GENERALIZED POLYTOPIC MATRIX FACTORIZATION

Gokcan Tatli and Alper T. Erdogan

EE Department and KUIS AI Lab
Koc University
Istanbul, Turkey

ABSTRACT

Polytopic Matrix Factorization (PMF) is introduced as a flexible data decomposition tool with potential applications in unsupervised learning. PMF assumes a generative model where observations are lossless linear mixtures of some samples drawn from a particular polytope. Assuming that these samples are sufficiently scattered inside the polytope, a determinant maximization based criterion is used to obtain latent polytopic factors from the corresponding observations. This article aims to characterize all eligible polytopic sets that are suitable for the PMF framework. In particular, we show that any polytope whose set of vertices have only permutation and/or sign invariances qualifies for PMF framework. Such a rich set of possibilities enables elastic modeling of independent/dependent latent factors with combination of features such as relatively sparse/antisparse subvectors, mixture of signed/nonnegative components with optionally prescribed domains.

***Index Terms*—** Polytopic Matrix Factorization, Nonnegative Matrix Factorization, Bounded Component Analysis

1. INTRODUCTION

Matrix factorization methods are instrumental tools for various signal processing and machine learning applications, especially those involving unsupervised learning. The goal is to decompose a data matrix into a certain product form, where one or more factors satisfy certain stochastic and/or structural properties reflecting some prior information. We can list Nonnegative Matrix Factorization (NMF) [1, 2, 3, 4, 5], Sparse Component Analysis (SCA) [6, 7], Independent Component Analysis (ICA) [8, 9, 10] and Bounded Component Analysis (BCA) [11, 12, 13, 14, 15] as some examples.

In [16], as an extension of NMF and BCA frameworks, Polytopic Matrix Factorization (PMF) was introduced as a flexible matrix decomposition approach enabling the extraction of sparse/antisparse and signed/nonnegative components as latent factors. The same reference also shows that if the original factors are sufficiently scattered in the presumed polytopic set, we can perfectly recover them from their linear and lossless mixtures through a determinant maximization based criterion. The idea of sufficient scatteredness of factors was initially put forward for NMF [17, 18], where the proposed criterion involved the coverage and tangency conditions of the some cones defined corresponding to the nonnegative orthant. The reference [16] proposes the use of Maximum Volume Inscribed Ellipsoid (MVIE) [19] of the polytopic set as a reference for sufficient scattering. We note that the use of MVIE for simplex structured matrix factorization algorithm construction was independently proposed in [20]. In [16], it is shown that if the convex hull of the original polytopic set vectors cover the MVIE of the polytope with some restricted tangency conditions, then their perfect recovery (with some potential sign and permutation ambiguities) can be achieved (see Definition 3.4 in Section 3). Figure 1 illustrates the 4 polytopes put forward as examples in [16], corresponding to sparse, antisparse, nonnegative sparse and nonnegative antisparse cases with sufficiently scattered samples. The same figure also includes the corresponding MVIEs which are used to determine the sufficient scattering condition.

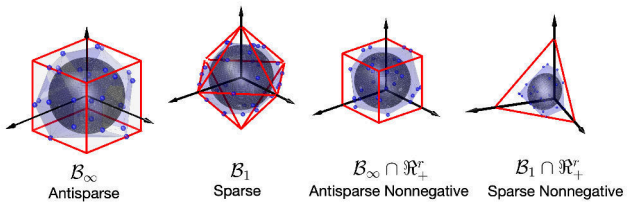


Fig. 1: Polytope examples from [16] : Vertices of polytopes (red), MVIE (black), sufficiently scattered samples from polytope (blue dots), convex hull of samples (blue (transparent)).

In this article, we explore the conditions to generalize the set of polytopes for which the identifiability condition extends beyond the four examples provided in [16], and illustrated in Figure 1. Specifically we derive the necessary and sufficient

This work is partially supported by an AI Fellowship provided by the KUIS AI Lab.

condition for a polytope to be identifiable in the sense that sufficiently scattered samples drawn from such polytopes can be perfectly recovered from their linear mixtures via determinant maximization approach.

Following is the outline of the article: Section 2 provides the setup for the PMF. Section 3 summarizes relevant results from [16] on PMF and provides basic definitions to be used in the generalized PMF treatment of Section 4. Section 4 is the main part where we provide the theorems that characterize all eligible polytopes for the PMF framework with respect to the determinant maximization criterion. Section 5 provides an example for the generalized PMF. Finally, Section 6 is the conclusion.

Table 1 outlines the basic notation used in the article:

Notation	Meaning
$\mathbf{S}_{j,:}$ ($\mathbf{S}_{:,j}$)	j^{th} row (column) of $\mathbf{S}$
$\langle \mathbf{x}, \mathbf{y} \rangle$	Euclidean inner product defined as $\mathbf{y}^T\mathbf{x}$
$\mathcal{B}_p$	unit ℓ_p-norm ball defined as $\{\mathbf{x} \mid \|\mathbf{x}\|_p \leq 1\}$
$\mathcal{S}^{*,\mathbf{d}}$	the polar of the set $\mathcal{S}$ with respect to the point $\mathbf{d}$
conv($\mathbf{H}$)	the convex hull of the columns of $\mathbf{H}$
ext($\mathcal{P}$)	the set of extreme points of the polytope $\mathcal{P}$
$\mathcal{E}_{\mathcal{P}}$	maximum volume inscribed ellipsoid of the polytope $\mathcal{P}$
bd($\mathcal{S}$)	the boundary of the set $\mathcal{S}$
$\mathbf{A}(S)$	image of the set S under transformation $\mathbf{A}$
$\Re^r_+$	the nonnegative orthant

Table 1: Notation

2. PMF DATA MODEL

For the PMF problem, we assume the generative model in [16]: The observation matrix $\mathbf{Y} \in \Re^{M\times N}$ is generated by

$$\mathbf{Y} = \mathbf{H}_g\mathbf{S}_g$$

where

- $\mathbf{H}_g \in \Re^{M\times r}$ is the ground truth for the left factor matrix which is assumed to be full rank and $M \geq r$.
- $\mathbf{S}_g \in \Re^{r\times N}$ is the ground truth for the right factor matrix. The underlying assumption of the PMF framework can be written as

  $$\mathbf{S}_{:,j} \in \mathcal{P},\ j = 1, \ldots, N,$$

  where $\mathcal{P}$ is a convex polytope. The proper choice of $\mathcal{P}$ is the central problem of the current article.

The goal is to obtain (the estimates of) $\mathbf{H}_g$ and $\mathbf{S}_g$ from the observation matrix $\mathbf{Y}$. In the article, "PMF Data Model" will refer to an instance of the triplet $(\mathcal{P}, \mathbf{S}_g, \mathbf{H}_g)$, i.e., a convex polytope $\mathcal{P}$, the samples drawn from this polytope in the columns of $\mathbf{S}_g$ and the (over)determined lossless mixing matrix $\mathbf{H}_g$.

3. FUNDAMENTAL DEFINITIONS AND PROPERTIES

In this section, we provide some fundamental definitions which follow from and extend the ones in [16], together with some useful facts, which will be used in next section to obtain a generalization of polytope choices.

As proposed in [16], the volume (determinant) maximization (for the columns of the factor $\mathbf{S}$) approach is the main criterion for obtaining desired PMF factorization outlined in the previous section:

Definition 3.1. *Det-Max Criterion: Given the PMF data model in Section 2, we define the following optimization problem*

$$\begin{aligned} &\underset{\mathbf{H}\in\Re^{M\times r},\mathbf{S}\in\Re^{r\times N}}{\textit{maximize}} && \det(\mathbf{S}\mathbf{S}^T) \quad \textit{Det-Max Problem} \\ &\textit{subject to} && \mathbf{Y} = \mathbf{H}\mathbf{S}, \\ & && \mathbf{S}_{:,j} \in \mathcal{P},\ j = 1, \ldots, N. \end{aligned}$$

as a criterion for obtaining original factors from the data matrix $\mathbf{Y}$. We refer to using above Det-Max Problem to finding solution to the matrix factorization problem in Section 2 as Det-Max Criterion.

We refer to the PMF problem as identifiable, if the optimal solution of the determinant maximization problem is equivalent to the desired factors with some permutation and sign ambiguities:

Definition 3.2. *Det-Max Identifiable PMF Data Model: Given the PMF data model in Section 2, if all the optimal solutions $\mathbf{S}$ of the Det-Max Problem satisfy*

$$\mathbf{S} = \mathbf{\Pi}\mathbf{D}\mathbf{S}_g$$

where $\mathbf{\Pi}$ is a permutation matrix and $\mathbf{D}$ is a full rank diagonal matrix, then the corresponding PMF data model is called to be identifiable with respect to the determinant maximization criterion.

The identifiability of the PMF Data Model is determined by the choice of $\mathcal{P}$ and the scattering of the columns of $\mathbf{S}_g$ in $\mathcal{P}$. The following provides a strong condition for scattering in $\mathcal{P}$:

Definition 3.3. *Strongly Scattered PMF Data Model: A PMF Data Model is called strongly scattered if the columns of $\mathbf{S}_g$ contains ext($\mathcal{P}$).*

Such a strong scattering assumption for the $\mathbf{S}_g$ in the generative model is typically too stringent for practical applications. Instead the following relaxed version was introduced which is proven to be useful in obtaining identifiability results [16]:

Definition 3.4. *Sufficiently Scattered PMF Data Model: A PMF Data Model is called Sufficiently Scattered if*

- $\mathcal{P} \supseteq conv(\mathbf{S}_g) \supset \mathcal{E}_\mathcal{P}$, *and,*
- $conv(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}} \cap bd(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) = ext(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})$.

where $\mathbf{g}_\mathcal{P}$ *is the center of* $\mathcal{E}_\mathcal{P}$.

The main scope of this article is investigation of the choices of $\mathcal{P}$ that lead to identifiable PMF Data Models. As shown in the next section, the characterization of such polytopes will be based on permutation-and/or-sign-only invariance property of their extremum points. Hence, we provide the following definition:

Definition 3.5. *Permutation-and/or-Sign-Only Invariant Set: Given a set E, we refer to it as Permutation-and/or-Sign-Only Invariant set, if and only if $\mathbf{A}(E) = E$ holds for only some $\mathbf{A} = \mathbf{D\Pi}$, where $\mathbf{\Pi}$ is a permutation matrix and $\mathbf{D}$ is a full rank diagonal matrix with diagonal entries in $\{-1, 1\}$.*

We now list some fundamental facts regarding the linear transformation of ellipsoids and polar sets, which will be useful on understanding of theorems in the next section:

Lemma 1. *On Linear Transformations of Ellipsoids: Given a full rank square transformation matrix $\mathbf{A}$ with real entries and an ellipsoid $\mathcal{E}$. $\mathbf{A}(\mathcal{E}) = \mathcal{E}$ implies $\mathbf{A}(bd(\mathcal{E})) = bd(\mathcal{E})$, and $\mathbf{A}(\mathcal{E}) = \mathcal{E}$ implies $\mathbf{Ag} = \mathbf{g}$, where $\mathbf{g}$ is the center of ellipsoid $\mathcal{E}$.*

Lemma 2. *Polar of the Transformed Set: Polar of a set S with respect to the point $\mathbf{d}$ is defined as*

$$S^{*,\mathbf{d}} = \{\mathbf{x} \in \Re^n | \langle \mathbf{x}, \mathbf{y} - \mathbf{d} \rangle \leq 1 \; \forall \mathbf{y} \in S\}$$

When $\mathbf{d} = \mathbf{0}$, we simplify the notation as S^*. In the light of this definition, we can characterize the polar of the transformed set $\mathbf{A}(S)$ in the following way,

$$(\mathbf{A}(S))^{*,\mathbf{d}} \quad = \quad \{\mathbf{x} \in \Re^n | \langle \mathbf{x}, \mathbf{Ay} - \mathbf{d} \rangle \leq 1 \; \forall \mathbf{y} \in S\}$$

which corresponds to the set $\mathbf{A}^{-T}(S^{*,\mathbf{d}})$, when $\mathbf{Ad} = \mathbf{d}$.

4. GENERALIZED PMF

In this section, based on the definitions in the previous section, we derive the characterization of polytopes suitable for PMF with respect to *Det-Max Criterion*. We start with the strongly scattered PMF models with the following theorem:

Theorem 1. *Characterization of Polytopes that are Det-Max Identifiable with respect to the Strongly Scattered PMF Data Model : All strongly scattered PMF data models corresponding to a non-degenerate polytope $\mathcal{P}$ are identifiable if and only if ext$(\mathcal{P})$ is a Permutation-and/or-Sign-Only Invariant set.*

Proof. "If" part: according to the setting in *Det-Max Criterion*, any feasible point $\mathbf{S}$ has the same row space as $\mathbf{Y}$, and therefore, $\mathbf{S}_g$, hence satisfy

$$\mathbf{S} = \mathbf{AS}_g,$$

for some full rank $\mathbf{A} \in \Re^{r \times r}$. Therefore, the optimal selection of $\mathbf{S}$ is equivalent to the optimal selection of $\mathbf{A}$. Using this equivalence, the objective function can be rewritten as $\det(\mathbf{SS}^T) = |\det(\mathbf{A})|^2 \det(\mathbf{S}_g\mathbf{S}_g^T)$. $\det(\mathbf{S}_g\mathbf{S}_g^T)$ term is constant, which simplifies optimization objective to $|\det(\mathbf{A})|$. Since identity matrix, $\mathbf{I}$, is clearly a feasible solution, its determinant corresponds to a lower bound. Therefore, for any optimal solution $\mathbf{A}_*$

$$|\text{det}(\mathbf{A}_*)| \geq 1. \tag{1}$$

Then, considering feasible region,

$$\mathbf{A}(\text{conv}(\mathbf{S}_g)) = \mathbf{A}(\mathcal{P}) \subseteq \mathcal{P}$$

implies the following upper bound

$$|\det(\mathbf{A})| \leq 1. \tag{2}$$

Now, (1) and (2) imply $|\text{det}(\mathbf{A}_*)| = 1$ and $\mathbf{A}_*(\mathcal{P}) = \mathcal{P}$ for optimal $\mathbf{A}_*$. Since $\mathcal{P}$ is a convex polytope, this is equivalent to

$$\mathbf{A}_*(\text{ext}(\mathcal{P})) = \text{ext}(\mathcal{P}). \tag{3}$$

Therefore, if ext$(\mathcal{P})$ is *Permutation-and/or-Sign-Only Invariant*, then the PMF Data Model is Det-Max identifiable.

"Only if" part: Let $\mathbf{V}_\mathcal{P} \in \Re^{r \times K}$ represent a matrix containing all K vertices of $\mathcal{P}$ in its columns. Then $\mathbf{S}_g = \mathbf{V}_\mathcal{P}\mathbf{L}_g$ for some $\mathbf{L}_g \in \Re^{K \times K}$ is a matrix with convex combination weights in its columns. If there exists an $\mathbf{A}_*$ matrix for which $\mathbf{A}_*(\text{ext}(\mathcal{P})) = \text{ext}(\mathcal{P})$ holds then $\mathbf{A}_*\mathbf{V}_\mathcal{P} = \mathbf{V}_\mathcal{P}\mathbf{\Gamma}$ for some permutation matrix $\mathbf{\Gamma}$. The corresponding $\mathbf{S} = \mathbf{A}_*\mathbf{S}_g = \mathbf{A}_*\mathbf{V}_\mathcal{P}\mathbf{L}_g = \mathbf{V}_\mathcal{P}\mathbf{\Gamma}\mathbf{L}_g = \mathbf{V}_\mathcal{P}\mathbf{L}'_g$ where $\mathbf{L}'_g = \mathbf{\Gamma}\mathbf{L}_g$ is another matrix with convex combination weights in its columns, obtained from $\mathbf{L}_g$ by shuffling in its rows. Therefore, the corresponding $\mathbf{S}$ is feasible. Full rank condition on $\mathbf{V}_\mathcal{P}$ from non-degeneracy of $\mathcal{P}$ and $\mathbf{A}_*\mathbf{V}_\mathcal{P} = \mathbf{V}_\mathcal{P}\mathbf{\Gamma}$ imply $|\det(\mathbf{A}_*)| = 1$. Therefore, all matrices satisfying (3) are optimal Det-Max solutions for strongly scattered PMF Data Models. Therefore, if there exists an $\mathbf{A}_*$ that satisfies (3) but not equal to the product of a diagonal sign and a permutation matrix, then all strongly scattered PMF Data Models corresponding to $\mathcal{P}$ will be unidentifiable. □

The next theorem shows that the use of the weaker scattering notion leads to the same condition of polytope eligibility for PMF Data Model identifiability:

Theorem 2. *Characterization of Polytopes that are Det-Max Identifiable with respect to the Sufficiently Scattered PMF Data Model : All sufficiently scattered PMF data models corresponding to a non-degenerate polytope $\mathcal{P}$ are identifiable if and only if ext$(\mathcal{P})$ is a Permutation-and/or-Sign-Only Invariant set.*

Proof. We will first show the "if" part: Let $(\mathcal{P}, \mathbf{S}_g, \mathbf{H}_g)$ be any sufficiently scattered PMF model for $\mathcal{P}$. Let $\mathbf{A}_*$ be any optimal solution of the corresponding Det-Max problem. We will show that $\mathbf{A}_*(\text{ext}(\mathcal{P})) = \text{ext}(\mathcal{P})$. We start by noting that the constraint part of Det-Max problem imposes

$$\text{conv}(\mathbf{A}_*\mathbf{S}_g) \subseteq \mathcal{P}, \tag{4}$$

and from the feasibility of the identity matrix $\mathbf{A} = \mathbf{I}$, the optimal solution satisfies $|\det(\mathbf{A}_*)| \geq \det(\mathbf{I}) = 1$. Then, due to the sufficient scattering of $\mathbf{S}_g$, the constraints in Definition 3.4 hold. The first constraint along with (4) imply that $\mathbf{A}_*(\mathcal{E}_\mathcal{P}) \subset \mathcal{P}$, i.e., an ellipsoid in $\mathcal{P}$ with

$$\text{vol}(\mathbf{A}_*(\mathcal{E}_\mathcal{P})) = |\det(\mathbf{A}_*)|\text{vol}(\mathcal{E}_P) \geq \text{vol}(\mathcal{E}_\mathcal{P}),$$

using the lower bound $|\det(\mathbf{A}_*)| \geq 1$. Due to the uniqueness of MVIE for $\mathcal{P}$ we deduce that $|\det(\mathbf{A}_*)| = 1$ and

$$\mathbf{A}_*(\mathcal{E}_\mathcal{P}) = \mathcal{E}_\mathcal{P}. \tag{5}$$

In other words, sufficient scattering condition places an upper bound on $|\det(\mathbf{A}_*)|$ and restricts optimal $\mathbf{A}_*$ to map $\mathcal{E}_\mathcal{P}$ to itself. Consequently, according to Lemma 2, (5) corresponds to

$$\mathbf{A}_*^{-T}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) = \mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}} \tag{6}$$

where $\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}$ is also an ellipsoid. Therefore, (6) implies

$$\mathbf{A}_*^{-T}(\text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})) = \text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}). \tag{7}$$

By the reversing property of polar sets given in Appendix of [16], (4) corresponds to $\mathcal{P}^{*,\mathbf{g}_\mathcal{P}} \subseteq \mathbf{A}_*^{-T}(\text{conv}(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}})$, which implies

$$\mathbf{A}_*^{-T}(\text{conv}(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}}) \supset \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}). \tag{8}$$

From the second constraint of sufficient scattering, we obtain $\text{conv}(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}} \supset \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})$, $\text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) \supset \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})$ and

$$\mathbf{A}^{-T}(\text{conv}(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}}) \cap \text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) = \mathbf{A}^{-T}(\text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})), \tag{9}$$

where we also inserted (7). In addition, (8) implies

$$\mathbf{A}^{-T}(\text{conv}(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}}) \cap \text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) \supseteq \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}). \tag{10}$$

Combining (9) and (10), we obtain

$$\mathbf{A}_*^{-T}(\text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})) \supseteq \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}).$$

Since the sets on both sides have the same cardinality, it boils down to the equality

$$\mathbf{A}_*^{-T}(\text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})) = \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}),$$

which corresponds to $\mathbf{A}_*^{-T}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) = \mathcal{P}^{*,\mathbf{g}_\mathcal{P}}$. This is equivalent to $\mathbf{A}_*(\mathcal{P}) = \mathcal{P}$, which corresponds to (3) by the convexity of $\mathcal{P}$. Therefore, the condition that $\text{ext}(\mathcal{P})$ is a *Permutation-and/or-Sign-Only Invariant Set* leads to the restriction that $\mathbf{A}_* = \mathbf{D}\mathbf{\Pi}$, with diagonal sign matrix $\mathbf{D}$ and the permutation matrix $\mathbf{\Pi}$, hence guaranteeing identifiability.

For the "only if" part: follows the same arguments as the "only if" part of Theorem 1. □

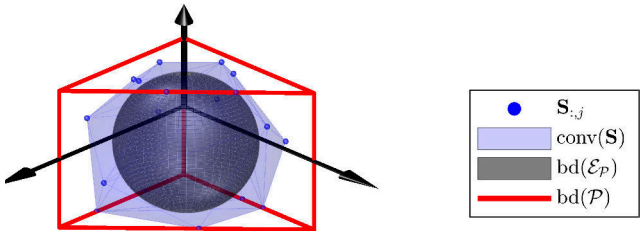


Fig. 2: An example polytope corresponding to one signed and two nonnegative-mutually sparse sources.

5. EXAMPLE

As asserted by Theorem 2 in the previous section, polytopes are only constrained to obey the restricted symmetry condition, i.e. only permutation and/or sign only invariance. This result offers a rich set of possibilities for selecting polytopes enabling high flexibility about shaping conditions on the desired factors.

In order to illustrate this point and provide the corresponding geometrical picture, we consider an example with three components where $s_3 \in [-1, 1]$ is signed and $s_1, s_2 \in [0, 1]$ are non-negative. Furthermore, we require $\| \begin{bmatrix} s_1 & s_2 \end{bmatrix} \|_1 \leq 1$, i.e., a sparsity enforcing condition on this subvector. The corresponding polytope, its MVIE and some sufficiently scattered sample points (with their convex hull) are shown on the left of Figure 2. Such sample points can be obtained from their lossless linear mixtures as the solution of the Det-Max optimization framework, through, for example, the algorithm proposed in [16].

6. CONCLUSION

In Polytopic Matrix Factorization, the choice of polytopes will determine the shaping of the latent vectors based on some prior information and/or desired properties attributed to them. Corresponding component feature choices may include a blend of individual properties such as being signed vs, nonnegative and/or having certain domain/range restrictions, as well as, defining some relational properties among different subsets of components such as ℓ_1-norm constraints promoting sparsity.

The results obtained through this article asserts that an infinite set of polytopes, with certain permutation/sign invariance constraints, are eligible choices for the PMF framework. This flexibility in the selection of polytopes can be utilized for shaping latent factors/features according to the goals of the data decomposition task. The previous section illustrates this fact through a simple example where the components with different features such as signed/nonnegative and partial mutual sparseness are defined.

7. REFERENCES

[1] Paris Smaragdis and Judith C Brown, "Non-negative matrix factorization for polyphonic music transcription," in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No. 03TH8684)*. IEEE, 2003, pp. 177–180.

[2] Wei Xu, Xin Liu, and Yihong Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, 2003, pp. 267–273.

[3] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[4] Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2011.

[5] Alexey Ozerov and Cédric Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 550–563, 2009.

[6] Pando Georgiev, Fabian Theis, and Andrzej Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 992–996, 2005.

[7] Michael Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer Science & Business Media, 2010.

[8] Pierre Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[9] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, "Independent component analysis, adaptive and learning systems for signal processing, communications, and control," *John Wiley & Sons, Inc*, vol. 1, pp. 11–14, 2001.

[10] Anthony J Bell and Terrence J Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[11] Sergio Cruces, "Bounded component analysis of linear mixtures: A criterion of minimum convex perimeter," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2141–2154, 2010.

[12] Alper T Erdogan, "A class of bounded component analysis algorithms for the separation of both independent and dependent sources," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5730–5743, 2013.

[13] Eren Babatas and Alper T Erdogan, "An algorithmic framework for sparse bounded component analysis," *IEEE Transactions on Signal Processing*, vol. 66, no. 19, pp. 5194–5205, 2018.

[14] Huseyin A Inan and Alper T Erdogan, "A convolutive bounded component analysis framework for potentially nonstationary independent and/or dependent sources," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 18–30, 2014.

[15] Eren Babatas and Alper T Erdogan, "Time and frequency based sparse bounded component analysis algorithms for convolutive mixtures," *Signal Processing*, p. 107590, 2020.

[16] Gokcan Tatli and Alper T. Erdogan, "Polytopic matrix factorization: Determinant maximization based criterion and identifiability," *Submitted to IEEE Transactions on Signal Processing*, 2020.

[17] Xiao Fu, Kejun Huang, and Nicholas D Sidiropoulos, "On identifiability of nonnegative matrix factorization," *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 328–332, 2018.

[18] Xiao Fu, Kejun Huang, Nicholas D Sidiropoulos, and Wing-Kin Ma, "Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications.," *IEEE Signal Process. Mag.*, vol. 36, no. 2, pp. 59–80, 2019.

[19] Stephen P Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[20] Chia-Hsiang Lin, Ruiyuan Wu, Wing-Kin Ma, Chong-Yung Chi, and Yue Wang, "Maximum volume inscribed ellipsoid: A new simplex-structured matrix factorization framework via facet enumeration and convex optimization," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 1651–1679, 2018.