H}$ *satisfy the conditions given in Theorem 1 and Corollary 2, and all their non-zero entries are drawn from an i.i.d. continuous distribution with support contained in* $\mathbb{R}_+$*, then there does not exist another simplicial cone* $\mathcal{W}$ *with* K *extreme rays that satisfies either cone*$(\mathbf{W}) \subset \mathcal{W} \subseteq \mathcal{P}_\mathcal{V}$ *or cone*$(\mathbf{S}) \subseteq \mathcal{W} \subset$ *cone*$(\mathbf{W})$ *with probability 1.*

Claiming that there does not exist another simplicial cone $\mathcal{W}$ with K extreme rays that satisfies either cone$(\mathbf{W}) \subset \mathcal{W} \subseteq \mathcal{P}_\mathcal{V}$ or cone$(\mathbf{S}) \subseteq \mathcal{W} \subset$ cone$(\mathbf{W})$ already rules out a lot of possibilities to find another NMF. However, it is still not enough to claim uniqueness, as we have seen in Example 2 when $\omega \geq 0.5$. This is why Corollary 2 is not a sufficient condition, but Theorem 2 is.

The sufficient condition for uniqueness of NMF given in Theorem 2 requires checking whether the conic hull of a set of vectors contains a specific second-order cone. Unfortunately, this turns out being a very hard problem:

Proposition 2. *Checking whether cone*$(\mathbf{W}^T) \supseteq \mathcal{C}$ *is true is NP-complete.*

Nevertheless, we have Theorem 1 and Corollary 2, which are easy to check, and Proposition 1, which rules out a great deal of possibilities for non-uniqueness.

One can check that Donoho and Stodden's "complete factorial sampling" condition implies the condition given in Corollary 2 applied to $\mathbf{W}$. Laurberg *et al.*'s "strongly boundary close" condition is a relaxed version of "complete factorial sampling", but still implies Corollary 2. However, neither "complete factorial sampling" nor "strongly boundary close" imply cone$(\mathbf{W}) \supseteq \mathcal{C}$, which means that their condition on $\mathbf{W}$ is weaker. Their condition on $\mathbf{H}$ is much stronger, however. Indeed, Donoho's "separability" condition requires that cone$(\mathbf{H}) = \mathbb{R}_+^K$, and Laurberg's "sufficiently spread" condition means that the above condition should be satisfied at least asymptotically. This certainly implies not only Corollary 2 but also Theorem 2, and is too strict on $\mathbf{H}$ compared to that on $\mathbf{W}$. The condition given in Theorem 2, on the other hand, imposes a tighter requirement on $\mathbf{W}$ and a looser one on $\mathbf{H}$. In fact $\mathbf{W}$ and $\mathbf{H}$ are treated equally – a symmetric condition, as we would normally expect.

3. SYMMETRIC NMF: ALGORITHM

Suppose there exists a symmetric NMF of $\mathbf{S}$ with $K = \text{rank}(\mathbf{S})$ 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 $\mathbf{BQ} = \mathbf{W}$. Therefore, after obtaining $\mathbf{B}$ via eigen-analysis, we can formulate the recovery of $\mathbf{S}$ as follows:

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

$$\text{subject to} \quad \mathbf{W} \geq 0, \mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I} \tag{2b}$$

The constraint $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ is not convex with respect to $\mathbf{Q}$, suggesting that (2) is a hard problem. We propose updating $\mathbf{W}$ 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 $\mathbf{W} \leftarrow \max(0, \mathbf{BQ})$. When updating $\mathbf{Q}$, it can be shown that the solution is given by the Procrustes projection [12]. This yields our new algorithm for the symmetric NMF given in Fig. 1. Since each step is optimal for $\mathbf{W}$ and $\mathbf{Q}$ respectively, iterations are guaranteed to reduce or maintain the cost function.

```
S = U_s Λ_s U_s^T                 ▷ Reduced eigenvalue decomposition
B ← U_s Λ_s^{1/2}, Q ← I
repeat
    W ← max(0, BQ)
    W^T B = U Σ V^T               ▷ Singular value decomposition
    Q ← V U^T
until convergence
```

Fig. 1: Proposed algorithm for symmetric NMF

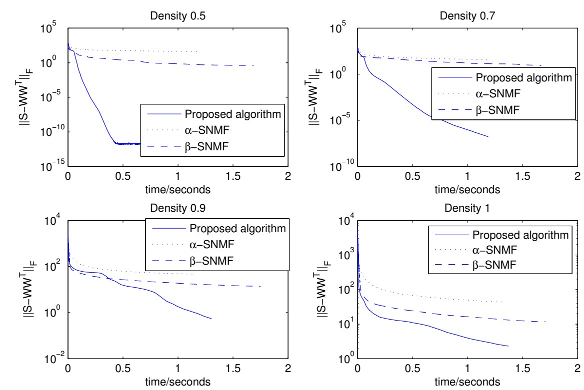

Fig. 2: Convergence of the proposed algorithm comparing to α-SNMF and β-SNMF [8] with $\alpha = \beta = 0.99$.

4. SIMULATIONS

• **Synthetic Data:** The matrix $\mathbf{S}$ is generated by taking $\mathbf{S} = \hat{\mathbf{W}}\hat{\mathbf{W}}^T$, where $\hat{\mathbf{W}}$ is a sparse non-negative matrix, with the non-zero entries drawn from an i.i.d. exponential distribution. We take the size of $\hat{\mathbf{W}}$ to be 100×20. The convergence of a single run of our proposed algorithm is illustrated in Fig. 2, comparing to α-SNMF and β-SNMF provided in [8] on the same $\mathbf{S}$, with $\alpha = \beta = 0.99$, since their experiments showed (and we verified) that this value gives faster convergence. The cost employed in both α-SNMF and β-SNMF is $\|\mathbf{S} - \mathbf{W}\mathbf{W}^T\|_F^2$, which is different from (2a), but we compare all three using $\|\mathbf{S} - \mathbf{W}\mathbf{W}^T\|_F$ on the y-axis as common basis.

In Fig. 2, approximately 50% (resp. 70%) of the entries of $\hat{\mathbf{W}}$ are nonzero for the top left (resp. top right) panel, in which cases our proposed algorithm performs the best, since it converges in less than 1 second, and we can see a nice linear convergence. The bottom two panels show rather dense cases, with approximately 90% nonzeros for the left, and fully dense on the right. In both cases, convergence is much slower compared to the sparse cases, although linear convergence eventually appears. An interesting observation is that our algorithm gets stuck at a saddle point around 0.1-0.3 seconds when density is 0.9. We are currently investigating line search schemes that can overcome such *swamps*. Still, our algorithm clearly outperforms both α-SNMF or β-SNMF, and by a big margin when the true latent factors are sparse.

We also compared the difference between $\mathbf{W}$ and $\hat{\mathbf{W}}$. According to Definition 5, there is no scaling ambiguity if the symmetric

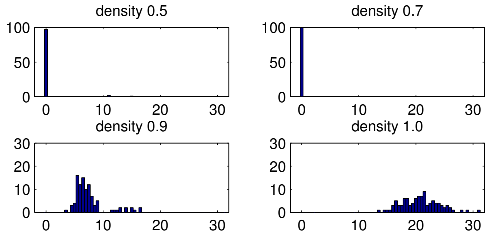

Fig. 3: Statistics of 100 Monte Carlo tests of Latent factor estimation error $\|\mathbf{W} - \hat{\mathbf{W}}\|_F$ for symmetric NMF with different latent density.

NMF is unique, so we only need to worry about the ordering of the columns. We first calculate $\mathbf{W}^T\hat{\mathbf{W}}$, and, for each column, the row index of the largest entry is picked up. If one row index has already been picked up, then following columns will pick the next largest row, until every row index is picked once. A permutation is constructed at the end of this process, and the columns of $\mathbf{W}$ are permuted accordingly. This greedy permutation matching procedure is generally suboptimal, but it is simple and works very well in practice. We run this experiment 100 times for different densities, and the statistics of $\|\mathbf{W} - \hat{\mathbf{W}}\|_F$ are given in Fig. 3. Apparently, if the latent factor is sparse enough, NMF is in most cases unique and our proposed algorithm finds the true latent factors.

• **ARL CTA Co-authorship Data:** We applied our proposed algorithm 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 run for 8 years, 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.* [7] have shown that k-means clustering can be approximated by 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 1 lists the top-10 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

Table 1: Top-10 contributors of the top-3 clusters for $K = 3$

cluster 1	cluster 2	cluster 3
G.B. Giannakis	L. Tong	M.A. Fecko
S. Zhou	A. Swami	S. Samtani
X. Ma	Q. Zhao	M.U. Uyar
P. Xia	B.M. Sadler	I. Hokelek
X. Cai	Y. Chen	J. Zou
T. Wang	M. Dong	J. Zheng
Q. Liu	Y. Sung	M.J. Lee
X. Wang	T. He	T.N. Saadawi
Z. Wang	P. Venkitasubramaniam	U.C. Kozat
A. Cano	Z. Xu	P.T. Conrad

Table 2: Top-10 contributors of the top-3 clusters for $K = 10$

cluster 1	cluster 2	cluster 3
G.B. Giannakis	L. Tong	M.A. Fecko
S. Zhou	A. Swami	S. Samtani
P. Xia	B.M. Sadler	M.U. Uyar
X. Cai	M. Dong	I. Hokelek
Q. Liu	T. He	J. Zou
T. Wang	Y. Sung	J. Zheng
X. Wang	P. Venkitasubramaniam	U.C. Kozat
Z. Wang	S. Adireddy	P.T. Conrad
Y. Yu	G. Mergen	A. Abdelal
A. Cano	A. Anandkumar	J. Sucec

been harder to decipher for someone without direct knowledge of the project. It primarily consists of Telcordia researchers, 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 in more tightly woven pieces).

5. CONCLUSIONS

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 from 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, 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. These were shown to be useful additions to our 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.

6. REFERENCES

[1] 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.

[2] 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.

[3] 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*, John Wiley & Sons Ltd, 2009.

[4] 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)*. MIT Press, 2003, vol. 16, pp. 1141–1148.

[5] H. Laurberg, M.G. Christensen, M.D. Plumbley, L.K. Hansen, and S.H. Jensen, "Theorems on positive data: On the uniqueness of NMF," *Computational Intelligence and Neuroscience*, vol. 2008, doi:10.1155/2008/764206, 2008.

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

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

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

[9] R.T. Rockafellar, *Convex analysis*, vol. 28, Princeton University Press, 1996.

[10] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[11] N. Gillis, "Sparse and unique nonnegative matrix factorization through data preprocessing," *Journal of Machine Learning Research*, vol. 13 (Nov), pp. 3349–3386, 2012.

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