Identifiability of Polytopic Matrix Factorization

Olivier Vu Thanh Nicolas Gillis
University of Mons
Rue de Houdain 9, 7000 Belgium
{olivier.vuthanh, nicolas.gillis}@umons.ac.be

***Abstract*—Polytopic matrix factorization (PMF) decomposes a given matrix as the product of two factors where the rows of the first factor belong to a given convex polytope and the columns of the second factor belong to another given convex polytope. In this paper we show that if the polytopes have certain invariant properties, and that if the rows of the first factor and the columns of the second factor are sufficiently scattered within their corresponding polytope, then this PMF is identifiable, that is, the factors are unique up to a signed permutation. The PMF framework is quite general, as it recovers other known structured matrix factorization models, and is highly customizable depending on the application. Hence, our result provides sufficient conditions that guarantee the identifiability of a large class of structured matrix factorization models.**

***Index Terms*—identifiability, polytopic matrix factorization, nonnegative matrix factorization**

I. Introduction

Low-rank matrix approximations allow one to obtain compressed representations of data while extracting automatically important features. Let $X \in \mathbb{R}^{m\times n}$ be a data matrix where each column is a data point, a standard low-rank matrix approximation decomposes X as follows: $X \approx WH$ where $W \in \mathbb{R}^{m\times r}$, $H \in \mathbb{R}^{r\times n}$, and $r < \min(m,n)$ is the factorization rank. If we interpret the columns of W as a set of basis elements, then the matrix H contains the weights necessary to reconstruct the columns of X through linear combinations of these basis elements. This simple, yet powerful, data representation technique is applied in many domains, e.g., facial feature extraction [1], document clustering [2], blind source separation [3], [4], community detection [5], gene expression analysis [6], and recommender systems [7]. Depending on the application and the goal at hand (e.g., clustering, denoising, feature extraction), one may impose different structures/constraints on the factors W and/or H.

For example, nonnegative matrix factorization (NMF) [1] requires every element of the two factors W and H to be nonnegative. In hyperspectral unmixing (HU), the data would be a vectorized hyperspectral cube: each column of X represents a pixel, and each row of X represents a spectral band. Since the sensed reflectance is nonnegative, it is meaningful to impose the extracted basis elements to be also nonnegative. Likewise, it is physically meaningful to impose the mixtures of these basis elements in a pixel to be nonnegative since linear combinations of spectral signatures should only be additive. Hence, NMF is quite relevant when it comes to HU to recover the underlying spectral signatures and the abundances of the materials in each pixel [3].

In many applications, it is important that the factorization is essentially unique, such as for HU since the goal is to find the true materials present in the hyperspectral image. For example, essential uniqueness for NMF, also known as *identifiability*, is achieved when given an NMF (W,H) of $X = WH$, the only matrices $\tilde{W} \in \mathbb{R}^{m\times r}_+$ and $\tilde{H} \in \mathbb{R}^{r\times n}_+$ such that $X = \tilde{W}\tilde{H}$ are of the form $\tilde{W} = WD\Pi$ and $\tilde{H} = \Pi^\top D^{-1}H$, where Π is a permutation matrix and D is a diagonal matrix whose diagonal is strictly positive. NMF is not essentially unique in general. However, it has been proven to be identifiable under the sufficiently scattered conditions (SSC):

Definition 1 (SSC)**.** *The matrix $H \in \mathbb{R}^{r\times n}_+$ is sufficiently scattered if the following two conditions are satisfied:*

[SSC1] $\mathcal{C} = \{x \in \mathbb{R}^r_+ \mid e^\top x \geq \sqrt{r-1}\|x\|_2\} \subseteq \operatorname{cone}(H)$.

[SSC2] There does not exist any orthogonal matrix Q such that $\operatorname{cone}(H) \subseteq \operatorname{cone}(Q)$, except for permutation matrices.

Theorem 1. *[8, Th. 4] If $W^\top \in \mathbb{R}^{r\times m}_+$ and $H \in \mathbb{R}^{r\times n}_+$ are sufficiently scattered then the NMF (W,H) of $X = WH$ is essentially unique.*

A geometric interpretation of these sufficient conditions is the following: while making sure that $X = WH$ and that W is nonnegative, it is not possible to decrease the "volume" of the cone of $W^\top$ without making the cone of H get out of the nonnegative orthant, and vice versa; see Section IV for details.

In this paper, we focus on the identifiability of polytopic matrix factorization (PMF). With NMF, the feasible domain is the nonnegative orthant. With PMF, the feasible domains are convex polytopes: the columns of $W^\top$ and H belong to the polytopes $\mathcal{P}_W$ and $\mathcal{P}_H$, respectively. A variant of PMF has already been studied in [9], [10] where the authors proposed a structured matrix factorization where: (i) the matrix W is unconstrained, (ii) the columns of H belong to a convex polytope, and (iii) the goal is to find a factorization maximizing the volume of the convex hull of the columns of H. This model, proposed in [9], [10], is also referred to as PMF, although it would have been more appropriate to refer to it as maximum-volume PMF. In fact, their proposed model could be viewed as a polytopic variant of minimum-volume semi-NMF [2], while our proposed model would rather be a polytopic variant of NMF.

This work was supported by the Fonds de la Recherche Scientifique - FNRS (F.R.S.-FNRS) under PDR no T.0097.22 and under a FRIA PhD grant for OVT, and by the Francqui Foundation.

Contribution and outline. Inspired by the identifiability conditions in [9] and similarly to Theorem 1, our main contribution in this paper is to show that if the convex hull of $W^\top$ and H are sufficiently scattered within their respective polytope, then the corresponding PMF is identifiable (Theorem 2).

In Section II we introduce PMF. Section III provides important definitions and properties. In Section IV we prove our main result. Section V presents known structured matrix factorization that are special cases of PMF, and how our theoretical finding relates to previous results.

II. Polytopic Matrix Factorization

In this paper, we consider convex polytopes, that is, bounded polyhedra. A convex polytope $\mathcal{P}$ can always be expressed in V-form, through a convex combination of its vertices:

$$\mathcal{P} = \text{conv}(V) = \{x \mid x = Vh, h \geq 0, \sum_i h_i = 1\}, \quad (1)$$

where the columns of V are the vertices, or the extremum points, of $\mathcal{P}$. We can now define PMF. Given a data matrix $X \in \mathbb{R}^{m \times n}$ and r, PMF computes W and H such that

$$\begin{aligned} X = WH \text{ s.t. } W(i,:) \in \mathcal{P}_W & \text{ for all } i \in 1,\dots,m, \\ H(:,j) \in \mathcal{P}_H & \text{ for all } j \in 1,\dots,n, \end{aligned} \quad (2)$$

where $W \in \mathbb{R}^{m \times r}$ is the basis matrix, $H \in \mathbb{R}^{r \times n}$ is the coefficient matrix, $\mathcal{P}_W$ and $\mathcal{P}_H$ are convex polytopes that respectively constrain the rows of W and the columns of H. This PMF is referred to as the quadruple $(W, H, \mathcal{P}_W, \mathcal{P}_H)$. This framework is quite general: it offers infinite varieties of structured matrix factorizations that promote different behaviors in the latent space, depending on the choice of $\mathcal{P}_W$ and $\mathcal{P}_H$. As we will show in Section V, PMF recovers factorizations that have been been studied in the literature.

III. Notation, Definitions and Properties

In this section, we give our notation, and provide important definitions and properties that are needed to achieve our main result on the identifiability of PMF (Theorem 2 in Section IV).

Notation. Here is some of our notation:

e	vector of all ones of appropriate dimension
$\mathcal{X}^{*,g}$	polar of the set $\mathcal{X} \subset \mathbb{R}^r$ with respect to g, that is, $\{x \in \mathbb{R}^r \mid \langle x, y - g\rangle \geq 0, \text{for all } y \in \mathcal{X}\}$
Δ^r	probability simplex $\{x \in \mathbb{R}^r \mid x \geq 0, e^\top x = 1\}$
$A^{-\top}$	inverse of the transpose of the square matrix A
$\text{cone}(A)$	conical hull of the columns of A
$\text{conv}(A)$	convex hull of the columns of A
$\text{ext}(\mathcal{X})$	set of extreme points of the set $\mathcal{X}$
$\text{bd}(\mathcal{X})$	boundary of the set $\mathcal{X}$

Identifiability. Let us clarify what is meant by identifiability. A PMF $(W, H, \mathcal{P}_W, \mathcal{P}_H)$ is identifiable if for any other PMF $(W_*, H_*, \mathcal{P}_W, \mathcal{P}_H)$ of X, there exist a permutation matrix Π and a diagonal matrix D with diagonal values in $\{-1, 1\}$ such that $W_* = W\Pi^\top D^{-1}$ and $H_* = D\Pi H$. We will refer to a matrix of the form $D\Pi$ as a signed permutation. As opposed to NMF, essential uniqueness of PMF is stronger as it only allows a sign ambiguity, while NMF allows a scaling ambiguity.

Minimum-volume ellipsoid and sufficient scatteredness. Our sufficient scatteredness conditions that guarantee identifiability heavily rely on the notion of ellipsoids. Given a center, $\bar{x} \in \mathbb{R}^r$, and a positive definite matrix E, an ellipsoid is defined as $\mathcal{E}(E, \bar{x}) := \{x \in \mathbb{R}^r \mid (x - \bar{x})^\top E (x - \bar{x}) \leq r\}$. Its volume is given by $\text{vol}(\mathcal{E}(E, \bar{x})) = \frac{r^{r/2}\Omega_r}{\sqrt{\det(E)}}$ where Ω_r is the volume of a ball of radius 1 in $\mathbb{R}^r$. The axis of the ellipsoid are given by the eigenvectors of E, and their length is inversely proportional to the square root of corresponding eigenvalues; see, e.g., [11]. Given an ellipsoid $\mathcal{E}(E, \bar{x})$ and an invertible matrix Q, it can be shown that $Q(\mathcal{E}(E, \bar{x})) = \{Qx \mid x \in \mathcal{E}\} = \mathcal{E}(Q^{-\top} E Q^{-1}, \bar{y})$ where $\bar{y} = Q\bar{x}$, and hence the volume of $Q\mathcal{E}$ equals the volume of $\mathcal{E}$ times $|\det(Q)|$. This will be useful in our identifiability proof.

The MVIE of a polytope $\mathcal{P}$, denoted $\mathcal{E}_\mathcal{P}$, is defined as the ellipsoid $\mathcal{E}_\mathcal{P} \subset \mathcal{P}$ with maximum volume $\text{vol}(\mathcal{E}(E, \bar{x}))$, that is, for which $\det(E)$ is maximized. It can be computed by solving a convex semidefinite program; see, e.g., [12, Chap. 8.4.2]. A convex set is said to be sufficiently scattered relative to a polytope when it is contained in that polytope while containing the MVIE of this polytope [9].

Our identifiability result will be based on the following sufficient scatteredness condition:

Definition 2 (Sufficiently Scattered Factor [9])**.** *The matrix $H \in \mathbb{R}^{r \times n}$ is called a sufficiently scattered factor (SSF) corresponding to $\mathcal{P}$ if*

[PMF.SSC1] $\mathcal{P} \supseteq \text{conv}(H) \supset \mathcal{E}_\mathcal{P}$, and

[PMF.SSC2] $\text{conv}(H)^{*,g_\mathcal{P}} \cap \text{bd}(\mathcal{E}_\mathcal{P}^{*,g_\mathcal{P}}) = \text{ext}(\mathcal{P}^{*,g_\mathcal{P}})$,

where $\mathcal{E}_\mathcal{P}$ is the MVIE of $\mathcal{P}$ centered at $g_\mathcal{P}$.

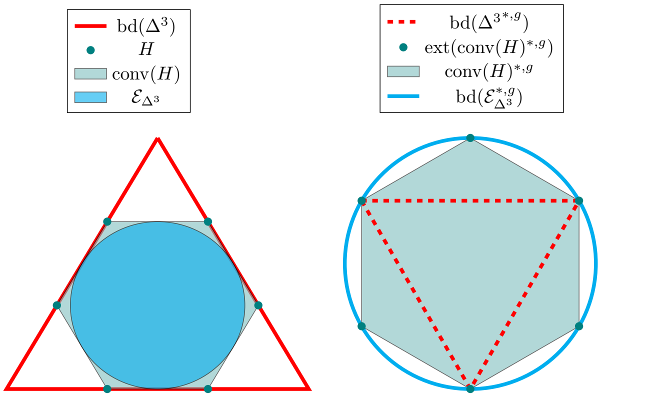
The idea behind the condition [PMF.SSC1] is similar to [SSC1] in Theorem 1, as both conditions ensure that the considered factor is sufficiently scattered within its feasible set. The MVIE acts like the second order cone $\mathcal{C}$ in [SSC1] which is the largest cone contained in the nonnegative orthant. Here, [PMF.SSC1] ensures that the convex hull of a factor H is contained in the polytope $\mathcal{P}$ and contains the MVIE of $\mathcal{P}$. The second condition [PMF.SSC2] makes sure that the MVIE is not contained too tightly. Let us illustrate why [PMF.SSC2] is important with the PMF $(H^\top, H, \Delta^3, \Delta^3)$ using Example 3 from [13], see also [8, Example 2]:

$$H = \frac{1}{3}\begin{pmatrix} 1 & 2 & 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 2 & 2 & 1 \end{pmatrix}. \quad (3)$$

As it can be seen on fig. 1a, H satisfies [PMF.SSC1]. However, fig. 1b exposes why H does not satisfy [PMF.SSC2] and it turns out that the PMF $(H^\top, H, \Delta^3, \Delta^3)$ is not identifiable:

$$Q = \frac{1}{3}\begin{pmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{pmatrix}$$

provides another PMF, $(H^\top Q^\top, QH, \Delta^3, \Delta^3)$, while QH is not a signed permutation of the rows of H.



(a) Visualization of why H satisfies [PMF.SSC1].

(b) Visualization of why H does not satisfy [PMF.SSC2].

Fig. 1: A small example, with H from eq. (3) and $g = \begin{pmatrix}1/3 & 1/3 & 1/3\end{pmatrix}^\top$, showing how [PMF.SSC1] can be satisfied without [PMF.SSC2] being satisfied.

Permutation-and/or-sign-only invariant sets. In addition to the sufficient scatteredness, the identifiability of PMF will rely on the following condition for the sets of vertices of $\mathcal{P}_W$ and $\mathcal{P}_H$.

Definition 3. *A set $\mathcal{X}$ is called a permutation-and/or-sign-only invariant (PSOI) set if, and only if, every linear transformation A such that $A(\mathcal{X}) = \mathcal{X}$ is a signed permutation, that is, $A = D\Pi$ where Π is a permutation matrix and D is a diagonal matrix with diagonal entries in $\{-1, 1\}$.*

The set of vertices of full-dimensional polytopes will in most cases be PSOI sets.

Lemma 1. *Let the columns of $V \in \mathbb{R}^{r\times n}$ contain the vertices of the polytope $\mathcal{V} \subset \mathbb{R}^r$ and such that $\operatorname{rank}(V) = r$ (this holds for full-dimensional polytopes). Let $A \in \mathbb{R}^{r\times r}$ be such that $AV = V(:,\Pi)$ for some permutation Π. Then A is an orthogonal matrix, that is, a rotation of $\mathbb{R}^r$.*

Proof. Since A permutes the columns of V, and the set of permutations is finite, there exists n such that $A^n V = V$. Since V has rank r, it admits a right inverse, so that $A^n = I_r$, where I_r is identity matrix of dimension r. This implies that the eigenvalues of A are roots of 1, and hence A is orthogonal, that is, $A^\top A = I_r$. □

In two dimensions, sets that are not PSOI are any regular polygon centered at the origin, except for the square (which is obtained by a rotation of 90 or 180 degrees in which case A is a signed permutation). For example, the vertices of the regular triangle given by the columns of

$$V = \begin{pmatrix} 0 & \sqrt{3}/2 & -\sqrt{3}/2 \\ 1 & -1/2 & -1/2 \end{pmatrix}$$

are preserved by a rotation of 120 degrees, corresponding to $A = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}$, and $AV = \begin{pmatrix} \sqrt{3}/2 & -\sqrt{3}/2 & 0 \\ -1/2 & 1/2 & 1 \end{pmatrix}$.

In Section V, we will use two polytopes: Δ^r and $[a,b]^r$ for $b > a$. Let us show that their vertices are PSOI sets. For Δ^r, this is trivial since $\Delta^r = \operatorname{conv}(I_r)$, hence any A that satisfies $AI_r = I_r(:,\Pi)$ for some permutation Π must be a permutation (note there is no sign ambiguity possible here). For the hypercube $[a,b]^r$, let us first prove the following lemma.

Lemma 2. *Let $a < b$ be scalars, and $d \in \mathbb{R}^r$ with $\|d\|_2 = 1$ be such that $d^\top x \in \{a, b\}$ for all $x \in \{a,b\}^r$. Then d is a unit vector, up to multiplication by -1.*

Proof. Let us prove the result by induction. For $r = 1$, the result is trivial, we must have $d = 1$. Assume the result holds for all $r' < r$, and let us denote $d = [d_{r-1}, d_r]$ with $d_{r-1} \in \mathbb{R}^{r-1}$, and similarly for x. We have for all $x \in \{a,b\}^r$ that

$$d^\top x = d_{r-1}^\top x_{r-1} + d_r x_r \in \{a, b\}.$$

If $d_r \in \{-1, 0, 1\}$, the result follows by induction since $\|d\|_2 = 1$. Hence assume $d_r \notin \{-1,0,1\}$. We have

$$d_{r-1}^\top x_{r-1} + d_r a \in \{a,b\} \quad \text{and} \quad d_{r-1}^\top x_{r-1} + d_r b \in \{a,b\}.$$

Let us denote $\alpha = d_{r-1}^\top x_{r-1}$, we have

$$\alpha \in \{a - d_r a, b - d_r a\} \quad \text{and} \quad \alpha \in \{a - d_r b, b - d_r b\}.$$

Since $a \neq b$, $a - d_r a \neq a - d_r b$ and $b - d_r a \neq b - d_r b$ as $d_r \neq 0$, $a - d_r a \neq b - d_r b$ as $d_r \neq 1$, and $b - d_r a \neq b - d_r b$ as $d_r \neq -1$. Hence α cannot exist for $x_r \in \{a,b\}$, a contradiction. □

Corollary 1. *The set of vertices of $[a,b]^r$ is a PSOI set.*

Proof. The set of vertices of $[a,b]^r$ are all vectors in $\{a,b\}^r$. Let the columns of $V \in \mathbb{R}^{r\times 2^r}$ contain the vertices of $[a,b]^r$, and the linear transformation A satisfy $AV = V(:,\Pi)$ for some permutation Π. By Lemma 1, A is orthogonal hence its rows have unit ℓ_2 norm. This implies that every row of A must satisfy the condition of Lemma 2 and hence are unit vectors. Since rows of A are orthogonal, A must be a signed permutation. □

IV. Identifiability

We can now state our main result: it fills a gap in the literature by combining the ideas of the identifiability of maximum-volume PMF in [9], and of NMF in [8].

Theorem 2. *Let $(W, H, \mathcal{P}_W, \mathcal{P}_H)$ be a PMF of X of size $r = \operatorname{rank}(X)$. If $W^\top$ and H are SSFs, and $\operatorname{ext}(\mathcal{P}_W)$ and $\operatorname{ext}(\mathcal{P}_H)$ are PSOI sets, then the PMF $(W, H, \mathcal{P}_W, \mathcal{P}_H)$ of $X = WH$ of size $r = \operatorname{rank}(X)$ is identifiable.*

Proof. This proof follows that from [9, Th. 6] where only H is required to be sufficiently scattered while its volume is maximized. Let $Q \in \mathbb{R}^{r\times r}$ be an invertible matrix such that (WQ^{-1}, QH) is a PMF of X with

$$\operatorname{conv}(Q^{-\top} W^\top) \subseteq \mathcal{P}_W \quad \text{and} \quad \operatorname{conv}(QH) \subseteq \mathcal{P}_H. \tag{4}$$

Since $W^\top$ and H are sufficiently scattered factors, their convex hull contains their corresponding MVIE:

$$\mathcal{E}_{\mathcal{P}_W} \subset \operatorname{conv}(W^\top) \quad \text{and} \quad \mathcal{E}_{\mathcal{P}_H} \subset \operatorname{conv}(H). \tag{5}$$

Then, eq. (4) leads to

$$Q^{-\top}(\mathcal{E}_{\mathcal{P}_W}) \subseteq \mathcal{P}_W \quad \text{and} \quad Q(\mathcal{E}_{\mathcal{P}_H}) \subseteq \mathcal{P}_H. \tag{6}$$

The set $Q^{-\top}(\mathcal{E}_{\mathcal{P}_W})$ (resp. $Q(\mathcal{E}_{\mathcal{P}_H})$) is still an ellipsoid of volume $|\det(Q^{-1})|\operatorname{vol}(\mathcal{E}_{\mathcal{P}_W})$ (resp. $|\det(Q)|\operatorname{vol}(\mathcal{E}_{\mathcal{P}_H})$). By definition of the MVIE, we have

$$\begin{aligned}&|\det(Q^{-1})|\operatorname{vol}(\mathcal{E}_{\mathcal{P}_W}) \leq \operatorname{vol}(\mathcal{E}_{\mathcal{P}_W})\\ &\quad \text{and } |\det(Q)|\operatorname{vol}(\mathcal{E}_{\mathcal{P}_H}) \leq \operatorname{vol}(\mathcal{E}_{\mathcal{P}_H})\\ \Leftrightarrow &|\det(Q^{-1})| \leq 1 \text{ and } |\det(Q)| \leq 1 \Leftrightarrow |\det(Q)| = 1.\end{aligned}$$

This implies that $Q^{-\top}$ and Q respectively map $\mathcal{E}_{\mathcal{P}_W}$ and $\mathcal{E}_{\mathcal{P}_H}$ onto themselves :

$$Q^{-\top}(\mathcal{E}_{\mathcal{P}_W}) = \mathcal{E}_{\mathcal{P}_W} \quad \text{and} \quad Q(\mathcal{E}_{\mathcal{P}_H}) = \mathcal{E}_{\mathcal{P}_H}. \tag{7}$$

The remaining of the proof is exactly like in the remaining proof of [9, Th. 6] by focusing on either H or $W^\top$. Focus on H for example, and using [PMF.SSC2], the idea is to show that $Q(\operatorname{ext}(\mathcal{P}_H)) = \operatorname{ext}(\mathcal{P}_H)$. Then, because $\operatorname{ext}(\mathcal{P}_H)$ is a PSOI set, Q has to be a signed permutation. □

The last part of the proof of Theorem 2 does not rely on both $W^\top$ and H satisfying [PMF.SSC2], and on both $\operatorname{ext}(\mathcal{P}_W)$ and $\operatorname{ext}(\mathcal{P}_H)$ being PSOI sets. Actually, Theorem 2 remains valid if only one the factors satisfies [PMF.SSC2] and if the vertices of its corresponding polytope form a PSOI set.

Corollary 2. *Let $W^\top$ and H satisfy [PMF.SSC1] and*
(i) $W^\top$ satisfy [PMF.SSC2] and $\operatorname{ext}(\mathcal{P}_W)$ be a PSOI set, or
(ii) H satisfy [PMF.SSC2] and $\operatorname{ext}(\mathcal{P}_H)$ be a PSOI set,

then the PMF $(W, H, \mathcal{P}_W, \mathcal{P}_H)$ of $X = WH$ of size $r = \operatorname{rank}(X)$ is identifiable.

Proof. The same proof as Theorem 2 applies. By symmetry, whether it is (i) or (ii) that is verified allows us to conclude that Q is a signed permutation. □

The PSOI set condition can be relaxed to sets that are "mutually" PSOI, that is, there cannot exist a matrix A which is not a signed permutation such that $A^{-\top}(\operatorname{ext}(\mathcal{P}_W)) = \operatorname{ext}(\mathcal{P}_W)$ and $A(\operatorname{ext}(\mathcal{P}_H)) = \operatorname{ext}(\mathcal{P}_H)$.

Corollary 3. *Let $(W, H, \mathcal{P}_W, \mathcal{P}_H)$ be a PMF of X of size $r = \operatorname{rank}(X)$. If $W^\top$ and H are SSFs, and $\operatorname{ext}(\mathcal{P}_W)$ and $\operatorname{ext}(\mathcal{P}_H)$ are mutually PSOI sets, then the PMF $(W, H, \mathcal{P}_W, \mathcal{P}_H)$ of $X = WH$ of size $r = \operatorname{rank}(X)$ is identifiable.*

Proof. The same proof as Theorem 2 applies up to eq. (7). Then, $W^\top$ and H satisfying [PMF.SSC2] leads to $Q^{-\top}(\operatorname{ext}(\mathcal{E}_{\mathcal{P}_W})) = \operatorname{ext}(\mathcal{E}_{\mathcal{P}_W})$ and $Q(\operatorname{ext}(\mathcal{E}_{\mathcal{P}_H})) = \operatorname{ext}(\mathcal{E}_{\mathcal{P}_H})$. Then, because $\operatorname{ext}(\mathcal{P}_W)$ and $\operatorname{ext}(\mathcal{P}_H)$ are mutually PSOI sets, Q has to be a signed permutation. □

V. Examples of PMF

In this section, we show that some known constrained matrix factorizations are special instances of PMF, and explain how Theorem 2 relates to known identifiability results for these special cases.

A. Nonnegative Matrix Factorization (NMF)

An NMF, $X = WH$, requires W and H to be component-wise nonnegative. This is not a PMF since the nonnegative orthant is unbounded. However, if $W^\top$ and H do not contain a column full of zeros (which can be assumed w.l.o.g.), then there exist two diagonal matrices, D_l and D_r, such that $D_l W e = e$ and $e^\top H D_r = e^\top$. Hence we can transform the NMF $X = WH$ into the PMF $(\tilde{W}, \tilde{H}, \Delta^r, \Delta^r)$ of $\tilde{X}$ with $\tilde{X} = D_l X D_r$, where $\tilde{W} = D_l W$ and $\tilde{H} = H D_r$.

Interestingly, the identifiability conditions for NMF in Theorem 1 and for PMF in Theorem 2 are equivalent, because $\tilde{H}$ satisfies the SSC in Def. 1 *if and only if* $\tilde{H}$ is an SSF according to Def. 2, while $\operatorname{ext}(\Delta^r)$ is a PSOI set (see Section III). This is due to the fact that $\mathcal{E}_{\mathcal{P}_W} = \mathcal{E}_{\mathcal{P}_H} = \mathcal{C} \cap \Delta^r$, since the MVIE of Δ^r is an $(r-1)$-dimensional ball centered at $\frac{1}{r}e$ of radius $\frac{1}{\sqrt{r(r-1)}}$, within the affine subspace $\{x \in \mathbb{R}^r,\ e^\top x = 1\}$. Indeed, the diagonal matrices are just rescaling the rows of W and the columns of H such that they belong to Δ^r. Hence, $\mathcal{C} \cap \Delta^r \subseteq \operatorname{conv}(\tilde{H})$ *if and only if* $\mathcal{C} \subseteq \operatorname{cone}(H)$, and by symmetry this also holds for $\tilde{W}^\top$ and $W^\top$.

B. Factor-Bounded Matrix Factorization

Factor-bounded matrix factorization (FBMF) requires the elements of each factor to be bounded. Given $a < b \in \mathbb{R}$, we write $a \leq W \leq b$ if $a \leq W(i,k) \leq b$ for all (i,k).

Definition 4 (Factor-Bounded MF)**.** *Let $X \in \mathbb{R}^{m \times n}$, r be an integer, $l_W < u_W \in \mathbb{R}$ and $l_H < u_H \in \mathbb{R}$. The pair $(W, H) \in \mathbb{R}^{m \times r} \times \mathbb{R}^{r \times n}$ is a FBMF of X of size r for the intervals $[l_W, u_W]$ and $[l_H, u_H]$ if*

$$X = WH \text{ such that } l_W \leq W \leq u_W, l_H \leq H \leq u_H. \tag{8}$$

This means that each row of W then belongs to the hypercube $[l_W, u_W]^r$ and each column of H belongs to the hypercube $[l_H, u_H]^r$. In [14], the authors propose a nonnegative FBMF (NFBMF), where $0 \leq l_W$ and $0 \leq l_H$ in eq. (8). They showed that NFBMF is particularly well suited for clustering tasks. To the best of our knowledge, FBMF has never been proven to be identifiable. Since eq. (8) is a PMF with the choice $\mathcal{P}_W = [l_W, u_W]^r$ and $\mathcal{P}_H = [l_H, u_H]^r$, Theorem 2 applies to FBMF. The MVIE $\mathcal{E}_{\mathcal{P}_W}$ is an r-dimensional ball centered at $\frac{u_W + 2l_W}{2}e$ of radius $\frac{u_W - l_W}{2}$, and similarly for $\mathcal{E}_{\mathcal{P}_H}$, while $\operatorname{ext}(\mathcal{P}_W)$ and $\operatorname{ext}(\mathcal{P}_H)$ are PSOI sets (Corollary 1).

C. Bounded Simplex-Structured Matrix Factorization

Bounded simplex-structured matrix factorization (BSSMF) was introduced in [15] to explain data that are convex combinations of vectors belonging to a hyper-rectangle. Given the vectors $a \leq b \in \mathbb{R}^m$, we write $W(:,k) \in [a,b]$ if $a_i \leq W(i,k) \leq b_i$ for all i.

Definition 5 (BSSMF)**.** *Let $X \in \mathbb{R}^{m \times n}$, r be an integer, and $a, b \in \mathbb{R}^m$ with $a \leq b$. The pair $(W, H) \in \mathbb{R}^{m \times r} \times \mathbb{R}^{r \times n}$ is a BSSMF of X of size r for the interval $[a,b]$ if*

$$X = WH \text{ s.t. } W(:,k) \in [a,b] \text{ and } H(:,j) \in \Delta^r \text{ for all } k, j. \tag{9}$$

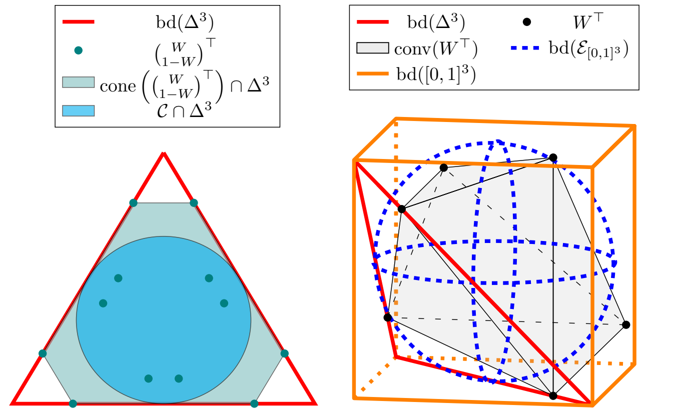
BSSMF does not belong to the class of PMFs. The hyperrectangle constraint on the columns of W cannot in general be expressed as a polytopic constraint on the rows of W. However, when all entries of a, and of b, are equal to one another, the hyperrectangle constraint becomes a hypercube constraint. For example, when X corresponds to a set of vectorized images, the intensity of a pixel belongs to $[0,1]$. If there is no specific pixel position that should be bounded differently than the others, every row of W is bounded in the same way. In other words, the rows of W belong to the hypercube $[0,1]^r$. Another example is when X is a rating matrix whose entries are ordinal, e.g., the Netflix matrix with entries in $\{1,2,3,4,5\} \in [1,5]$. In these cases, BSSMF uses a hypercube $[a,b]^m$ and is equivalent to PMF since eq. (9) is equivalent to eq. (2) with $\mathcal{P}_W = [a,b]^r$ and $\mathcal{P}_H = \Delta^r$. BSSMF was shown to be identifiable under the following conditions, different than in Theorem 2:

Theorem 3. *[15, Th. 2] Let $a \leq b \in \mathbb{R}^m$, $W \in \mathbb{R}^{m\times r}$ and $H \in \mathbb{R}^{r\times n}$ be such that $W(:,k) \in [a,b]$ for all k, $H(:,j) \in \Delta^r$ for all j. If $\binom{W-ae^\top}{be^\top - W}^\top \in \mathbb{R}^{r\times 2m}$ and $H \in \mathbb{R}^{r\times n}$ satisfy the SSC, then this BSSMF is essentially unique.*

When BSSMF and PMF are equivalent, which identifiability theorem is the strongest? Since BSSMF is invariant by translation along e, we can assume w.l.o.g. that $a = 0$ for the sake of simplicity. Also, we do not need to focus on the conditions for H. Indeed, when $H(:,j) \in \Delta^r$ for all j, [SSC1] is equivalent to [PMF.SSC1] because the MVIE of $\mathcal{P}_H = \Delta^r$ is equal to $\mathcal{C} \cap \Delta^r$. We then focus on the sufficient scatteredness of $W^\top$. The MVIE of $[0,b]^r$ is a ball $\mathcal{E}_{[0,b]^r}$ centered at $\frac{b}{2}e$ of radius $\frac{b}{2}$. This ball is tightly contained by $\mathcal{C}$, which means that for any convex set A that contains $\mathcal{E}_{[0,b]^r}$, $\mathcal{C} \subseteq \text{cone}(A)$. As a consequence, if $W^\top$ satisfies [PMF.SSC1], $W^\top$ satisfies [SSC1], which implies that $\binom{W}{be^\top - W}^\top$ satisfies [SSC1]. However, it is possible that $\binom{W}{be^\top - W}^\top$ satisfies [SSC1] while $W^\top$ does not satisfy [PMF.SSC1]. Here is an example with $\mathcal{P}_W = [0,1]^3$:

$$W^\top = \begin{pmatrix} 0.8 & 0 & 0.2 & 0.2 & 0.8 & 1 \\ 0.2 & 0.8 & 0 & 1 & 0.2 & 0.8 \\ 0 & 0.2 & 0.8 & 0.8 & 1 & 0.2 \end{pmatrix}. \quad (10)$$

As it can be seen in fig. 2a, the cone of $\binom{W}{1-W}^\top$ contains $\mathcal{C}$ because W reaches enough times the minimum and maximum bounds 0 and 1. However, in Figure 2b the convex hull of $W^\top$ does not contain the MVIE of $[0,1]^3$. Therefore, Theorem 2 is quite general but is not as strong as Theorem 3 for BSSMF.



(a) Visualization of [SSC1] being satisfied.
(b) Visualization of [PMF.SSC1] not being satisfied.

Fig. 2: Visualization of $\binom{W}{1-W}^\top$ from eq. (10) satisfying [SSC1] while $W^\top$ does not satisfy [PMF.SSC1]. The cone of $\binom{W}{1-W}^\top$ contains $\mathcal{C}$, while the convex hull of $W^\top$ does not contain the ball $\mathcal{E}_{[0,1]^3}$.

VI. Conclusion

We presented PMF, a structured matrix factorization model where the latent space of the factors is constrained by given polytopes. The choice of the polytopes should depend on the data and the application at hand. When the polytopes have certain invariant properties, we derived some sufficient conditions under which the identifiability of a PMF is guaranteed. Geometrically, these conditions are based on the scatteredness of the factors within the constraining polytopes.

References

[1] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, pp. 788–791, 1999.
[2] X. Fu, K. Huang, B. Yang, W.-K. Ma, and N. D. Sidiropoulos, “Robust volume minimization-based matrix factorization for remote sensing and document clustering,” *IEEE Trans. Signal Process.*, vol. 64 (23), pp. 6254–6268, 2016.
[3] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi, “A signal processing perspective on hyperspectral unmixing: Insights from remote sensing,” *IEEE Signal Process. Mag.*, vol. 31 (1), pp. 67–81, 2014.
[4] A. Ozerov and C. Févotte, “Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 18 (3), pp. 550–563, 2009.
[5] M. Sørensen, N. D. Sidiropoulos, and A. Swami, “Overlapping community detection via semi-binary matrix factorization: Identifiability and algorithms,” *IEEE Trans. Signal Process.*, vol. 70, pp. 4321–4336, 2022.
[6] Z.-Y. Zhang, T. Li, C. Ding, X.-W. Ren, and X.-S. Zhang, “Binary matrix factorization for analyzing gene expression data,” *Data Mining and Knowledge Discovery*, vol. 20, pp. 28–52, 2010.
[7] S. Rendle, W. Krichene, L. Zhang, and Y. Koren, “Revisiting the performance of ials on item recommendation benchmarks,” in *Proceedings of the 16th ACM Conference on Recommender Systems*, 2022, pp. 427–435.
[8] K. Huang, N. D. Sidiropoulos, and A. Swami, “Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition,” *IEEE Trans. Signal Process.*, vol. 62 (1), pp. 211–224, 2014.
[9] G. Tatli and A. T. Erdogan, “Polytopic matrix factorization: Determinant maximization based criterion and identifiability,” *IEEE Trans. Signal Process.*, vol. 69, pp. 5431–5447, 2021.
[10] ——, “Generalized polytopic matrix factorization,” in *ICASSP*, 2021.
[11] M. J. Todd, *Minimum-volume ellipsoids: Theory and algorithms*. SIAM, 2016.
[12] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
[13] 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, 2008.
[14] K. Liu, X. Li, Z. Zhu, L. Brand, and H. Wang, “Factor-bounded nonnegative matrix factorization,” *ACM Trans. Knowl. Discov. Data*, vol. 15 (6), May 2021.
[15] O. Vu Thanh, N. Gillis, and F. Lecron, “Bounded simplex-structured matrix factorization,” in *ICASSP*, 2022.