# Polytopic Matrix Factorization: Determinant Maximization Based Criterion and Identifiability

Gokcan Tatli Alper T. Erdogan

***Abstract*—We introduce Polytopic Matrix Factorization (PMF) as a novel data decomposition approach. In this new framework, we model input data as unknown linear transformations of some latent vectors drawn from a polytope. In this sense, the article considers a semi-structured data model, in which the input matrix is modeled as the product of a full column rank matrix and a matrix containing samples from a polytope as its column vectors. The choice of polytope reflects the presumed features of the latent components and their mutual relationships. As the factorization criterion, we propose the determinant maximization (Det-Max) for the sample autocorrelation matrix of the latent vectors. We introduce a sufficient condition for identifiability, which requires that the convex hull of the latent vectors contains the maximum volume inscribed ellipsoid of the polytope with a particular tightness constraint. Based on the Det-Max criterion and the proposed identifiability condition, we show that all polytopes that satisfy a particular symmetry restriction qualify for the PMF framework. Having infinitely many polytope choices provides a form of flexibility in characterizing latent vectors. In particular, it is possible to define latent vectors with heterogeneous features, enabling the assignment of attributes such as nonnegativity and sparsity at the subvector level. The article offers examples illustrating the connection between polytope choices and the corresponding feature representations.**

***Index Terms*—Polytopic Matrix Factorization, Nonnegative Matrix Factorization, Sparse Component Analysis, Independent Component Analysis, Bounded Component Analysis, Blind Source Separation.**

## I. Introduction

MATRIX factorization methods are fundamental algorithmic tools for both signal processing and machine learning (e.g., [1]–[7]). Revealing information hidden inside input data is a central problem in several applications. A common solution approach is to model the input matrix as the product of two factors.

In unsupervised settings, both factors are unknown, and there is no available training information for their estimation. Structured matrix factorization (SMF) methods utilize prior information or assumptions on both factors, such as rank, nonnegativity, sparsity, and antisparsity, to achieve the desired decomposition. In the semi-structured matrix factorization that we pursue in this article, the left factor is simply a full column rank matrix with no additional structure. In this setting, we refer to the columns of the right-factor as latent vectors, which have some presumed structure. The left factor is the linear transformation matrix that maps latent vectors to inputs. Due to the full column rank assumption on the left-factor matrix, the scope of the article is limited to the (over)determined case.

We can define the attributes of latent vectors through the choice of their domain. The topology of this set determines both the individual properties of latent vector components and their relationships. For example, for Nonnegative Matrix Factorization (NMF) [2], [8], [9], the domain choice is the nonnegative orthant, and for a related approach, Simplex Structured Matrix Factorization (SSMF) [5], it is the unit simplex. In addition, we can list two polytopic sets, namely the $\ell_\infty$-norm-ball for the antisparse version of Bounded Component Analysis (BCA) [6], [10] and the $\ell_1$-norm-ball for Sparse Component Analysis (SCA) [7], [11], [12], as further examples.

These matrix factorization frameworks have found successful applications in different domains. Basic applications of NMF include document mining [4], feature extraction for natural images [9], source separation for hyperspectral images [13], community detection [14] and audio demixing [3], [15], [16]. The main application area for SCA has been sparse dictionary learning, which has laid the foundation for the *sparse coding* principle, utilized in both computational neuroscience [17] and machine learning [18]. BCA has both dependent source and short-data-length separation capabilities with applications in natural image separation and digital communications [6], [10], [12].

Identifiability is a crucial concept in determining the applicability of matrix factorization methods. It concerns the ability to obtain unique factors of the input data up to some acceptable ambiguities, such as sign and permutation. All of the aforementioned domain choices, i.e., the nonnegative orthant, unit simplex, $\ell_\infty$, and $\ell_1$-norm balls, have been shown to lead to identifiable data models [6], [12], [19], [20].

A fundamental question is addressed in this article: *Can we extend domains enabling identifiability beyond these existing examples?* We indeed provide a positive answer and show that all polytopes that comply with a particular symmetry restriction qualify. We refer to the associated framework as Polytopic Matrix Factorization (PMF) and the polytopes that qualify for this framework as "identifiable polytopes". The availability of

Manuscript received February 13, 2021; revised July 5, 2021 and September 6, 2021; accepted September 6, 2021. Date of publication September 16, 2021; date of current version October 8, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Michael Muma. This work was supported in part by an AI Fellowship provided by the KUIS AI Lab. *(Corresponding author: Alper T. Erdogan.)*

The authors are with the Electrical-Electronics Engineering and KUIS AI Lab, Koc University College of Engineering, 34450 Istanbul, Turkey (e-mail: gtatli13@ku.edu.tr; alper@stanfordalumni.org).

Digital Object Identifier 10.1109/TSP.2021.3112918

infinite identifiable polytope choices offers a degree of freedom in generating a diverse set of feature attributes for latent vectors. For example, as shown in Section III, we can combine different attributes, such as nonnegativity, sparsity, and antisparsity, which separately exist in the SMF frameworks listed above, through proper selection of polytopes. Furthermore, as illustrated by the example in Section VI, we can define latent vectors with heterogeneous features by performing attribute assignments at the subvector level. In other words, it is possible, for example, to define latent vectors in which only a fraction of the components are nonnegative, and the sparsity is imposed on subsets of components. Part of this work, mainly Theorem 6 in Section IV on the characterization of identifiable polytopes for PMF, was presented in [21].

For the identification of the factor matrices in the PMF model, we propose the use of *determinant maximization criterion*, which has been successfully employed in both the NMF [2], [22] and BCA [6], [12] approaches. The determinant of the sample correlation matrix acts as a scattering measure for the set of latent vectors and its maximization targets to exploit the presumed spread of the latent vectors inside the polytope.

The success of the Det-Max criterion for the perfect recovery of factor matrices is dependent on the scattering of the latent vectors inside the polytope. Intuitively, if they are concentrated in a relatively small subregion of the polytope, they would fail to reflect its topology. Therefore, any criterion exploiting polytope membership information would fail in such a case. This article offers a sufficient condition on the spread of latent vectors inside the polytope to enable identifiability for the Det-Max criterion, providing theoretical grounds for this intuition.

We can position the proposed identifiability condition on PMF as an extension of the existing results in other SMF frameworks. The existing BCA approaches use two particular polytopes: $\ell_\infty$-norm-ball for antisparse and $\ell_1$-norm-ball for sparse components. The identifiability results for BCA assume that latent vectors contain the vertices of these polytopes [6], [12]. Similarly, the early identification results for NMF used the condition that latent vectors include the scaled corner points of the unit simplex [19], [23], [24]. This condition is referred as the separability/pure pixel condition or the groundedness assumption. These assumptions in both frameworks require the inclusion of specific points in a random collection of latent vectors, which is too restrictive for practical plausibility. The "sufficiently scattered" condition proposed for NMF in [25] significantly relaxed the corner inclusion assumption. This new condition requires that the conic hull of latent vectors contains a specific "reference cone" [2], [25]–[27]. Lin *et al.* [20] proposed a related sufficient scattering condition for SSMF.

Using geometric principles similar to those proposed for NMF and SSMF, we introduce a novel "sufficiently scattered" condition for the PMF framework. This new condition is much weaker than the vertex inclusion assumption used in BCA identifiability analysis for $\ell_1$ and $\ell_\infty$-norm balls [6], [12]. Furthermore, it is applicable to the class of all identifiable polytopes. The proposed criterion uses the maximum volume inscribed ellipsoid (MVIE) of the polytope, which can be considered its best inscribed ellipsoidal approximation. According to this new criterion, the samples should be sufficiently spread across the polytope such that their convex hull, i.e., the smallest convex set that contains these samples, also contains the MVIE of the polytope with a particular tightness constraint. In other words, they can be used to construct a more accurate model of the polytope than its best ellipsoidal approximation. In Section III, we illustrate latent vector sets satisfying the proposed identifiability condition for some particular polytopes. As demonstrated by these examples, this new condition leads to more practically plausible identifiable data models, compared to the vertex inclusion assumption. Furthermore, this condition forms the basis for the generalized polytope identifiability result offered in Section IV. We note that Lin *et al.* [5] proposed *an algorithm for SSMF* based on the MVIE of *the convex hull of the input vectors*. In our context, we use the MVIE of *the polytope* to define a *sufficient condition* for the *PMF identifiability*. Therefore, the MVIE concept is used in different domains (input vs latent vector spaces) and for different purposes (algorithm vs identifiability analysis).

We can summarize the main contributions of the article as follows:

- We offer a new, unsupervised data decomposition framework called Polytopic Matrix Factorization (PMF).
- We propose the use of a Det-Max criterion and introduce a novel geometric identifiability condition for PMF based on the MVIE of the polytope.
- We provide a characterization of identifiable polytopes.
- We illustrate the potential of the proposed PMF framework in terms of flexible description of latent vectors with heterogeneous features.

The following describes the organization of this article. In Section II, we provide the data model and the Det-Max optimization criterion for PMF. We also introduce the proposed sufficient scattering-based identifiability condition for the PMF framework. In Section III, we focus on four special polytopes corresponding to the combinations of antisparse/sparse and nonnegative/signed attributes and provide their identifiability results. In Section IV, we offer a theorem on the characterization of polytopes that qualify for the PMF framework. Section V presents a PMF algorithm adopted from the NMF literature. Section VI contains numerical examples for PMF. Finally, Section VII concludes the study.

Table I outlines the basic notations used throughout the article.

## II. Polytopic Matrix Factorization Problem

In this section, we introduce Polytopic Matrix Factorization as a new unsupervised data decomposition framework. We start by describing the PMF problem in Section II-A. Then, we define the determinant maximization based criterion for the PMF problem in Section II-B. In connection with this criterion, we provide the proposed PMF identifiability condition in Section II-C.

### A. PMF Problem

For the PMF problem, we assume the following generative data model: the input matrix $\mathbf{Y} \in \mathbb{R}^{M \times N}$ is given by

$$\mathbf{Y} = \mathbf{H}_g \mathbf{S}_g, \tag{1}$$

TABLE I
NOTATION

| Notation | Meaning |
|---|---|
| $\mathbf{S}_{j,:}$ ($\mathbf{S}_{:,j}$) | $j^{\text{th}}$ row (column) of matrix $\mathbf{S}$ |
| $\mathbf{e}_k$ | Standard basis vector which is all zeros except 1 at the $k^{\text{th}}$ index |
| $\langle \mathbf{x}, \mathbf{y} \rangle$ | Euclidean inner product between the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ defined as $\mathbf{y}^T\mathbf{x}$ |
| $\mathcal{B}_p$ | unit $\ell_p$-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 $\mathbf{d}$ (See Appendix A) |
| conv($\mathbf{S}$) | the convex hull of the columns of $\mathbf{S}$ |
| cone($\mathbf{S}$) | the conic hull of the columns of $\mathbf{S}$ |
| $\mathcal{K}^d$ | the dual cone of the cone $\mathcal{K}$ which is defined as $\{\mathbf{x} \mid \langle \mathbf{x}, \mathbf{y} \rangle \geq 0, \forall \mathbf{y} \in \mathcal{K}\}$ |
| ext($\mathcal{P}$) | the set of extreme points (vertices, or corner points) of the polytope $\mathcal{P}$ |
| bd($\mathcal{S}$) | the boundary of the set $\mathcal{S}$ |
| $\mathbf{A}(S)$ | the image of the set $S$ under the linear transformation with matrix $\mathbf{A}$ |
| $\mathbf{1}(\mathbf{0})$ | a vector or a matrix with all ones (zeros) of the appropriate dimensions |
| $\mathbb{R}^r_+$ | The $r$-dimensional nonnegative orthant |
| $\mathbf{A} \succ (\succeq) \mathbf{0}$ | $\mathbf{A}$ is a positive (semi)definite matrix |
| $\alpha\mathcal{S} + \mathbf{d}$ | The image of the set $\mathcal{S}$ under the transformation $f(\mathbf{x}) = \alpha\mathbf{x} + \mathbf{d}$ |

where

- $\mathbf{H}_g \in \mathbb{R}^{M \times r}$ is the ground truth of the left-factor matrix, which is assumed to be full column rank; and
- $\mathbf{S}_g \in \mathbb{R}^{r \times N}$ is the ground truth of the right-factor matrix, where we assume $r \leq \min(M, N)$. The underlying assumption of the PMF framework can be written as

$$\mathbf{S}_{g_{:,j}} \in \mathcal{P},\ j = 1, \ldots, N, \tag{2}$$

  where $\mathcal{P}$ is a convex polytope.
- There are two canonical forms to describe $\mathcal{P}$:
  - **H-Form** (*Intersections of Half-spaces*): A convex polytope $\mathcal{P}$ can be defined in the form

$$\mathcal{P} = \{\mathbf{x} \mid \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i, i = 1, \ldots, f\}, \tag{3}$$

    where $f$ is the number of faces, and vectors $\mathbf{a}_i$ are the face normals. Each inequality in (3) represents a half-space, and the intersection of these half-spaces forms a convex polyhedron. If $\mathcal{P}$ is bounded, we refer to it as a convex polytope.
  - **V-Form** (*Convex Hull of Vertices*): Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ represent the vertices, or the extreme points, of a convex polytope, where $m$ is the number of vertices, and then we can use

$$\mathcal{P} = \text{conv}(\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_m \end{bmatrix}), \tag{4}$$

    for the representation of the corresponding polytope.

  The conversion between these canonical forms is referred to as the *polyhedral representation conversion problem* [28].

The goal of PMF is to obtain a factorization of the input data $\mathbf{Y}$ in the form $\mathbf{Y} = \mathbf{HS}$ such that these factors satisfy

$$\mathbf{H} = \mathbf{H}_g\mathbf{\Pi}^T\mathbf{D}^{-1}, \tag{5}$$

$$\mathbf{S} = \mathbf{D}\mathbf{\Pi}\mathbf{S}_g, \tag{6}$$

where $\mathbf{\Pi} \in \mathbb{R}^{r \times r}$ is a permutation matrix that represents unresolvable ambiguity in obtaining the ordering of the columns (rows) of $\mathbf{H}_g$ ($\mathbf{S}_g$) and $\mathbf{D} \in \mathbb{R}^{r \times r}$ is a full rank diagonal matrix that corresponds to the scaling ambiguity.

In Section III, we provide PMF identifiability results using four particular polytopes of practical interest:

i. $\mathcal{P} = \mathcal{B}_\infty$, i.e., the unit $\ell_\infty$-norm-ball, which we refer to as the "antisparse" PMF case hereafter;
ii. $\mathcal{P} = \mathcal{B}_1$, i.e., the unit $\ell_1$-norm-ball, which we refer to as the "sparse" PMF case;
iii. $\mathcal{P} = \mathcal{B}_\infty \cap \mathbb{R}^r_+ = \{\mathbf{x} \mid \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}\}$, which is referred to as the "antisparse nonnegative" PMF case; and
iv. $\mathcal{P} = \mathcal{B}_1 \cap \mathbb{R}^r_+ = \{\mathbf{x} \mid \mathbf{1}^T\mathbf{x} \leq 1, \mathbf{x} \geq 0\}$, which is referred to as the "sparse nonnegative" PMF case.

We extend these results for a wider range (infinite number) of polytopes in Section IV.

### *B. The Criterion for Identification*

For the PMF problem outlined in Section II, we propose the use of the determinant maximization (Det-Max) approach, which has been successfully utilized in both the NMF [2], [27], [29] and antisparse/sparse BCA [6], [12], [30] frameworks. Therefore, the following serves as the prototype optimization problem throughout the article:

$$\underset{\mathbf{H} \in \mathbb{R}^{M \times r}, \mathbf{S} \in \mathbb{R}^{r \times N}}{\text{maximize}} \quad \det(\mathbf{S}\mathbf{S}^T) \tag{7a}$$

$$\text{subject to} \quad \mathbf{Y} = \mathbf{HS} \tag{7b}$$

$$\mathbf{S}_{:,j} \in \mathcal{P}, \quad j = 1, \ldots, N. \tag{7c}$$

The objective function of the Det-Max Optimization Problem in (7a) is equal to the determinant of the (scaled) sample correlation matrix $\mathbf{R_s} = \frac{1}{N}\mathbf{S}\mathbf{S}^T = \frac{1}{N}\sum_{j=1}^N \mathbf{S}_{:,j}\mathbf{S}_{:,j}^T$, which is a measure of nondegenerate scattering. In the zero mean case, the objective function boils down to "generalized variance," defined as the product of the eigenvalues of the covariance matrix [31]. These eigenvalues are the variances for the principal directions. Due to its product form, generalized variance is sensitive to the existence of directions with small variations. Therefore, its maximization corresponds to a nondegenerate spreading of the corresponding samples in all directions.

The determinant minimization criterion employed in the NMF approaches of [2], [22] corresponds to the minimization of $\det(\mathbf{H}^T\mathbf{H})$ in our problem. However, we utilize a dual approach that maximizes $\det(\mathbf{S}\mathbf{S}^T)$ for our identifiability results. The following definition classifies generative PMF settings with respect to the determinant maximization criterion:

*Definition II.1: Det-Max Identifiable PMF Generative Setting:* The generative data model described by (1) and (2) is called "Det-Max identifiable PMF setting" if all of the solutions of the

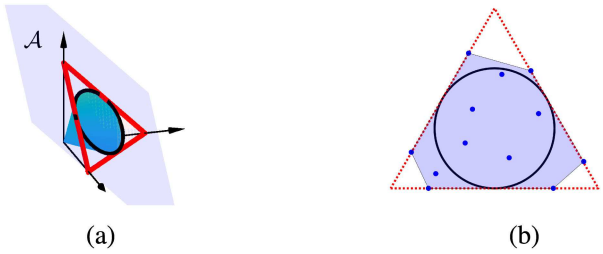



Fig. 1. Sufficient scattering condition for NMF: (a) 3D illustration of the unit simplex $\Delta_r$ (triangle with the red boundary), the second order cone $\mathcal{C}$, and the hyperplane $\mathcal{A}$. (b) an illustration of sufficiently scattered samples (dots) and their convex hull in relation to $\mathcal{C} \cap \mathcal{A}$.

corresponding *Det-Max optimization problem* in (7) satisfy the forms in (5) and (6).

### C. Proposed PMF Identifiability Condition

One of the basic premises of this article is to provide an identifiability condition for the Det-Max optimization problem introduced in Section II-B. This criterion assumes that the columns of the generative model matrix $\mathbf{S}_g$ are well spread inside $\mathcal{P}$. The Det-Max optimization problem in (7) targets the dispersal of the columns of $\mathbf{S}$ to exploit this assumption. The identifiability condition offered in this section is a geometric condition on the columns of $\mathbf{S}_g$, to guarantee their sufficient scattering in $\mathcal{P}$.

Earlier Det-Max optimization based BCA approaches used the inclusion of the polytope vertices as the sufficient identifiability condition for $\mathcal{B}_\infty$ [6] and $\mathcal{B}_1$ [12]. This assumption resembles the "separability," "pure pixel" or "groundedness" sufficient condition of the NMF/SSMF frameworks [19], [23], requiring that latent vectors include the vertices of the unit simplex or their scaled versions. If we assume that the latent vectors are randomly drawn from their domains, the probability of the vertex inclusion is very low. Therefore, from the practical plausibility standpoint, we desire less stringent sufficient conditions.

To address this issue, weaker "sufficiently scattered conditions" were introduced for NMF. These conditions are mainly based on the enclosure of the second order cone $\mathcal{C} = \{\mathbf{x} \mid \mathbf{x}^T\mathbf{1} \geq \sqrt{r-1}\|\mathbf{x}\|_2, \mathbf{x} \in \mathbb{R}^r\}$ as a measure of spread inside the nonnegative orthant. Fig. 1(a) illustrates $\mathcal{C}$ for the three-dimensional case. A common approach employed in NMF algorithms is to preprocess inputs to enforce unit $\ell_1$-norm constraints on the nonnegative latent vectors. As a result of this normalization, the original vectors in nonnegative orthant $\mathbb{R}^r_+$ are mapped to the unit simplex $\Delta_r = \{\mathbf{x} \mid \mathbf{1}^T\mathbf{x} = 1, \mathbf{x} \geq 0, \mathbf{x} \in \mathbb{R}^r\}$, which is the region with the triangular boundary in Fig. 1(a). Due to this mapping, we can focus on the $r-1$ dimensional affine subspace $\mathcal{A} = \{\mathbf{x} \mid \mathbf{1}^T\mathbf{x} = 1, \mathbf{x} \in \mathbb{R}^r\}$. Fig. 1(b) illustrates the restriction to $\mathcal{A}$, where the latent vectors are represented with the dots, and $\text{bd}(\mathcal{C} \cap \mathcal{A})$ is the circle.

Based on this geometric setting, Fu *et al.* [27] proposed the combination of the following conditions for sufficient scattering:

(NMF.SS.i) $\text{cone}(\mathbf{S}) \supseteq \mathcal{C}$,

(NMF.SS.ii) $\text{cone}(\mathbf{S})^d \cap \text{bd}(\mathcal{C}^d) = \{\gamma\mathbf{e}_k \mid \gamma > 0, k = 1, \dots r\}$.

The notation $\mathcal{K}^d$ in (NMF.SS.ii) represents the dual cone of $\mathcal{K}$, as defined in Table I. The first condition, (NMF.SS.i), ensures that the conic hull of the columns of $\mathbf{S}$ contains $\mathcal{C}$. Restricted to the affine subspace $\mathcal{A}$, this condition is equivalent to that in which the convex hull of the (normalized) latent vectors, the purple shaded region in Fig. 1(b), contains $\mathcal{C} \cap \mathcal{A}$. The second condition, (NMF.SS.ii), limits the tightness of the enclosure by constraining the points of tangency between $\mathcal{C}$ and $\text{cone}(\mathbf{S})$.

Lin *et al.* [20] introduced an alternative but related [32] condition for SSMF, which is based on the set $\mathcal{R}(a) = (a\mathcal{B}_2) \cap \Delta_r$, i.e., the intersection of the origin centered hypersphere with radius $a$ and the unit simplex. The constant $\gamma = \sup\{a \leq 1 \mid \mathcal{R}(a) \subseteq \text{conv}(\mathbf{S})\}$, where the columns of $\mathbf{S}$ are in $\Delta_r$, is defined as the uniform pixel purity level. The sufficiently scattered condition in [20] requires that $\gamma > \frac{1}{r}$.

In this article, we extend the sufficient scattering condition approach introduced for NMF in [27] to PMF. For this purpose, we replace the second order cone $\mathcal{C}$ in NMF with the MVIE of the polytope. The MVIE serves as the reference object to measure the spread of the latent vectors inside $\mathcal{P}$. The MVIE of a polytope $\mathcal{P}$ can be represented with (see Section 8.4.2, Page 400 of [33])

$$\mathcal{E}_\mathcal{P} = \{\mathbf{C}_\mathcal{P}\mathbf{u} + \mathbf{g}_\mathcal{P} \mid \|\mathbf{u}\|_2 \leq 1\}, \tag{8}$$

where, for a polytope defined by (3), the pair $(\mathbf{C}_\mathcal{P} \in \mathbb{R}^{r\times r}, \mathbf{g}_\mathcal{P} \in \mathbb{R}^r)$ is obtained as the optimal solution of the optimization problem:

$$\underset{\mathbf{C}\in\mathbb{R}^{r\times r}, \mathbf{g}\in\mathbb{R}^r}{\text{minimize}} \quad -\log\det\mathbf{C} \tag{9a}$$

$$\text{subject to} \quad \|\mathbf{C}\mathbf{a}_i\|_2 + \mathbf{a}_i^T\mathbf{g} \leq b_i, \quad i = 1, \dots, f \tag{9b}$$

$$\mathbf{C} \succeq \mathbf{0}. \tag{9c}$$

The following theorem by Fritz John [34] is useful for identifying spherical MVIEs:

*Theorem 1:* $\mathcal{B}_2$ is the ellipsoid of maximal volume contained in the convex body $C \subset \mathbb{R}^r$ if and only if $\mathcal{B}_2 \subset C$ and, for some $m \geq r$, there are unit 2-norm vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_m\} \subset \mathbb{R}^r$ on the boundary of $C$, and positive numbers $\{c_i, i = 1, \dots, m\}$ for which $\sum_{i=1}^m c_i\mathbf{u}_i = \mathbf{0}$ and $\sum_{i=1}^m c_i\mathbf{u}_i\mathbf{u}_i^T = \mathbf{I}_r$.

Based on Theorem 1, we can show that the special polytope $\mathcal{B}_\infty$ (item (i) in Section II.) has $\mathcal{B}_2$ as its MVIE, with the choices of $\mathbf{u}_i = \mathbf{e}_i, i = 1, \dots, r$ and $\mathbf{u}_i = -\mathbf{e}_{i-r}, i = r+1, \dots 2r$, $c_i = 1, i = 1, \dots, r$. Similarly, the MVIE of $\sqrt{r}\mathcal{B}_1$ is also $\mathcal{B}_2$, which can be justified by Theorem 1, through the choice of $\mathbf{u}_i = \frac{1}{\sqrt{r}}\mathbf{q}_i, i = 1, \dots, 2^r$, where vectors $\mathbf{q}_i$ are all possible distinct sign vectors with $\pm 1$ entries.

We propose the following MVIE-based sufficiently scattering condition to be used in the identifiability results.

*Definition II.2: Sufficiently Scattered Factor:* $\mathbf{S} \in \mathbb{R}^{r\times N}$ is called a sufficiently scattered factor corresponding to $\mathcal{P}$ if

*(PMF.SS.i)* $\mathcal{P} \supseteq \text{conv}(\mathbf{S}) \supset \mathcal{E}_\mathcal{P}$, and

*(PMF.SS.ii)* $\text{conv}(\mathbf{S})^{*,\mathbf{g}_\mathcal{P}} \cap \text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) = \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})$,

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

The condition (PMF.SS.i) in Definition II.2 guarantees that the convex hull of the columns of $\mathbf{S}$ contains the MVIE of $\mathcal{P}$. Fig. 2 illustrates a set of sufficiently scattered samples for $\mathcal{P} = \mathcal{B}_\infty \subset$

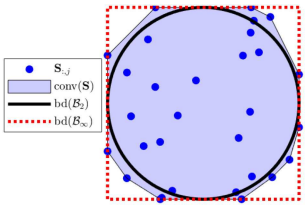



Fig. 2. *Sufficiently scattered case* example for 2D antisparse PMF.

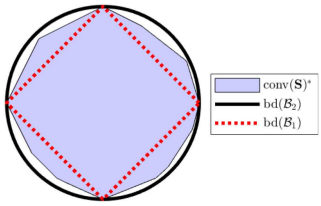



Fig. 3. Polar sets for the example in Fig. 2.

$\mathbb{R}^2$. Here the square region (with red borders) is $\mathcal{B}_\infty$ polytope, the circle is the boundary of its MVIE $\mathcal{B}_2$, the dots represent the sufficiently scattered samples, and the purple shaded region is the convex hull of these samples. One can consider MVIE in Fig. 2 to be an ellipsoidal approximation of the polytope. The condition (PMF.SS.i) essentially ensures that the convex hull of the samples forms a better approximation of $\mathcal{P}$ than its MVIE.

Furthermore, the polar domain constraint (PMF.SS.ii) in Definition II.2 places a restriction on the intersection between $\text{conv}(\mathbf{S})^{*,\mathbf{g}_\mathcal{P}}$ and $\text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})$. Fig. 3 provides the polar domain picture corresponding to the example in Fig. 2. The polar of $\mathcal{B}_\infty$ is $\mathcal{B}_1$, the boundary of which is shown with the dotted (red) lines. This is due to the fact that the face normals of $\mathcal{B}_\infty$, which are the standard basis vectors and their negatives, are the vertices of the polar polytope (see Appendix A). The polar of the MVIE $\mathcal{B}_2$ is equal to itself. Here, we can clearly observe the reversal of the inclusion relationship $\mathcal{P} \supseteq \text{conv}(\mathbf{S}) \supset \mathcal{E}_\mathcal{P}$ in the sample domain as $\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}} \supset \text{conv}(\mathbf{S})^{*,\mathbf{g}_\mathcal{P}} \supseteq \mathcal{P}^{*,\mathbf{g}_\mathcal{P}}$ in the polar domain. Furthermore, we observe from Fig. 3 that the boundary of $\mathcal{E}_{\mathcal{B}_\infty}^*$ intersects $\text{conv}(\mathbf{S})^*$ at the vertices of $\mathcal{B}_1$, i.e., the standard basis vectors and their negatives. These intersection points in the polar domain correspond to the normals of hyperplanes where $\text{conv}(\mathbf{S})$ and $\mathcal{B}_\infty$ are tangent to the boundary of the MVIE $\mathcal{B}_2$ in Fig. 2. As illustrated by this example in Figs. 2 and 3, the polar domain constraint (PMF.SS.ii) limits the points of tangency between $\text{conv}(\mathbf{S})$ and $\mathcal{E}_\mathcal{P}$ to the intersection of the polytope $\mathcal{P}$ and the boundary of its MVIE $\text{bd}(\mathcal{E}_\mathcal{P})$. Therefore, we can consider (PMF.SS.ii) to be a constraint on how tightly $\text{conv}(\mathbf{S})$ can enclose $\mathcal{E}_\mathcal{P}$.

## III. Special PMF Cases

This section focuses on the special polytopes introduced in Section II-A corresponding to the combination of multiple component attributes such as sparse/antisparse and nonnegative/signed due to their practical relevance in existing applications. We provide the generalization of all of the identifiable polytopes in Section IV.

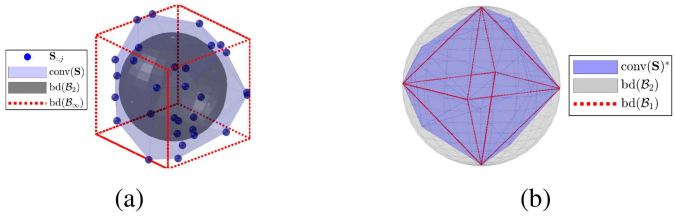



Fig. 4. Sufficiently scattered case example for 3D antisparse PMF: (a) sample domain, (b) polar domain.

### A. Antisparse PMF

The antisparse case corresponds to the setting in which the columns of the $\mathbf{S}_g$ matrix are distributed inside the $\ell_\infty$-norm-ball; i.e., $\mathcal{P} = \mathcal{B}_\infty$ as defined in item (i) in Section II-A. In the sufficiently scattered case, $\mathbf{S}_g$ has columns for which near-maximum magnitude values are simultaneously achieved for all of their components, hence the name antisparse [35], [36]. Such factorization has also been referred to as "democratic representations" [37]. The reference [6] proposed a BCA framework that exploited the use of $\mathcal{B}_\infty$ as the domain of latent vectors. This approach also employs the determinant maximization criterion. However, instead of defining an optimization problem with a $\mathcal{B}_\infty$ constraint, it proposes an unconstrained optimization problem. The antisparse BCA objective function contains a penalty term corresponding to the "size" of the minimum volume $\mathcal{B}_\infty$ polytope enclosing latent vectors. The identifiability results offered in [6] assumed that the latent vectors in the generative model contain all of the vertices of $\mathcal{B}_\infty$. In this section, we show that we can replace the vertex-inclusion assumption with the less stringent sufficiently scattering assumption in Definition II.2 for $\mathcal{B}_\infty$. In Section II-C, using Theorem 1, we showed that the MVIE for $\mathcal{B}_\infty$ is $\mathcal{B}_2$. Therefore, the corresponding MVIE parameters in description (8) are $\mathbf{C}_{\mathcal{B}_\infty} = \mathbf{I}$ and $\mathbf{g}_{\mathcal{B}_\infty} = \mathbf{0}$.

In Section II-C, Figs. 2 and 3 offered illustrations for the convex hull of sufficiently scattered samples in $\mathcal{B}_\infty$, the polytope $\mathcal{B}_\infty$, its MVIE $\mathcal{E}_{\mathcal{B}_\infty}$ and their polars, in a two-dimensional setting. Fig. 4(a) illustrates a sufficiently scattered selection of the columns of $\mathbf{S}$ for the three-dimensional case ($r = 3$), where the vertices of $\mathcal{B}_\infty$ are not included in the samples. The corresponding three-dimensional polar domain picture is provided in Fig. 4(b).

The following theorem characterizes the identifiability of the antisparse PMF problem under the proposed sufficiently scattered condition on $\mathbf{S}_g$.

*Theorem 2:* Given the general PMF setting outlined in Section II-A, if $\mathbf{S}_g$ is a sufficiently scattered factor for *antisparse PMF* according to Definition II.2, then all global optima $\mathbf{H}_*, \mathbf{S}_*$ of the Det-Max optimization problem in (7) for $\mathcal{P} = \mathcal{B}_\infty$ satisfy

$$\mathbf{H}_* = \mathbf{H}_g \mathbf{\Pi}^T \mathbf{D}, \tag{10}$$

$$\mathbf{S}_* = \mathbf{D} \mathbf{\Pi} \mathbf{S}_g, \tag{11}$$

where $\mathbf{\Pi} \in \mathbb{R}^{r \times r}$ is a permutation matrix and $\mathbf{D} \in \mathbb{R}^{r \times r}$ is an invertible diagonal matrix with $\pm 1$ entries on its diagonal.

*Proof:* Due to the full column rank condition on $\mathbf{H}_g$ and the constraint $\mathbf{Y} = \mathbf{H}\mathbf{S}$ in (7b), any feasible point $\mathbf{S}$ has the same

row space as $\mathbf{S}_g$, which implies

$$\mathbf{S} = \mathbf{A}\mathbf{S}_g, \tag{12}$$

for some full rank $\mathbf{A} \in \mathbb{R}^{r\times r}$ matrix. Therefore, finding the optimal choice of $\mathbf{S}$ boils down to finding the optimal $\mathbf{A}$. Using the parametrization in (12), the objective function in (7a) of the Det-Max optimization problem can be written as

$$\det(\mathbf{S}\mathbf{S}^T) = |\det(\mathbf{A})|^2 \det(\mathbf{S}_g\mathbf{S}_g^T).$$

Since the second term on the right-hand side is constant, the optimization objective function can be reduced to $|\det(\mathbf{A})|$. Therefore, we can write the equivalent problem of (7) for the antisparse PMF case as

$$\underset{\mathbf{A}\in\mathbb{R}^{r\times r}}{\text{maximize}} \quad |\det(\mathbf{A})| \tag{13a}$$

$$\text{subject to} \quad \|\mathbf{A}\mathbf{S}_{g_{:,j}}\|_\infty \leq 1, \quad j = 1,\dots,N. \tag{13b}$$

The remaining proof consists of three main steps.

The first step:> *Using the constraint in (13b), we first show that the rows of any feasible* $\mathbf{A}$ *are in conv*$(\mathbf{S}_g)^*$.

For this purpose, using (13b), we can write $\mathbf{A}_{i,:}\mathbf{S}_{g_{:,j}} \leq 1$, for all $(i,j)$ index pairs, which further implies that $\mathbf{A}_{i,:}$ satisfies

$$\mathbf{A}_{i,:}\mathbf{s} = \mathbf{A}_{i,:}\mathbf{S}_g\boldsymbol{\lambda} \leq 1, \quad \boldsymbol{\lambda} \in \mathbb{R}_+^N,\ \mathbf{1}^T\boldsymbol{\lambda} = 1,$$

for all $\mathbf{s} \in \text{conv}(\mathbf{S}_g)$ and $i \in \{1,\dots,r\}$. This condition is equivalent to each row of $\mathbf{A}$ lying in the polar of $\text{conv}(\mathbf{S}_g)$, i.e.,

$$\mathbf{A}_{i,:}^T \in \text{conv}(\mathbf{S}_g)^*, \quad \forall i \in \{1,\dots,r\}.$$

In reference to the three-dimensional polar domain picture in Fig. 4(b), the rows of $\mathbf{A}$ lie in the polytopic shaded region corresponding to $\text{conv}(\mathbf{S}_g)^*$.

The second step: *Using (PMF.SS.i) and Hadamard's inequality, we show that any optimal solution* $\mathbf{A}_*$ *of (13) should be a real orthogonal matrix.*

The polar version of the sufficient scattering condition (PMF.SS.i), i.e., $\text{conv}(\mathbf{S}_g) \supset \mathcal{E}_\mathcal{P}$, is equivalent to $\text{conv}(\mathbf{S}_g)^* \subset \mathcal{B}_2$. Therefore, we have $\|\mathbf{A}_{i,:}\|_2 \leq 1$ for all rows of $\mathbf{A}$. This finding implies that the rows of $\mathbf{A}$ lie inside the unit sphere in Fig. 4(b). Hadamard's inequality-based bound on the objective function in (13a),

$$|\det(\mathbf{A})| \leq \|\mathbf{A}_{1,:}^T\|_2\|\mathbf{A}_{2,:}^T\|_2 \dots \|\mathbf{A}_{r,:}^T\|_2 \leq 1,$$

is achieved if and only if the rows of $\mathbf{A}$ are on $\text{bd}(\mathcal{B}_2)$ and they are mutually orthogonal. In other words, any optimal solution $\mathbf{A}_*$ is a real orthogonal matrix. Therefore, for the example case in Fig. 4, the rows of $\mathbf{A}_*$ should lie on the boundary of the unit sphere in Fig. 4(b). At the same time, they should be members of $\text{conv}(\mathbf{S}_g)^*$, the purple shaded region in the same figure.

The third step: *Using the polar domain constraint (PMF.SS.ii), we show that any optimal real orthogonal* $\mathbf{A}_*$ *has only one nonzero element at each row (column).*

The sufficiently scattered condition (PMF.SS.ii), $\text{conv}(\mathbf{S}_g)^* \cap \text{bd}(\mathcal{B}_2) = \text{ext}(\mathcal{B}_1)$, restricts the rows of any optimal solution $\mathbf{A}_*$, which are located on the boundary of the unit sphere in Fig. 4(b), to the vertices of $\mathcal{B}_1$. This condition implies that the rows of $\mathbf{A}_*$ are standard basis vectors (or their negatives). Therefore, we can write $\mathbf{A}_* = \mathbf{D}\boldsymbol{\Pi}$, where $\mathbf{D} \in \mathbb{R}^{r\times r}$ is a diagonal matrix with $\pm 1$ entries on its diagonal and $\boldsymbol{\Pi} \in \mathbb{R}^{r\times r}$ is a permutation matrix. Due to the equality constraint in (7b), $\mathbf{H}_* = \mathbf{H}_g\mathbf{A}^{-1}$; therefore, (10) and (11) follow. ■

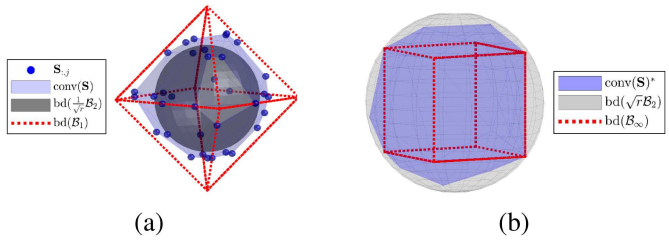



Fig. 5. Sufficiently scattered case example for 3D sparse PMF: (a) sample domain, (b) polar domain.

### B. Sparse PMF

In the sparse PMF setting defined in item (ii) in Section II-A, the columns of $\mathbf{S}_g$ are located inside the $\ell_1$-norm-ball; i.e., $\mathcal{P} = \mathcal{B}_1$. The connection between sparsity and $\ell_1$-norm constraints was well established in [38], [39]. It was shown that, under some practically plausible assumptions, the $\ell_1$-norm acts as a convex surrogate for the $\ell_0$-norm which counts the number of non-zero elements in a given vector. The sparsity property has also been exploited in different unsupervised approaches (see, for example, [7], [11], [40] and references therein). In particular, [12] adopted the determinant maximization based antisparse BCA approach in [6] to the sparse case by replacing the minimum volume enclosing $\mathcal{B}_\infty$ with its $\mathcal{B}_1$ counterpart. The identifiability result for the sparse BCA in [12] is based on the assumption that $\mathbf{S}_g$ contains all vertices of $\mathcal{B}_1$. In this section, we show that we can relax this condition using the sufficiently scattered condition in Definition II.2.

In Section II-C, using Theorem 1, we show that the MVIE of $\sqrt{r}\mathcal{B}_1$ is $\mathcal{B}_2$. Therefore, the MVIE of $\mathcal{B}_1$ is $\mathcal{E}_{\mathcal{B}_1} = \frac{1}{\sqrt{r}}\mathcal{B}_2$. The polar of a hypersphere is another hypersphere with the reciprocal radius. Therefore, $\mathcal{E}_{\mathcal{B}_1}^* = \sqrt{r}\mathcal{B}_2$.

For a visual illustration of the sufficient scattering condition for $\mathcal{P} = \mathcal{B}_1$, we consider the example in Fig. 5(a) for $r = 3$. The sample points, represented by dots, in Fig. 5(a) do not contain the vertices of $\mathcal{B}_1$. Furthermore, both $\text{bd}(\mathcal{B}_1)$ and the edges of $\text{conv}(\mathbf{S})$ intersect $\text{bd}(\frac{1}{\sqrt{r}}\mathcal{B}_2)$ at identical points due to the polar domain sufficient scattering constraint (PMF.SS.ii), as illustrated in Fig. 5(b). The polar of $\mathcal{B}_1$ is $\mathcal{B}_\infty$, the boundary of which is plotted as the red cube in Fig. 5(b). Its vertices and $\text{conv}(\mathbf{S})^*$ intersect the boundary of $\mathcal{E}_\mathcal{P}^*$ at identical points.

The following theorem characterizes the identifiability of the sparse PMF problem, based on the sufficiently scattering condition in Definition II.2:

*Theorem 3:* Given the general PMF setting outlined in Section II-A, if $\mathbf{S}_g$ is a sufficiently scattered factor for *sparse PMF* according to Definition II.2, then all global optima $\mathbf{H}_*, \mathbf{S}_*$ of the Det-Max optimization problem in (7) with $\mathcal{P} = \mathcal{B}_1$ satisfy

$$\mathbf{H}_* = \mathbf{H}_g\boldsymbol{\Pi}^T\mathbf{D},$$

$$\mathbf{S}_* = \mathbf{D}\boldsymbol{\Pi}\mathbf{S}_g,$$

where $\mathbf{\Pi} \in \mathbb{R}^{r\times r}$ is a permutation matrix and $\mathbf{D} \in \mathbb{R}^{r\times r}$ is a diagonal matrix with $\pm 1$ entries on its diagonal.

*Proof:* Using the same arguments in the proof of Theorem 2, we write the optimization problem equivalent to (7) for the sparse PMF case as

$$\underset{\mathbf{A}\in\mathbb{R}^{r\times r}}{\text{maximize}} \quad |\det(\mathbf{A})| \tag{14a}$$

$$\text{subject to} \quad \|\mathbf{A}\mathbf{S}_{g_{:,j}}\|_1 \leq 1, \quad j = 1,\dots,N. \tag{14b}$$

The proof consists of three fundamental steps:

The first step: *We show that (14b) and (PMF.SS.i) imply any feasible $\mathbf{A}^T$ maps $\mathcal{B}_\infty$ into $\sqrt{r}\mathcal{B}_2$.*

We start by noting that $\|\mathbf{A}\mathbf{x}\|_1 = \langle \mathbf{x}, \mathbf{A}^T \text{sign}(\mathbf{A}\mathbf{x})\rangle$, for any $\mathbf{x} \in \mathbb{R}^r$. Therefore, from (14b), we can write

$$\langle \mathbf{S}_{g_{:,j}}, \mathbf{A}^T\mathbf{q}_i\rangle \leq 1, \quad j = 1,\dots,N, \ \forall \mathbf{q}_i \in \text{ext}(\mathcal{B}_\infty), \tag{15}$$

where we used $\text{ext}(\mathcal{B}_\infty)$ as the set of all possible sign vectors. From (15), we conclude that

$$\mathbf{A}^T\mathbf{q}_i \in \text{conv}(\mathbf{S}_g)^*, \qquad \forall \mathbf{q}_i \in \text{ext}(\mathcal{B}_\infty). \tag{16}$$

In other words, any feasible $\mathbf{A}^T$ maps the vertices of $\mathcal{B}_\infty$ into $\text{conv}(\mathbf{S}_g)^*$, purple shaded polytopic region in Fig. 5(b). Since $\mathbf{S}_g$ is a sufficiently scattered factor, we have $\text{conv}(\mathbf{S}_g)^* \subset \sqrt{r}\mathcal{B}_2$ by (PMF.SS.i). Therefore,

$$\mathbf{A}^T\mathbf{q}_i \in \sqrt{r}\mathcal{B}_2, \qquad \forall \mathbf{q}_i \in \text{ext}(\mathcal{B}_\infty). \tag{17}$$

(17) further implies $\mathbf{A}(\mathcal{B}_\infty) \subset \sqrt{r}\mathcal{B}_2$, using the convexity of $\mathcal{B}_2$ and that $\mathcal{B}_\infty$ is the convex hull of $\text{ext}(\mathcal{B}_\infty)$. Therefore, the image of $\mathcal{B}_\infty$ under $\mathbf{A}^T$ lies inside the spherical region in Fig. 5(b).

The second step: *We show that any $\mathbf{A}$ satisfying (17) and maximizing $|\det(\mathbf{A})|$ is a real orthogonal matrix*

Replacing $\text{conv}(\mathbf{S}_g)^*$ in (16) with the larger set $\sqrt{r}\mathcal{B}_2$, we obtain the following optimization problem, the solution of which provides an upper bound for (14):

$$\underset{\mathbf{A}\in\mathbb{R}^{r\times r}}{\text{maximize}} \quad |\det(\mathbf{A})| \tag{18a}$$

$$\text{subject to} \quad \|\mathbf{A}^T\mathbf{q}_i\|_2^2 \leq r, \quad \forall \mathbf{q}_i \in \text{ext}(\mathcal{B}_\infty). \tag{18b}$$

Further relaxing (18b) by totaling its individual constraints, we obtain an alternative optimization for obtaining another upper bound:

$$\underset{\mathbf{A}\in\mathbb{R}^{r\times r}}{\text{maximize}} \quad |\det(\mathbf{A})| \tag{19a}$$

$$\text{subject to} \quad \sum_{\mathbf{q}_i\in\text{ext}(\mathcal{B}_\infty)} \|\mathbf{A}^T\mathbf{q}_i\|_2^2 \leq 2^r r. \tag{19b}$$

We now show that globally optimal solutions for (19) are real orthogonal matrices, which are also feasible, and therefore optimal, for the problem in (18). First, we note that the constraint (19b) can be rewritten more compactly as $Tr(\mathbf{A}\mathbf{A}^T\mathbf{Q}\mathbf{Q}^T) \leq 2^r r$, where $\mathbf{Q} = [\,\mathbf{q}_1\ \mathbf{q}_2\ \dots\ \mathbf{q}_{2^r}\,] \in \mathbb{R}^{r\times 2^r}$. Due to the symmetry of the extreme points of $\text{ext}(\mathcal{B}_\infty)$, we have $\mathbf{Q}\mathbf{Q}^T = 2^r\mathbf{I}$. Therefore, the constraint (19b) is further simplified to $\|\mathbf{A}\|_F^2 \leq r$, which can be written in terms of the singular values of $\mathbf{A}$, $\{\sigma_i(\mathbf{A}), i \in \{1,\dots,r\}\}$, as

$$\sqrt{\sum_{i=1}^{r}\sigma_i^2(\mathbf{A})} \leq \sqrt{r}. \tag{20}$$

For the objective function in (19a), we can write

$$\begin{aligned} |\det(\mathbf{A})| &= \prod_{i=1}^{r}\sigma_i(\mathbf{A}), \\ &\leq \left(\frac{1}{r}\sum_{i=1}^{r}\sigma_i(\mathbf{A})\right)^r, &(21)\\ &\leq \left(\frac{1}{r}\sqrt{\sum_{i=1}^{r}\sigma_i^2(\mathbf{A})}\sqrt{r}\right)^r, &(22)\\ &\leq 1, &(23)\end{aligned}$$

where (21) is due to the arithmetic-geometric mean inequality, (22) is due to the Cauchy-Schwarz inequality, and (23) is due to (20). The equality holds if and only if $\sigma_1(\mathbf{A}) = \sigma_2(\mathbf{A}) = \dots = \sigma_r(\mathbf{A}) = 1$, which is equivalent to the condition that $\mathbf{A}$ is real orthogonal. We note that the real orthogonal matrix $\mathbf{A} = \mathbf{I}$ is a feasible point for (14). Therefore, the upper bound by (18) and (19), is achievable by the optimization in (14). Thus, $\mathbf{A}_*$ is optimal solution of (14a) only if it is real orthogonal. What remains to be shown is that all of the global optima of (14) are real orthogonal matrices with the desired form.

The third step: *We use the sufficient scattering condition (PMF.SS.ii) together with the orthogonality of optimal $\mathbf{A}_*$ to show that any optimal point of (14) has the desired form, i.e., the product of a permutation and a diagonal matrix.*

Given $\mathbf{q}_i \in \text{ext}(\mathcal{B}_\infty)$, we have $\|\mathbf{q}_i\|_2 = \sqrt{r}$. Since optimal $\mathbf{A}_*$ is a real orthogonal matrix, we have $\mathbf{A}_*^T\mathbf{q}_i \in \text{bd}(\sqrt{r}\mathcal{B}_2)$, $\forall \mathbf{q}_i \in \text{ext}(\mathcal{B}_\infty)$. Combining it with (16), we obtain $\mathbf{A}_*^T\mathbf{q}_i \in \text{conv}(\mathbf{S}_g)^* \cap \text{bd}(\sqrt{r}\mathcal{B}_2)$, $\forall \mathbf{q}_i \in \text{ext}(\mathcal{B}_\infty)$. Since the assumption (PMF.SS.ii) restricts $\text{conv}(\mathbf{S}_g)^* \cap \text{bd}(\sqrt{r}\mathcal{B}_2)$ to $\text{ext}(\mathcal{B}_\infty)$, the equivalent condition for the global optimality of $\mathbf{A}_*$ for (14) can be written as

$$\mathbf{A}_*^T\mathbf{A}_* = \mathbf{I}, \tag{24a}$$

$$\mathbf{A}_*^T\mathbf{q}_i \in \text{ext}(\mathcal{B}_\infty) \quad \forall \mathbf{q}_i \in \text{ext}(\mathcal{B}_\infty). \tag{24b}$$

In other words, $\mathbf{A}_*$ is a global optimum if and only if it is real orthogonal and its transpose maps the vertices of $\mathcal{B}_\infty$ in Fig. 5(b) to itself. Using (24), we conclude that $\mathbf{A}_*$ has only one nonzero entry in each column(row) as follows:

i. (24a) implies $\|\mathbf{A}_{*:,i}\|_2 = 1$,
ii. (24b) implies $\|\mathbf{A}_{*:,i}\|_1 = 1$, since for any column of $\mathbf{A}_*$, $\text{sign}(\mathbf{A}_{*:,i}) \in \text{ext}(\mathcal{B}_\infty)$. Therefore, $\mathbf{A}_*^T\text{sign}(\mathbf{A}_{*:,i}) = \mathbf{q}_j$, for some $\mathbf{q}_j \in \text{ext}(\mathcal{B}_\infty)$, due to (24b). The $i^{\text{th}}$ row of $\mathbf{q}_j$ is $(\mathbf{A}_{*:,i})^T\text{sign}(\mathbf{A}_{*:,i}) = \|\mathbf{A}_{*:,i}\|_1$. Since all components of $\mathbf{q}_j$ have magnitude 1, we have $\|\mathbf{A}_{*:,i}\|_1 = 1$.

The statements (i) and (ii) above are true if and only if $\mathbf{A}_{*:,i}$ has only one non-zero element with unit magnitude. Since the rows of $\mathbf{A}_*$ are orthonormal, the global optima characterization

is given by $\mathbf{A}_* = \mathbf{D}\mathbf{\Pi}$, where $\mathbf{D}$ and $\mathbf{\Pi}$ are as stated in the theorem. ■

### C. Antisparse Nonnegative PMF

As defined in item (iii) in Section II-A, this special case refers to the polytope choice

$$\begin{aligned}\mathcal{P} = \mathcal{B}_{\infty,+} =& \ \{\mathbf{x} \mid \mathbf{0} \le \mathbf{x} \le \mathbf{1}, \mathbf{x} \in \mathbb{R}^r_+\},\\ =& \ \mathcal{B}_\infty \cap \mathbb{R}^r_+ = 0.5\mathcal{B}_\infty + 0.5\mathbf{1},\end{aligned}$$

i.e., in essence, a scaled and translated version of $\mathcal{B}_\infty$. We apply the same affine transformation to $\mathcal{E}_{\mathcal{B}_\infty}$ to obtain the MVIE of $\mathcal{B}_{\infty,+}$: $\mathcal{E}_{\mathcal{B}_{\infty,+}} = 0.5\mathcal{E}_{\mathcal{B}_\infty} + 0.5\mathbf{1}$. Therefore, the parameters of the MVIE corresponding to this polytope are given by $\mathbf{C}_{\mathcal{B}_{\infty,+}} = 0.5\mathbf{I}$ and $\mathbf{g}_{\mathcal{B}_{\infty,+}} = 0.5\mathbf{1}$.

This case is a special case of antisparse BCA covered in [6], in which the existing identifiability condition is based on the vertex inclusion assumption. In this section, we provide the characterization of the identifiability condition for nonnegative antisparse PMF based on the weaker sufficient scattering assumption through the following theorem.

*Theorem 4:* Given the general PMF setting outlined in Section II-A, if $\mathbf{S}_g$ is a sufficiently scattered factor for *antisparse nonnegative PMF* according to Definition II.2, then all global optima $\mathbf{H}_*, \mathbf{S}_*$ of the Det-Max optimization problem in (7) with $\mathcal{P} = B_{\infty,+}$ satisfy

$$\begin{aligned}\mathbf{H}_* =& \ \mathbf{H}_g\mathbf{\Pi}^T,\\ \mathbf{S}_* =& \ \mathbf{\Pi}\mathbf{S}_g,\end{aligned}$$

where $\mathbf{\Pi} \in \mathbb{R}^{r\times r}$ is a permutation matrix.

*Proof:* Following the same treatment in the proof of Theorem 2, we can write the equivalent form of the optimization in (7) for nonnegative antisparse PMF as

$$\underset{\mathbf{A}\in\mathbb{R}^{r\times r}}{\text{maximize}} \quad |\det(\mathbf{A})| \tag{25a}$$

$$\text{subject to} \quad 0 \le \mathbf{A}_{i,:}\mathbf{S}_{g_{:,j}} \le 1, \quad i = 1,\dots,r \quad j = 1,\dots,N. \tag{25b}$$

The proof consists of three major steps.

The first step: *We use (25b) and (PMF.SS.i) to show that the columns(rows) of any feasible* $\mathbf{A}$ *of (25) lie in* $\mathcal{B}_2$.

Due to (PMF.SS.i), which is $\mathcal{E}_{\mathcal{B}_{\infty,+}} \subset \text{conv}(\mathbf{S}_g)$, for all $\mathbf{s} \in \mathcal{E}_{\mathcal{B}_{\infty,+}}$ the constraint (25b) holds. Therefore, for any $i \in \{1,\dots,r\}$, we have

$$0 \le 0.5\mathbf{A}_{i,:}\mathbf{u} + 0.5\mathbf{A}_{i,:}\mathbf{1} \le 1, \quad \forall \mathbf{u} \in \mathcal{B}_2, \tag{26}$$

where we used $\mathcal{E}_{\mathcal{B}_{\infty,+}} = \{0.5\mathbf{u} + 0.5\mathbf{1}, \|\mathbf{u}\|_2 \le 1\}$. If we substitute $\mathbf{u} = \frac{(\mathbf{A}_{i,:})^T}{\|\mathbf{A}_{i,:}\|_2}$ and $\mathbf{u} = -\frac{(\mathbf{A}_{i,:})^T}{\|\mathbf{A}_{i,:}\|_2}$ in (26), we obtain

$$0 \le 0.5\|\mathbf{A}_{i,:}\|_2 + 0.5\mathbf{A}_{i,:}\mathbf{1} \ \le \ 1, \text{ and,} \tag{27}$$

$$0 \le -0.5\|\mathbf{A}_{i,:}\|_2 + 0.5\mathbf{A}_{i,:}\mathbf{1} \ \le \ 1, \tag{28}$$

respectively, for all $i = 1,\dots,r$. The summation of (27) and the sign reversed (28) lead to $\|\mathbf{A}_{i,:}\|_2 \le 1$ for all rows of $\mathbf{A}$. In other words, the rows(columns) of any feasible $\mathbf{A}$ should be in $\mathcal{B}_2$.

The second step: *We now show any optimal solution* $\mathbf{A}_*$ *of (25) should be a real orthogonal matrix.*

Using Hadamard's inequality and $\|\mathbf{A}_{i,:}\|_2 \le 1$ from the previous step, we can place an upper bound on the objective function in (25a) as $|\det(\mathbf{A})| \le \|\mathbf{A}_{1,:}\|_2\|\mathbf{A}_{2,:}\|_2 \dots \|\mathbf{A}_{r,:}\|_2 \le 1$. This bound is achieved if and only if the rows form an orthonormal set. Therefore, any optimal solution $\mathbf{A}_*$ of (25) is a real orthogonal matrix.

The third step: *We use (PMF.SS.ii) to show that all global optima of (25) are permutation matrices.*

If we substitute a real orthogonal $\mathbf{A}_*$ in (28), we obtain $\mathbf{A}_{*i,:}\mathbf{1} \ge 1$. Combining this inequality with (25b), we can write $2\mathbf{A}_{*i,:}\mathbf{S}_{g_{:,j}} \le 2 \le 1 + \mathbf{A}_{*i,:}\mathbf{1}$. Reorganizing the left and right terms of this expression, we obtain $2\mathbf{A}_{*i,:}(\mathbf{S}_{g_{:,j}} - 0.5\mathbf{1}) \le 1$ for all the columns of $\mathbf{S}_g$. Based on this inequality, we conclude that $2\mathbf{A}_{*i,:}$ lies in $\text{conv}(\mathbf{S}_g)^{*,0.5\mathbf{1}}$ for all the rows of $\mathbf{A}_*$. Furthermore, since $\mathbf{A}_*$ is also real orthogonal, $2\mathbf{A}_{*i,:} \in \text{conv}(\mathbf{S}_g)^{*,0.5\mathbf{1}} \cap \text{bd}(2\mathcal{B}_2)$. Due to the sufficiently scattered constraint (PMF.SS.ii), which is $\text{conv}(\mathbf{S}_g)^{*,0.5\mathbf{1}} \cap \text{bd}(2\mathcal{B}_2) = \text{ext}(2\mathcal{B}_1)$, $\mathbf{A}_{*i,:} \in \text{ext}(\mathcal{B}_1)$. Furthermore, due to the condition $\mathbf{A}_{*i,:}\mathbf{1} \ge 1$, $\mathbf{A}_*$ is a global optimum if and only if its rows are positive standard basis vectors that are orthogonal to each other, i.e., $\mathbf{A}_*$ is a permutation matrix. ■

### D. Sparse Nonnegative PMF

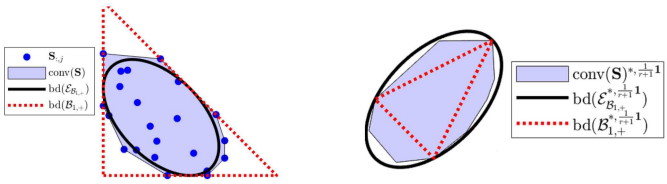
As defined in item (iv) in Section II-A, the polytope for sparse nonnegative factors is given by

$$\mathcal{P} = \mathcal{B}_{1,+} = \{\mathbf{x} \mid \mathbf{x} \ge \mathbf{0}, \mathbf{1}^T\mathbf{x} \le 1\} = \mathcal{B}_1 \cap \mathbb{R}^r_+. \tag{29}$$

There are various matrix factorization algorithms combining nonnegativity and sparsity. However, we are not aware of any approaches that make explicit mention of the polytope $\mathcal{B}_{1,+}$, and provide the corresponding identifiability conditions. In this section, we provide identifiability results for this polytope, again based on the sufficient scattering assumption in Definition II.2. For this purpose, we derive the MVIE $\mathcal{E}_{\mathcal{B}_{1,+}}$ in Appendix B. The MVIE derivation in this case is relatively more involved compared to other special cases (in items (i)-(iii) in Section II-A). This is due to the fact the MVIE in this case is not spherical, and therefore, Theorem 1 is not applicable. The inscribed ellipsoid parameters obtained in Appendix B are

$$\begin{aligned}\mathbf{C}_{\mathcal{B}_{1,+}} &= \frac{1}{\sqrt{r}}\left(\frac{1}{\sqrt{r+1}}\mathbf{I} - \frac{\sqrt{r+1}-1}{r^2+r}\mathbf{1}\mathbf{1}^\mathbf{T}\right),\\ \mathbf{g}_{\mathcal{B}_{1,+}} &= \frac{1}{r+1}\mathbf{1}.\end{aligned} \tag{30}$$

To obtain the polar of the polytope $\mathcal{P}$, we first use the canonical form for the polytope shifted by $-\mathbf{g}_{\mathcal{B}_{1,+}}$ in the form $\mathcal{P} - \mathbf{g}_{\mathcal{B}_{1,+}} = \{\mathbf{x} \mid \mathbf{1}^T(\mathbf{x} + \frac{1}{r+1}\mathbf{1}) \le 1, -\mathbf{e}_i^T(\mathbf{x} + \frac{1}{r+1}\mathbf{1}) \le 0, i \in \{1,\dots,r\}\}$. After some algebraic manipulations, we can convert it into standard form (58) in Appendix A: $\mathcal{P} - \mathbf{g}_{\mathcal{B}_{1,+}} = \{\mathbf{x} \mid (r+1)\mathbf{1}^T\mathbf{x} \le 1, -(r+1)\mathbf{e}_i^T\mathbf{x} \le 1, i \in \{1,\dots,r\}\}$. Therefore, based on the procedure in Appendix A,




(a) (b)

Fig. 6. Sufficiently scattered case example for 2D sparse nonnegative PMF: (a) sample domain, (b) polar domain.

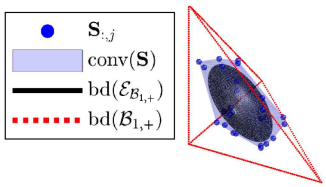



Fig. 7. 3D Sufficiently scattered case example.

the polar of $\mathcal{B}_{1,+}$ can be written as

$$\mathcal{B}_{1,+}^{*,\frac{1}{r+1}\mathbf{1}} = \text{conv}(-(r+1)\mathbf{e}_1,\ldots,-(r+1)\mathbf{e}_r,(r+1)\mathbf{1}), \tag{31}$$

which is a polytope with $(r+1)$ vertices (Note that $\mathcal{B}_{1,+}$ has $(r+1)$ faces).

Fig. 6(a) illustrates $\mathcal{B}_{1,+}$, its MVIE and an example set of sufficiently scattered samples for $r=2$. These samples clearly do not contain the vertices of $\mathcal{B}_{1,+}$. The convex hull of the samples contains the nonspherical MVIE and is tangent to its boundary at the points where the polytope is tangent.

The picture corresponding to the polars of the sets in Fig. 6(a) is provided in Fig. 6(b). The polar of the polytope $\mathcal{B}_{1,+}$ is a triangular region, the boundary of which is shown with dashed-red lines. We note that as expected from (PMF.SS.ii), conv$(\mathbf{S})^{*,\frac{1}{r+1}\mathbf{1}}$ intersects the boundary of $\mathcal{E}^{*,\frac{1}{r+1}\mathbf{1}}_{\mathcal{B}_{1,+}}$ only at the vertices of $\mathcal{B}^{*,\frac{1}{r+1}\mathbf{1}}_{1,+}$.

Fig. 7 illustrates a sufficiently scattered distribution of columns of $\mathbf{S}$ for the three-dimensional case.

We characterize the identifiability condition for the nonnegative sparse PMF problem through the following theorem.

*Theorem 5:* Given the general PMF setting outlined in Section II-A, if $\mathbf{S}_g$ is a sufficiently scattered factor for *sparse nonnegative PMF* according to Definition II.2, then all global optima $\mathbf{H}_*, \mathbf{S}_*$ of the *Det-Max optimization problem in (7) for* $\mathcal{P}=\mathcal{B}_{1,+}$ *satisfy*

$$\mathbf{H}_* = \mathbf{H}_g\mathbf{\Pi}^T, \tag{32}$$

$$\mathbf{S}_* = \mathbf{\Pi}\mathbf{S}_g, \tag{33}$$

*where* $\mathbf{\Pi}\in\mathbb{R}^{r\times r}$ *is a permutation matrix.*

*Proof:* Following the same arguments in the proof of Theorem 2, we start with the equivalent determinant maximization formulation for the nonnegative sparse PMF:

$$\underset{\mathbf{A}\in\mathbb{R}^{r\times r}}{\text{maximize}} \quad |\det(\mathbf{A})| \tag{34a}$$

$$\text{subject to} \quad 0\le \mathbf{A}_{i,:}\mathbf{S}_{g_{:,j}}\le 1, \quad i=1,\ldots,r \quad j=1,\ldots,N \tag{34b}$$

$$\mathbf{1}^T\mathbf{A}\mathbf{S}_{g_{:,j}}\le 1 \quad j=1,\ldots,N. \tag{34c}$$

The proof consists of four steps.

The first step: *We use the polytopic constraints (34b) and (34c), and the sufficient scattering condition (PMF.SS.i) to show that optimal* $\mathbf{A}_*$ *is a real orthogonal matrix with* $\mathbf{1}^T\mathbf{A}_*\mathbf{1}=r$ *and* $\|\mathbf{A}_{*i,:}\|_2=\mathbf{A}_{*i,:}\mathbf{1}=1$ *for all* $i=1,\ldots,r$.

Based on the sufficient scattering condition (PMF.SS.i), which is conv$(\mathbf{S}_g)\supset\mathcal{E}_{\mathcal{B}_{1,+}}$ (as illustrated in Figs. 6(a) and 7), any member of the MVIE should satisfy the constraints (34b) and (34c). Therefore, we have

$$0\le \mathbf{A}_{i,:}\mathbf{C}_{\mathcal{B}_{1,+}}\mathbf{u}+\frac{1}{r+1}\mathbf{A}_{i,:}\mathbf{1}\le 1, \quad \forall\mathbf{u}\in\mathcal{B}_2, \tag{35}$$

$$\mathbf{1}^T\mathbf{A}\mathbf{C}_{\mathcal{B}_{1,+}}\mathbf{u}+\frac{1}{r+1}\mathbf{1}^T\mathbf{A}\mathbf{1}\le 1, \quad \forall\mathbf{u}\in\mathcal{B}_2, \tag{36}$$

Specifically, if we substitute $\mathbf{u}=-\frac{(\mathbf{A}_{i,:}\mathbf{C}_{\mathcal{B}_{1,+}})^T}{\|\mathbf{C}_{\mathbf{A}_{i,:}\mathcal{B}_{1,+}}\|_2}$ in (35), we obtain $\|\mathbf{A}_{i,:}\mathbf{C}_{\mathcal{B}_{1,+}}\|_2\le\frac{1}{r+1}\mathbf{A}_{i,:}\mathbf{1}$, which further implies

$$\mathbf{A}_{i,:}\mathbf{C}^2_{\mathcal{B}_{1,+}}(\mathbf{A}_{i,:})^T\le\left(\frac{1}{r+1}\mathbf{A}_{i,:}\mathbf{1}\right)^2, \tag{37}$$

for all $i=1,\ldots,r$. Inserting (77) for $\mathbf{C}^2_{\mathcal{B}_{1,+}}$ (derived in Appendix B) in (37), we obtain

$$\frac{1}{r(r+1)}\left(\|\mathbf{A}_{i,:}\|_2^2-\frac{(\mathbf{A}_{i,:}\mathbf{1})^2}{r+1}\right)\le\frac{(\mathbf{A}_{i,:}\mathbf{1})^2}{(r+1)^2}.$$

Simplifying this expression, we have

$$\|\mathbf{A}_{i,:}\|_2^2\le(\mathbf{A}_{i,:}\mathbf{1})^2, \tag{38}$$

for all $i=1,\ldots,r$. If we substitute $\mathbf{u}=\mathbf{0}$ in (35), we obtain $\mathbf{A}_{i,:}\mathbf{1}\ge 0$. Therefore, we can rewrite (38) as $\|\mathbf{A}_{i,:}\|_2\le\mathbf{A}_{i,:}\mathbf{1}$. Furthermore, if we substitute $\mathbf{u}=\frac{1}{\sqrt{r}}\mathbf{1}$ in (36), we have

$$\frac{1}{\sqrt{r}}\mathbf{1}^T\mathbf{A}\mathbf{C}_{\mathcal{B}_{1,+}}\mathbf{1}+\frac{1}{r+1}\mathbf{1}^T\mathbf{A}\mathbf{1}\le 1. \tag{39}$$

Inserting (30) for $\mathbf{C}_{\mathcal{B}_{1,+}}$ in (39), and after applying some simplifications, we obtain

$$\mathbf{1}^T\mathbf{A}\mathbf{1}\le r. \tag{40}$$

Replacing the constraints of (34) with (38) and (40), we obtain the following optimization problem, the solution of which provides an upper bound for (34):

$$\underset{\mathbf{A}\in\mathbb{R}^{r\times r}}{\text{maximize}} \quad |\det(\mathbf{A})| \tag{41a}$$

$$\text{subject to} \quad \mathbf{A}_{i,:}\mathbf{1}\ge\|\mathbf{A}_{i,:}\|_2, \quad i=1,\ldots,r \tag{41b}$$

$$\mathbf{1}^T\mathbf{A}\mathbf{1}\le r. \tag{41c}$$

The optimal value of the objective function (41a) is bounded from above by

$$|\det(\mathbf{A})|\le\|\mathbf{A}^T_{1,:}\|_2\|\mathbf{A}^T_{2,:}\|_2\ldots\|\mathbf{A}^T_{r,:}\|_2, \tag{42}$$

$$\le(\mathbf{A}_{1,:}\mathbf{1})(\mathbf{A}_{2,:}\mathbf{1})\ldots(\mathbf{A}_{r,:}\mathbf{1}), \tag{43}$$

$$\leq (\frac{1}{r}\mathbf{1}^T\mathbf{A}\mathbf{1})^r, \tag{44}$$

$$\leq 1, \tag{45}$$

where the expression in (42) is Hadamard's inequality, (44) is due to the arithmetic-geometric mean inequality, and (45) is due to (41c). Therefore, for any optimal solution $\mathbf{A}_*$ of (34), we have

$$|\det(\mathbf{A}_*)| \leq 1. \tag{46}$$

The equality in (46) is achieved if and only if $\mathbf{A}_*$ is a real orthogonal matrix with

$$\mathbf{1}^T\mathbf{A}_*\mathbf{1} = r, \text{ and } \|\mathbf{A}_{*i,:}\|_2 = \mathbf{A}_{*i,:}\mathbf{1} = 1, i = 1, \ldots, r. \tag{47}$$

Furthermore, since the identity matrix, $\mathbf{I}$, is clearly a feasible point for (34), its determinant corresponds to a lower bound. Therefore, for any optimal solution $\mathbf{A}_*$ of (34), we obtain

$$|\det(\mathbf{A}_*)| \geq 1. \tag{48}$$

Combining the lower bound in (48) and the upper bound in (46) and with the equality conditions in (42)–(45), we conclude that $\mathbf{A}_*$ is an optimal solution for the problem in (34) only if $\mathbf{A}_*$ is a real orthogonal matrix satisfying (47).

The second step: *We show that the scaled rows of the global optima of (34) are in conv*$(\mathbf{S}_g)^{*,\frac{1}{r+1}\mathbf{1}}$.

A global optimal point $\mathbf{A}_*$ satisfies (34b), therefore, we can write $\mathbf{A}_{*i,:}\mathbf{S}_{g_{:,j}} \geq 0$ for all possible $(i, j)$ pairs. Multiplying this expression by $-(r+1)$, we obtain $-(r+1)\mathbf{A}_{*i,:}\mathbf{S}_{g_{:,j}} \leq 0$. From the previous step, we also have $\mathbf{A}_{*i,:}\mathbf{1} = 1$. Combining these expressions, we can write

$$-(r+1)\mathbf{A}_{*i,:}\left(\mathbf{S}_{g_{:,j}} - \frac{1}{r+1}\mathbf{1}\right) \leq 1,$$

which implies $-(r+1)(\mathbf{A}_{*i,:})^T \in \text{conv}(\mathbf{S}_g)^{*,\frac{1}{r+1}\mathbf{1}}$ for all $i \in \{1, \ldots, r\}$. We will show in the fourth step that these scaled rows of $\mathbf{A}_*$ are also the vertices of $\mathcal{B}_{1,+}^{*,\frac{1}{r+1}\mathbf{1}}$.

Furthermore, scaling (34c) with $(r+1)$, we can write $((r+1)\mathbf{A}_*^T\mathbf{1})^T\mathbf{S}_{g_{:,j}} \leq r+1$ for all $j \in \{1, \ldots, N\}$. Replacing $r$ on the right with $\mathbf{1}^T\mathbf{A}_*\mathbf{1}$ (based on (47)), we have

$$((r+1)\mathbf{A}_*^T\mathbf{1})^T\mathbf{S}_{g_{:,j}} \leq \mathbf{1}^T\mathbf{A}_*\mathbf{1} + 1. \tag{49}$$

Reorganizing (49), leads to

$$((r+1)\mathbf{A}_*^T\mathbf{1})^T\left(\mathbf{S}_{g_{:,j}} - \frac{1}{r+1}\mathbf{1}\right) \leq 1,$$

from which we conclude $(r+1){\mathbf{A}_*}^T\mathbf{1} \in \text{conv}(\mathbf{S}_g)^{*,\frac{1}{r+1}\mathbf{1}}$.

The third step: *We show that the* $-(r+1)$ *scaled rows of the global optima* $\mathbf{A}_*$ *of (34) are on the boundary of the polar ellipsoid* $\mathcal{E}_{\mathcal{P}}^{*,\frac{1}{r+1}\mathbf{1}}$.

From Appendix A, the polar of the ellipsoid in (8) is given by

$$\mathcal{E}_{\mathcal{P}}^{*,\mathbf{g}_P} = \{\mathbf{C}_P^{-1}\mathbf{u} \mid \|\mathbf{u}\|_2 \leq 1, \mathbf{u} \in \mathbb{R}^r\},$$

the boundary of which can be written as

$$\text{bd}(\mathcal{E}_{\mathcal{P}}^{*,\mathbf{g}_P}) = \{\mathbf{x} \mid \|\mathbf{x}^T\mathbf{C}_P\|_2 = 1\}. \tag{50}$$

To check whether $-(r+1)(\mathbf{A}_{*i,:})^T$ is on the boundary of the polar of the MVIE, we evaluate the norm expression in (50):

$$\|(r+1)\mathbf{A}_{*i,:}\mathbf{C}_{\mathcal{B}_{1,+}}\|_2^2 = (r+1)^2\mathbf{A}_{*i,:}\mathbf{C}_{\mathcal{B}_{1,+}}^2{\mathbf{A}_{*i,:}}^T,$$

$$= \frac{r+1}{r}\left(\|\mathbf{A}_{*i,:}\|_2^2 - \frac{(\mathbf{A}_{*i,:}\mathbf{1})^2}{r+1}\right), \tag{51}$$

$$= 1. \tag{52}$$

We used (77) to obtain (51), and (47) to simplify (51) into (52). As a result, $-(r+1)(\mathbf{A}_{*i,:})^T \in \text{bd}(\mathcal{E}_{\mathcal{P}}^{*,\mathbf{g}_P})$. Therefore, the rows of the global optimum $\mathbf{A}_*$ lie in the boundary of the polar domain ellipsoid illustrated in Fig. 6(b).

Similarly, we can show $(r+1){\mathbf{A}_*}^T\mathbf{1} \in \text{bd}(\mathcal{E}_{\mathcal{P}}^{*,\mathbf{g}_P})$ through

$$\|(r+1)\mathbf{1}^T\mathbf{A}_*\mathbf{C}_{\mathcal{B}_{1,+}}\|_2^2 = (r+1)^2\mathbf{1}^T\mathbf{A}_*\mathbf{C}_{\mathcal{B}_{1,+}}^2{\mathbf{A}_*}^T\mathbf{1},$$

$$= \frac{r+1}{r}\left(\mathbf{1}^T\mathbf{A}_*{\mathbf{A}_*}^T\mathbf{1} - \frac{(\mathbf{1}^T\mathbf{A}_*\mathbf{1})^2}{r+1}\right), \tag{53}$$

$$= 1. \tag{54}$$

We again used (77) to obtain (53), and (47) to simplify (53) into (54).

The fourth step: *We combine the results of the previous steps and (PMF.SS.ii) to show that all global optima for (34) are permutation matrices.*

Combining the results of the second and third steps of the proof, we deduce that $-(r+1)(\mathbf{A}_{*i,:})^T \in \text{conv}(\mathbf{S}_g)^{*,\frac{1}{r+1}\mathbf{1}} \cap \text{bd}(\mathcal{E}_{\mathcal{P}}^{*,\frac{1}{r+1}\mathbf{1}})$ for all $i \in \{1, \ldots, r\}$ and $(r+1){\mathbf{A}_*}^T\mathbf{1} \in \text{conv}(\mathbf{S}_g)^{*,\frac{1}{r+1}\mathbf{1}} \cap \text{bd}(\mathcal{E}_{\mathcal{P}}^{*,\frac{1}{r+1}\mathbf{1}})$. The sufficient scattering condition (PMF.SS.ii) dictates that $\text{conv}(\mathbf{S}_g)^{*,\frac{1}{r+1}\mathbf{1}} \cap \text{bd}(\mathcal{E}_{\mathcal{B}_{1,+}}^{*,\frac{1}{r+1}\mathbf{1}}) = \text{ext}(\mathcal{B}_{1,+}^{*,\frac{1}{r+1}\mathbf{1}})$. Due to the canonical description in (31),

$$\text{ext}(\mathcal{B}_{1,+}^{*,\frac{1}{r+1}\mathbf{1}}) = \{-(r+1)\mathbf{e}_1, \ldots, -(r+1)\mathbf{e}_r, (r+1)\mathbf{1}\}. \tag{55}$$

This condition is visible in Fig. 6(b) as a triangle intersecting the boundaries of the ellipsoid and $\text{conv}(\mathbf{S}_g)^{*,\frac{1}{r+1}\mathbf{1}}$ at its $(r+1)$ vertices only. Therefore, we conclude that the set $\{(r+1)\mathbf{A}_*^T\mathbf{1}, (r+1)(\mathbf{A}_{*i,:})^T, i \in \{1, \ldots, r\}\}$ are all vertices of $\mathcal{B}_{1,+}^{*,\frac{1}{r+1}\mathbf{1}}$ given by (55). This equivalence together with (47) lead to the conclusion that the rows of $\mathbf{A}_*$ are the standard basis vectors; therefore, $\mathbf{A}_*$ is a permutation matrix. ■

## IV. Generalized PMF

In the previous section, we concentrated on four particular polytope examples that correspond to the cross combinations of sparse/antisparse and nonnegative/signed attributes. For these examples, the resulting features are globally applied to all latent vector components. We showed that PMF settings based on these particular polytopes are always identifiable using Det-Max optimization problem in (7) if the sufficient scattering condition in Definition II.2 holds.

It is interesting to explore whether we can develop alternative polytopes that lead to Det-Max identifiable PMF generative settings as defined in Definition II.1. In this section, we provide a positive answer and show that the set of polytopes that qualify for the PMF framework is infinitely rich [21]. In particular, we show that, as long as a polytope conforms with a specific symmetry restriction, it would always lead to identifiable generative models under the sufficiently scattering assumption. We now formalize the definition of "identifiable polytopes".

*Definition IV.1: Identifiable Polytopes with respect to the Det-Max Criterion:* We refer to a polytope $\mathcal{P}$ as "Det-Max identifiable" (or simply "identifiable") if all generative models in (1) and (2) based on sufficiently scattered samples from $\mathcal{P}$, according to Definition II.2, are "Det-Max identifiable"

We show that the identifiable polytopes should satisfy a particular symmetry condition, which is laid out in the following definition:

*Definition IV.2: Permutation-and/or-Sign-Only Invariant Set.* A set $\mathcal{F}$ is called permutation-and/or sign-only invariant if and only if any linear transformation that satisfies

$$\mathbf{A}(\mathcal{F}) = \mathcal{F}$$

has the form $\mathbf{A} = \mathbf{\Pi D}$, where $\mathbf{\Pi} \in \mathbb{R}^{r\times r}$ is a permutation matrix, and $\mathbf{D} \in \mathbb{R}^{r\times r}$ is a full rank diagonal matrix with its diagonal entries in $\{-1, 1\}$.

We note that Definition IV.2 defines a symmetry restriction: a set satisfying the condition in this definition can not be mapped to itself under any linear transformation other than the combination of permutation-sign scaling transformations. The following theorem, the proof of which is provided in [21] and Appendix C, characterizes all "Det-Max identifiable" polytopes based on the symmetry restriction in Definition IV.2 [21]:

*Theorem 6 (Det-Max Identifiable Polytope):* A polytope $\mathcal{P}$ is "Det-Max identifiable" if and only if its set of vertices, $ext(\mathcal{P})$, is a permutation-and/or-sign-only invariant set.

The symmetry condition imposed by this theorem is satisfied by infinitely many polytopes. The abundance of polytope choices implies a degree of freedom for defining a diverse set of feature descriptions for latent vectors. In particular, this diversity can be exploited to render latent vector features with heterogeneous structures without resorting to any stochastic assumption such as independence. This property contrasts with the existing deterministic matrix factorization frameworks, such as NMF, SCA, and BCA, which impose a common attribute, such as nonnegativity, antisparsity or sparsity, over the whole vector. Using the PMF framework, it is possible to choose only a subset of the components to be nonnegative. Furthermore, we can impose sparsity constraints on potentially overlapping multiple subsets of components. In the numerical examples section (Section VI-A), we provide an example of such a heterogeneous latent vector design.

## V. Algorithm

The main emphasis of the current article is laying out the PMF framework and the corresponding identifiability analysis.

**Algorithm 1:** Det-Min Algorithm for PMF.

**Input:** $\mathbf{Y}$; $r$; Initial $\mathbf{H}, \mathbf{S}$; $\tau$.

$t = 0$;
$\mathbf{X}^{(t)} = \mathbf{S}, \mathbf{F}^{(t)} = \mathbf{I}, \mathbf{H}^{(t)} = \mathbf{H}, \mathbf{S}^{(t)} = \mathbf{S}, q^{(t)} = 1$;
**repeat**
select $L^{(t)}$;
$\mathbf{S}_{:,l}^{(t+1)} \leftarrow P_{\mathcal{P}}(\mathbf{X}_{:,l}^{(t)} - (\mathbf{H}^{(t)})^T(\mathbf{Y}_{:,l} - \mathbf{H}^{(t)}\mathbf{X}_{:,l}^{(t)}))$ for $l = 1, \ldots, N$;
$q^{(t+1)} \leftarrow \frac{1+\sqrt{1+(q^{(t)})^2}}{2}$;
$\mathbf{X}_{:,l}^{(t+1)} \leftarrow \mathbf{S}_{:,l}^{(t+1)} + \frac{q^{(t)}-1}{q^{(t+1)}}(\mathbf{S}_{:,l}^{(t+1)} - \mathbf{S}_{:,l}^{(t)})$ for $l = 1, \ldots, N$;
$\mathbf{H}^{(t+1)} \leftarrow \mathbf{Y}(\mathbf{S}^{(t+1)})^T(\mathbf{S}^{(t+1)}(\mathbf{S}^{(t+1)})^T + \lambda\mathbf{F}^{(t)})$;
$t \leftarrow t + 1$;
$\mathbf{F}^{(t)} \leftarrow ((\mathbf{H}^{(t)})^T\mathbf{H}^{(t)} + \tau\mathbf{I})^{-1}$.
**until** some stopping criterion is reached.

**Output:** $\mathbf{H}^{(t)}$;$\mathbf{S}^{(t)}$.

To illustrate its use, we adopt the iterative algorithm in [32] which is originally proposed for the SSMF framework.

We start by introducing the Determinant Minimization Det-Min problem, equivalent to the Det-Max optimization problem in (7) under the equality constraint in (7b) [2]:

$$\underset{\mathbf{H}\in\mathbb{R}^{M\times r},\mathbf{S}\in\mathbb{R}^{r\times N}}{\text{minimize}} \quad \det(\mathbf{H}^T\mathbf{H}) \tag{56a}$$

$$\text{subject to} \quad \mathbf{Y} = \mathbf{HS} \tag{56b}$$

$$\mathbf{S}_{:,j} \in \mathcal{P}, \quad j = 1, \ldots, N. \tag{56c}$$

Similar to [32], we employ the Lagrangian optimization,

$$\underset{\mathbf{H},\mathbf{S}}{\text{minimize}} \quad \|\mathbf{Y} - \mathbf{HS}\|_F^2 + \lambda \log\det(\mathbf{H}^T\mathbf{H} + \tau\mathbf{I}) \tag{57a}$$

$$\text{subject to} \quad \mathbf{S}_{:,j} \in \mathcal{P}, \quad j = 1, \ldots, N. \tag{57b}$$

corresponding to (56), where $\tau > 0$ is a hyperparameter to ensure that the objective function is bounded from below.

The corresponding steps are provided in Algorithm 1, which is the algorithm in [32], except that the projection onto the unit simplex is replaced with the projection to the polytope, $P_{\mathcal{P}}(\cdot)$. The following are examples of this projection operator:

- *Antisparse Case:* $\mathbf{X} = P_{\mathcal{B}_\infty}(\bar{\mathbf{X}})$ defines an elementwise projection operator to $\mathcal{B}_\infty$, which can be written as

$$X_{ij} = \begin{cases} \bar{X}_{ij} & \text{if } |\bar{X}_{ij}| < 1, \\ \text{sign}(\bar{X}_{ij}) & \text{otherwise.} \end{cases}$$

- *Sparse Case:* For $P_{\mathcal{B}_1}(\cdot)$, the projection onto the $\ell_1$-norm-ball has no closed-form solution; however, efficient iterative algorithms, such as [41], exist.
- *Antisparse Nonnegative Case:* The projection operator is a simple modification of $P_{\mathcal{B}_\infty}(\cdot)$, where elementwise projections are performed over [0,1] instead of $[-1, 1]$.
- *Sparse Nonnegative Case:* We can simplify the projection operator $\mathcal{B}_1$ to obtain the projection onto $\mathcal{B}_{1,+}$ (see [41]).

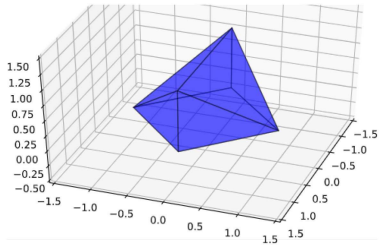


Fig. 8. Polytope $\mathcal{P}_{\text{ex}}$.

## VI. Numerical Experiments

### A. Polytope With Local Features

To illustrate the feature shaping flexibility provided by the PMF framework, we consider the following example polytope:

$$\mathcal{P}_{\text{ex}} = \left\{ \mathbf{x} \in \mathbb{R}^3 \;\middle|\; \begin{array}{l} x_1, x_2 \in [-1,1], x_3 \in [0,1], \\ \left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_1 \leq 1, \left\| \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} \right\|_1 \leq 1 \end{array} \right\},$$

which corresponds to the following local attributes:

- $x_3$ is nonnegative and $x_1, x_2$ are signed; and
- $[\, x_1 \; x_2 \,]^T$ and $[\, x_2 \; x_3 \,]^T$ are sparse subvectors.

The polytope $\mathcal{P}_{\text{ex}}$, shown in Fig. 8, has 6 vertices placed in the columns of the following matrix:

$$\mathbf{V}_{\mathcal{P}_{\text{ex}}} = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Checking all of the possible permutations of the columns of this matrix reveals that the vertex set satisfies the desired symmetry restriction; therefore, $\mathcal{P}_{\text{ex}}$ is identifiable. Due to its factorial complexity, this check can become computationally demanding for polytopes with a large number of vertices. To illustrate both the identifiability of $\mathcal{P}_{\text{ex}}$ and the convergence behavior of the algorithm for sufficiently scattered samples, we conducted the following experiment. With each run of the experiment, we generate a set of sufficiently scattered samples from $\mathcal{P}_{\text{ex}}$ using the procedure described below.

S1. Generate the polar domain set $\text{conv}(\mathbf{S})^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}}$ in **V-Form** (4): For this purpose we generate $L$ samples from $\text{conv}(\mathbf{S})^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}}$, the columns of $\mathbf{K} \in \mathbb{R}^{r \times L}$, and define $\text{conv}(\mathbf{S})^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}} = \text{conv}(\mathbf{K})$:

i. According to the condition (PMF.SS.ii) in Definition II.2, the elements of $\text{ext}(\mathcal{P}^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}})$ should be the vertices of $\text{conv}(\mathbf{S})^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}}$. Therefore, we first include elements of $\text{ext}(\mathcal{P}_{\text{ex}}^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}})$ in $\mathbf{K}$, by setting:

$$\mathbf{K}_{:,1:7} = \begin{bmatrix} 1 & 1 & -1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & -1 & 1.6 & -1.6 & 0 \\ 0 & 0 & 0 & 0 & 1.6 & 1.6 & -8/3 \end{bmatrix}.$$

Note that $\text{conv}(\mathbf{K}_{,1:7}) = \mathcal{P}_{\text{ex}}{}^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}}$. If we set $L = 7$, and therefore, skip the next step (S1.ii), our procedure would generate the vertices of $\mathcal{P}_{\text{ex}}$, which is a sufficiently scattered set.

ii. According to (PMF.SS.ii), the remaining vertices of $\mathcal{P}^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}}$ should be in the interior of $\mathcal{E}^*_{\mathcal{P}_{\text{ex}}}$. Therefore, we generate $L-7$ random points in the interior of $\mathcal{E}^*_{\mathcal{P}_{\text{ex}}}$. For this purpose, we generate $L-7$ i.i.d. $r$-dimensional random samples in $0.9\mathcal{B}_2$. Then we multiply these vectors with $\mathbf{C}_P^{-1}$ to obtain the remaining columns of $\mathbf{K}$.

iii. We find **V-Form** for $\text{conv}(\mathbf{S})^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}}$ by applying a numerical convex hull algorithm (such as ConvexHull function of Python's Scipy library [42]) to the columns of $\mathbf{K}$. The output of this step is the matrix $\mathbf{V} \in \mathbb{R}^{r \times L'}$ containing the vertices of $\mathcal{P}^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}}$.

S2. Convert the representation of $\text{conv}(\mathbf{S})^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}}$ from **V-Form** to **H-Form**: We convert the representation of $\text{conv}(\mathbf{S})^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}}$ from **V-Form** in (4) to **H-Form** in (3). For this purpose, we use the PYPOMAN (a PYthon module for POlyhedral MANipulations) software package (`duality.compute_polytope_halfspaces` function.) [43]. The output of this stage are the hyperplane parameters $(\mathbf{a}_i, b_i), i = 1, \ldots, f$, for the **H-Form**.

S3. Find the vertices of $\text{conv}(\mathbf{S}_g)$: We first perform the normalization on the hyperplane parameters to obtain $(\mathbf{a}_i/b_i, 1), i = 1, \ldots, f$. Note that according to the polar conversion described in Appendix A, obtaining **V-Form** for $\mathcal{P}_{\text{ex}}$ in (59) from the **H-Form** for $\text{conv}(\mathbf{S})^{*,\mathbf{g}_{\mathcal{P}_{\text{ex}}}}$ in (58), the vectors $\{\mathbf{a}_i/b_i + \mathbf{g}_{\mathcal{P}_{\text{ex}}}, i = 1, \ldots, f\}$ are the vertices of $\text{conv}(\mathbf{S}_g)$, where $\mathbf{g}_{\mathcal{P}_{\text{ex}}} = [0, 0, 0.375]$. Therefore, we can set

$$\mathbf{S}_{g_{:,1:f}} = \left[ \tfrac{\mathbf{a}_1}{b_1} + \mathbf{g}_{\mathcal{P}_{\text{ex}}} \; \tfrac{\mathbf{a}_2}{b_2} + \mathbf{g}_{\mathcal{P}_{\text{ex}}} \; \cdots \; \tfrac{\mathbf{a}_f}{b_f} + \mathbf{g}_{\mathcal{P}_{\text{ex}}} \right].$$

Note that $f$, the number of vertices of conv $(\mathbf{S}_g)$ (or faces of its polar), is a random quantity.

In the following experiments, we chose $L = 30$. The $\mathbf{H}_g$ matrix is generated as a $4 \times 3$ i.i.d. Gaussian matrix (with zero mean and unity variance). We added zero mean i.i.d. Gaussian noise to the input matrix $\mathbf{Y} = \mathbf{H}_g\mathbf{S}_g$. The projection operation $P_{\mathcal{P}_{\text{ex}}}$ onto $\mathcal{P}_{\text{ex}}$ is implemented through 5 alternating iterations of the following.

- the projection onto the rectangle corresponding to the range constraints, implemented by elementwise clipping operations;
- the projection onto the $\ell_1$-norm-ball for $[\, x_1 \; x_2 \,]^T$;
- the projection onto the $\ell_1$-norm-ball for $[\, x_2 \; x_3 \,]^T$.

For the algorithm hyperparameters, we selected $\tau = 10^{-8}$, $L^{(t)} = 5\|(\mathbf{H}^{(t)})^T\mathbf{H}^{(t)}\|_2$ and $\lambda = 0.01$.

The convergence of the algorithm in terms of the signal-to-interference ratio (SIR), averaged over 100 realizations, as a function of the iterations is shown in Fig. 9 for different signal-to-noise-ratio (SNR) levels. These experiments confirm the identifiability under the sufficiently scattered condition.

### B. Special Polytopes in Section III

In this section, we provide experiments for the special polytopes in Section III and illustrate an alternative sample generation method. For high dimensions, the sufficiently scattered sample generation procedure that we proposed in Section VI-A may not be feasible due to the computational complexity of the polytope representation conversion step in Step S2. Instead, we use another method which is based on principles similar to those

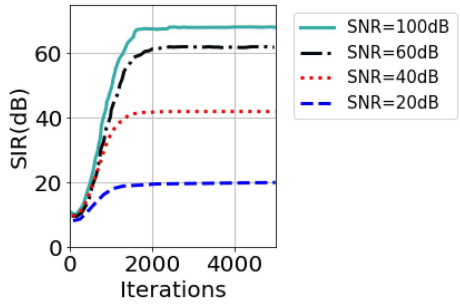
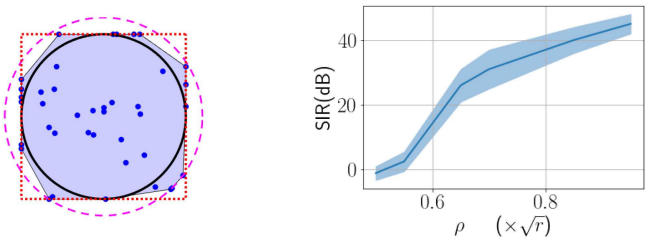



Fig. 9. Average SIR convergence curves for $\mathcal{P}_{\text{ex}}$.




(a) (b)

Fig. 10. (a) Sample generation example for Section VI-B: $\mathcal{P} = \mathcal{B}_\infty$ ($r = 2$), the dashed circle represents the boundary of the inflated MVIE for $\rho = 0.85\sqrt{r}$. (b) SIR (mean-solid line with std. envelope) as a function of $\rho$ for $\mathcal{B}_\infty$, $r = 10$ and $N = 500$.

used in [20], [32]. In this technique, we generate random samples inside the inflated version of the MVIE, and project them onto the polytope. Therefore, the procedure consists of two steps:

S1 Generate $N$ samples from the inflated version of the MVIE of the polytope: we first generate $N$ random i.i.d. Gaussian vector samples $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \frac{\rho^2}{r}\mathbf{I})$, $i = 1, \dots N$, in $\mathbb{R}^r$, the norms of which concentrate around the mean $E(\|\mathbf{w}_i\|_2) = \rho$. We call $\rho$ "inflation constant" and choose $\rho > 1$. Then, we saturate vectors with norm greater than $\rho$ by defining $\mathbf{u}_i = \rho\mathbf{w}_i/\|\mathbf{w}_i\|_2$ if $\|\mathbf{w}_i\|_2 > \rho$ and $\mathbf{u}_i = \mathbf{w}_i$ otherwise. The resulting $\mathbf{u}_i$ vectors are random samples in the hypersphere $\rho\mathcal{B}_2$, i.e., expanded unit hypersphere. Finally, we map the samples $\mathbf{u}_i$ into the inflated version of $\mathcal{E}_\mathcal{P}$ using $\mathbf{z}_i = \mathbf{C}_\mathcal{P}\mathbf{u}_i + \mathbf{g}_\mathcal{P}$. Therefore, the resulting samples lie in $\rho(\mathcal{E}_\mathcal{P} - \mathbf{g}_\mathcal{P}) + \mathbf{g}_\mathcal{P}$, the $\rho$-inflated version of the MVIE.

S2. Project the samples $\mathbf{z}_i$ onto the polytope: i.e., $\mathbf{S}_{g_{:,i}} = P_\mathcal{P}(\mathbf{z}_i), i = 1, \dots, N$.

Fig. 10(a) illustrates the proposed sample generation for $\mathcal{P} = \mathcal{B}_\infty$ and $r = 2$. The inflation constant is selected as $\rho = 0.85\sqrt{r}$. Note that the choice $\rho = 1$ corresponds to the MVIE, and $\rho = \sqrt{r}$ corresponds to the minimum volume enclosing sphere of $\mathcal{B}_\infty$. Therefore, when $\rho < \sqrt{r}$, the vertices of $\mathcal{B}_\infty$ are not covered by the inflated MVIE.

For the experiments with polytopes in Section III, we took $r = 10$ and $M = 20$. At each realization, we independently generated $\mathbf{H}_g$ as a $20 \times 10$ i.i.d. Gaussian matrix with zero mean and unity variance. We used different empirical $\lambda$ parameter choices for different polytope and sample size selections to improve the SIR performance. We conducted this experiment

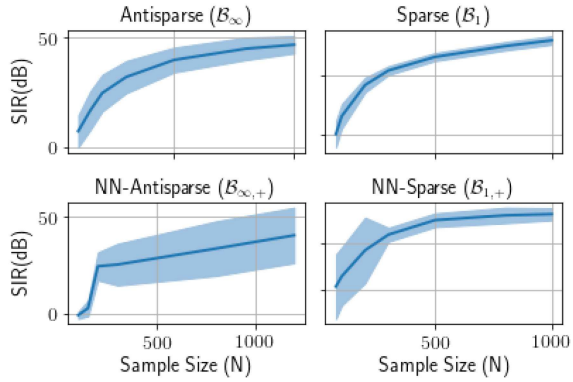



Fig. 11. SIR (mean-solid line with std. envelope) as a function of sample size $N$ for the polytopes in Section III.

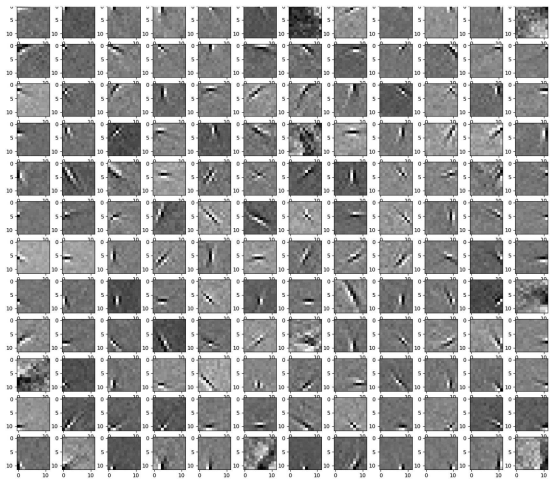


Fig. 12. Dictionary obtained by sparse PMF for prewhitened natural image patches.

for 300 realizations. Fig. 10(b) shows the SIR obtained as a function of $\rho$ for $\mathcal{P} = \mathcal{B}_\infty$ and $N = 500$. In Fig. 11, for the choice $\rho = 0.85\sqrt{r}$, we show the SIR as a function of the sample size $N$ for all four polytopes in Section III. Both Fig. 10 and 11 confirm that the sufficiently scattered set generation probability increases with the increasing values of $\rho$ and $N$ as expected.

### C. Sparse Dictionary Learning for Natural Image Patches

As the second example, we applied sparse PMF to $12 \times 12$ image patches obtained from Olshausen's prewhitened natural images (available at http://www.rctn.org/bruno/sparsenet/). These patches are vectorized (into $144 \times 1$ vectors) and placed in the columns of the $\mathbf{Y}$ matrix. The columns of the $\mathbf{H} \in \mathbb{R}^{144\times144}$ matrix obtained from the sparse PMF algorithm (with $\lambda = 1$, $\tau = 10^{-6}$ and $L^{(t)} = 4\|(\mathbf{H}^{(t)})^T\mathbf{H}^{(t)}\|_2$) are reshaped as $12 \times 12$ images (rescaled to the $0 - 1$ range) are shown in Fig. 12. It is interesting to note that, although we used a different normative approach (based on determinant maximization) than [17], we obtained similar Gabor-like edge features for the natural image patches.

## VII. Conclusion

In this article, we introduced PMF as a novel structured matrix factorization approach. Having infinite choices of identifiable polytopes positions PMF as a general framework with a diverse set of feature representation selections. We also proposed a geometric approach for identifiability analysis, which provided practically plausible conditions for the applicability of the PMF framework. We can foresee various future extensions including underdetermined data models, fully structured matrix factorization, noise/outlier analysis and efficient algorithms.

## Appendix A
## Relevant Convex Analysis Preliminaries

The polar of a set $C \subset \mathbb{R}^r$ with respect to the point $\mathbf{d} \in \mathbb{R}^r$ is defined as

$$C^{*,\mathbf{d}} = \{\mathbf{x} \in \mathbb{R}^r | \langle \mathbf{x}, \mathbf{y} - \mathbf{d} \rangle \leq 1 \; \forall \mathbf{y} \in C\}.$$

When $\mathbf{d} = \mathbf{0}$, we simplify the notation as $C^*$. Based on this notation, we can write $C^{*,\mathbf{d}} = (C - \mathbf{d})^*$. A polytope containing the origin as an interior point can be written as

$$\mathcal{P} = \{\mathbf{x} \,|\langle \mathbf{a}_i, \mathbf{x} \rangle \leq 1, i = 1, \ldots, f\}, \tag{58}$$

the polar of which is given by (Theorem 9.1 in [44])

$$\mathcal{P}^* = \text{conv}(\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_f \end{bmatrix}). \tag{59}$$

Therefore, the face normals $\mathbf{a}_i$ of the polytope $\mathcal{P}$ are the vertices of its polar $\mathcal{P}^*$. Note that if $\mathbf{d}$ is an interior point of $\mathcal{P}$, to calculate $\mathcal{P}^{*,\mathbf{d}}$, one can write $\mathcal{P} - \mathbf{d}$ in the form (58) and use $\mathcal{P}^{*,\mathbf{d}} = (\mathcal{P} - \mathbf{d})^*$ and (59).

For an ellipsoid $\mathcal{E}_P \subset \mathbb{R}^r$ defined by

$$\mathcal{E}_\mathcal{P} = \{\mathbf{C}_\mathcal{P} \mathbf{u} + \mathbf{g}_\mathcal{P} \mid \|\mathbf{u}\|_2 \leq 1, \mathbf{u} \in \mathbb{R}^r\}, \tag{60}$$

where $\mathbf{C} \in \mathbb{R}^{r \times r}$ and $\mathbf{g} \in \mathbf{R}^r$, its polar with respect to its center $\mathbf{g}_p$, can be written as

$$\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_P} = \; \{\mathbf{x} \mid \langle \mathbf{x}, \mathbf{C}_P \mathbf{u} \rangle \leq 1, \forall \|\mathbf{u}\|_2 \leq 1, \mathbf{u} \in \mathbb{R}^r\},$$

which can be simplified to

$$\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_P} = \{\mathbf{C}_P^{-1} \mathbf{u} \mid \|\mathbf{u}\|_2 \leq 1, \mathbf{u} \in \mathbb{R}^r\}. \tag{61}$$

A particular property of polar operation that we use in the article is the reversal of the set inclusion: if $A \subset B$ holds then we have $B^{*,\mathbf{d}} \subset A^{*,\mathbf{d}}$.

## Appendix B
## MVIE for Sparse Nonnegative PMF

We can obtain the MVIE parameters $(\mathbf{C}_P, \mathbf{g}_P)$ of the polytope $\mathcal{B}_{1,+}$, as the optimal solution of the following optimization problem:

$$\begin{aligned}
\underset{\mathbf{C} \in \mathbb{R}^{r \times r}, \mathbf{g} \in \mathbb{R}^r}{\text{minimize}} \quad & -\log\det \mathbf{C} && \text{(62a)}\\
\text{subject to} \quad & \|\mathbf{C}\mathbf{e}_i\|_2 - \mathbf{e}_i^T \mathbf{g} \leq 0, \quad i = 1, \ldots, r && \text{(62b)}\\
& \|\mathbf{C}\mathbf{1}\|_2 + \mathbf{1}^T \mathbf{g} \leq 1 && \text{(62c)}\\
& \mathbf{C} \succeq \mathbf{0}, && \text{(62d)}
\end{aligned}$$

which is obtained by applying the description of $\mathcal{B}_{1,+}$ in (29) to the generic MVIE optimization problem in (9b). To find the solution of (62), we utilize the following Karush-Kuhn-Tucker (KKT) optimality conditions [33] for the solution pair $(\mathbf{C}_P, \mathbf{g}_P)$:

$$-\mathbf{C}_P^{-1} + \sum_{i=1}^{r} \lambda_i \frac{\mathbf{C}_P \mathbf{e}_i \mathbf{e}_i^T}{\|\mathbf{C}_P \mathbf{e}_i\|_2} + \lambda_{r+1} \frac{\mathbf{C}_P \mathbf{1}\mathbf{1}^T}{\|\mathbf{C}_P \mathbf{1}\|_2} - \mathbf{W} = \mathbf{0}, \tag{63}$$

$$\lambda_i(\|\mathbf{C}_P \mathbf{e}_i\|_2 - \mathbf{e}_i^T \mathbf{g}_P) = 0, \text{ for all } i = 1, \ldots, r, \tag{64}$$

$$\lambda_{r+1}(\|\mathbf{C}_P \mathbf{1}\|_2 + \mathbf{1}^T \mathbf{g}_P - 1) = 0, \tag{65}$$

$$-\sum_{i=1}^{r} \lambda_i \mathbf{e}_i + \lambda_{r+1} \mathbf{1} = \mathbf{0}, \tag{66}$$

$$Tr(\mathbf{W}\mathbf{C}_P) = 0, \tag{67}$$

where (63) and (66) represent stationarity conditions of $\mathbf{C}_P$ and $\mathbf{g}_P$, respectively. The remaining KKT equations – (64), (65) and (67) – are known as complementary slackness conditions, which involve optimal dual variables $\lambda_1, \lambda_2, \ldots, \lambda_{r+1} \in \mathbb{R}_+$ and $\mathbf{W} \succeq \mathbf{0}$ corresponding to the inequality constraints in (62b), (62c) and (62d), respectively.

Based on the optimization in (62) and the corresponding optimality conditions in (63)-(67), we can deduce the following:

- $\mathbf{C}_P$ is nonsingular, i.e., $\mathbf{C}_P \succ \mathbf{0}$, otherwise the objective function $-\log\det \mathbf{C}_P$ becomes infinite under the primal feasibility condition $\mathbf{C} \succeq \mathbf{0}$ in (62d). Therefore, the nonsingularity of $\mathbf{C}$ implies $\mathbf{W} = \mathbf{0}$ due the KKT condition given in (67).
- The condition in (66) is equivalent to $\lambda_1 = \lambda_2 = \ldots = \lambda_{r+1} = \lambda$, further implying that either all or none of the corresponding primal feasibility conditions are binding, i.e., the inequalities corresponding to these dual variables, (62b) and (62c), are equalities at the optimal point. If we assume none of them are binding, the KKT conditions (65) and (66) imply that $\lambda_1 = \lambda_2 = \ldots = \lambda_{r+1} = 0$, which further implies $\mathbf{W} \neq \mathbf{0}$ due to (63). This outcome contradicts our earlier finding that $\mathbf{W} = \mathbf{0}$. Therefore, we conclude that

  $$\lambda_1 = \lambda_2 = \ldots = \lambda_{r+1} > 0, \tag{68}$$

  and all corresponding inequalities are binding, i.e., $\|\mathbf{C}_P \mathbf{e}_i\|_2 - \mathbf{e}_i^T \mathbf{g}_P = 0$ for $i = 1, \ldots, r$ and $\|\mathbf{C}_P \mathbf{1}\|_2 + \mathbf{1}^T \mathbf{g}_P = 1$ hold.
- Using (68) and $\mathbf{W} = \mathbf{0}$, the expression in (63) can be rewritten as

  $$-\mathbf{C}_P^{-1} + \lambda \left( \sum_{i=1}^{r} \frac{\mathbf{C}_P \mathbf{e}_i \mathbf{e}_i^T}{\|\mathbf{C}_P \mathbf{e}_i\|_2} + \frac{\mathbf{C}_P \mathbf{1}\mathbf{1}^T}{\|\mathbf{C}_P \mathbf{1}\|_2} \right) = \mathbf{0}. \tag{69}$$

  If we multiply both sides of (69) by $\mathbf{C}_P^{-1}$, and rearrange the terms, we obtain

  $$\mathbf{C}_P^{-2} = \lambda \left( \sum_{i=1}^{r} \frac{\mathbf{e}_i \mathbf{e}_i^T}{\|\mathbf{C}_P \mathbf{e}_i\|_2} + \frac{\mathbf{1}\mathbf{1}^T}{\|\mathbf{C}_P \mathbf{1}\|_2} \right). \tag{70}$$

Using the binding constraints $\|\mathbf{C}_P\mathbf{e}_i\|_2 = \mathbf{e}_i^T\mathbf{g}_P$ and $\|\mathbf{C}_P\mathbf{1}\|_2 = 1 - \mathbf{1}^T\mathbf{g}_P$, (70) can be rewritten as

$$\mathbf{C}_P^{-2} = \lambda\left(\mathbf{G}^{-1} + \frac{\mathbf{1}\mathbf{1}^T}{1-\mathbf{1}^T\mathbf{g}_P}\right), \tag{71}$$

where $\mathbf{G} \in \mathbb{R}^{r\times r}$ is a diagonal matrix, the $i^{\text{th}}$ diagonal entry of which is equal to $\mathbf{e}_i^T\mathbf{g}$ for all i=1, ..., r. Applying the matrix inversion lemma to the right hand side of (71), we have

$$\mathbf{C}_P^2 = \lambda^{-1}\left(\mathbf{G} - \mathbf{g}_P\mathbf{g}_P^T\right). \tag{72}$$

Multiplying (72) by the ones-vector from both the left and right yields

$$\mathbf{1}^T\mathbf{C}_P^2\mathbf{1} = \lambda^{-1}\mathbf{1}^T\mathbf{g}_P\left(1 - \mathbf{1}^T\mathbf{g}_P\right). \tag{73}$$

The binding constraint $\|\mathbf{C}_P\mathbf{1}\|_2 = 1 - \mathbf{1}^T\mathbf{g}_P$ can be used to obtain an alternative expression

$$\mathbf{1}^T\mathbf{C}_P^2\mathbf{1} = (1 - \mathbf{1}^T\mathbf{g}_P)^2. \tag{74}$$

Equating the right hand sides of (73) and (74), we have

$$\lambda = \frac{\mathbf{1}^T\mathbf{g}_P}{1-\mathbf{1}^T\mathbf{g}_P} = \frac{1}{1-\mathbf{1}^T\mathbf{g}_P} - 1. \tag{75}$$

- Squaring both sides of the binding constraint $\|\mathbf{C}_P\mathbf{e}_i\|_2 = \mathbf{e}_i^T\mathbf{g}$, we obtain $\mathbf{e}_i^T\mathbf{C}_P^2\mathbf{e}_i = (\mathbf{e}_i^T\mathbf{g}_P)^2 = g_{P,i}^2$, where $g_{P,i}$ stands for $\mathbf{e}_i^T\mathbf{g}_P$. Inserting the expression for $\mathbf{C}_P^2$ in (72), we obtain

$$\lambda^{-1}(g_{P,i} - g_{P,i}^2) = g_{P,i}^2,$$

which leads to

$$\lambda = \frac{1-g_{P,i}}{g_{P,i}} = \frac{1}{g_{P,i}} - 1, \quad \forall i \in \{1,\dots,r\}. \tag{76}$$

Combining (75) and (76), we can write

$$1 - \mathbf{1}^T\mathbf{g} = g_{P,i}, \quad \forall i \in \{1,\dots,r\}.$$

From this expression and (76), we obtain $\mathbf{g}_P = \frac{1}{r+1}\mathbf{1}$ and $\lambda = r$. Inserting these values into (72) yields

$$\mathbf{C}_P^2 = \frac{1}{r}\left(\frac{1}{r+1}\mathbf{I} - \frac{1}{(r+1)^2}\mathbf{1}\mathbf{1}^T\right), \tag{77}$$

the square root of which is given by

$$\mathbf{C}_P = \frac{1}{\sqrt{r}}\left(\frac{1}{\sqrt{r+1}}\mathbf{I} - \frac{\sqrt{r+1}-1}{r(r+1)}\mathbf{1}\mathbf{1}^T\right).$$

## APPENDIX C
## PROOF OF THEOREM 6

The following lemma from [21] is used in the proof of Theorem 6.

*Lemma 1: Polar of the Transformed Set:* We can characterize the polar of the transformed set $\mathbf{A}(S)$ in the following way,

$$(\mathbf{A}(S))^{*,\mathbf{d}} = \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{x}, \mathbf{A}\mathbf{y} - \mathbf{d}\rangle \le 1 \ \forall \mathbf{y} \in S\}$$

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

*Proof of Theorem 6:* Using the same arguments in the proof of Theorem 2, we write the optimization problem equivalent to (7) as

$$\underset{\mathbf{A}\in\mathbb{R}^{r\times r}}{\text{maximize}} \ |\det(\mathbf{A})| \tag{78a}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{S}_{g\,,j} \in \mathcal{P}, \quad j = 1,\dots,N. \tag{78b}$$

We first show the "if" part: Let $\mathbf{S}_g$ be any sufficiently scattered factor for $\mathcal{P}$ and $\mathbf{H}_g$ be any full column rank matrix. Let $\mathbf{A}_*$ represent any global optimum of the equivalent Det-Max optimization problem in (78). Below we show that $\mathbf{A}_*(\text{ext}(\mathcal{P})) = \text{ext}(\mathcal{P})$. We first note that due to the constraint (78b), we have

$$\text{conv}(\mathbf{A}_*\mathbf{S}_g) \subseteq \mathcal{P}. \tag{79}$$

Furthermore, since $\mathbf{A} = \mathbf{I}$ is a trivial feasible solution of (78), $\mathbf{A}_*$ ought to satisfy $|\det(\mathbf{A}_*)| \ge \det(\mathbf{I}) = 1$. The inclusion of the MVIE of $\mathcal{P}$ in $\text{conv}(\mathbf{S}_g)$, due to the sufficient scattering condition (PMF.SS.i), and (79) lead to $\mathbf{A}_*(\mathcal{E}_\mathcal{P}) \subset \mathcal{P}$. We note that $\mathbf{A}_*(\mathcal{E}_\mathcal{P})$ is an ellipsoid in $\mathcal{P}$, for which

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

which uses the lower bound $|\det(\mathbf{A}_*)| \ge 1$. Conversely, based on the uniqueness of the MVIE for $\mathcal{P}$, we can write $\text{vol}(\mathbf{A}_*(\mathcal{E}_\mathcal{P})) \le \text{vol}(\mathcal{E}_\mathcal{P})$. As a result, we conclude that $|\det(\mathbf{A}_*)| = 1$ and

$$\mathbf{A}_*(\mathcal{E}_\mathcal{P}) = \mathcal{E}_\mathcal{P}. \tag{80}$$

In other words, the constraint in (78b) and the sufficient scattering condition (PMF.SS.i) together imply $|\det(\mathbf{A}_*)| = 1$ and restrict $\mathbf{A}_*$ to map $\mathcal{E}_\mathcal{P}$ onto itself. Thus, we can write $\mathbf{A}_*\mathbf{g}_P = \mathbf{g}_P$ for the center of $\mathcal{E}_\mathcal{P}$. As a consequence, according to Lemma 1, (80) is identical to

$$\mathbf{A}_*^{-T}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) = \mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}, \tag{81}$$

where $\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}$ is also an ellipsoid. Therefore, (81) implies

$$\mathbf{A}_*^{-T}(\text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})) = \text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}). \tag{82}$$

Using the reversal of the set inclusion property of the polar operation, given in Appendix A, and (79), we can write $\mathcal{P}^{*,\mathbf{g}_P} \subseteq \text{conv}(\mathbf{A}_*\mathbf{S}_g)^{*,\mathbf{g}_P}$. Applying Lemma 1 to the right hand side of this expression, we obtain $\mathcal{P}^{*,\mathbf{g}_\mathcal{P}} \subseteq \mathbf{A}_*^{-T}(\text{conv}(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}})$, which implies

$$\mathbf{A}_*^{-T}(\text{conv}(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}}) \supseteq \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}). \tag{83}$$

Based on the sufficient scattering condition (PMF.SS.ii), we can write the inclusion expressions $\text{conv}(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}} \supseteq \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})$, $\text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) \supseteq \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})$, and

$$\mathbf{A}_*^{-T}(\text{conv}(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}}) \cap \text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) = \mathbf{A}_*^{-T}(\text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})), \tag{84}$$

where we also inserted (82). Furthermore, based on (83), we can write

$$\mathbf{A}^{-T}(\text{conv}(\mathbf{S}_g)^{*,\mathbf{g}_\mathcal{P}}) \cap \text{bd}(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) \supseteq \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}). \tag{85}$$

The expressions in (84) and (85) together imply

$$\mathbf{A}_*^{-T}(\text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})) \supseteq \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}). \tag{86}$$

Since the cardinality of the sets on both sides are equal, we obtain

$$\mathbf{A}_*^{-T}(\text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})) = \text{ext}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}), \tag{87}$$

which implies $\mathbf{A}_*^{-T}(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) = \mathcal{P}^{*,\mathbf{g}_\mathcal{P}}$. Invoking Lemma 1, we obtain $\mathbf{A}_*(\mathcal{P}) = \mathcal{P}$ and deduce that $\mathbf{A}_*(\text{ext}(\mathcal{P})) = \text{ext}(\mathcal{P})$ due to the convexity of $\mathcal{P}$. As a result, the condition that $\text{ext}(\mathcal{P})$ is a *permutation-and/or-sign-only invariant set* implies $\mathbf{A}_* = \mathbf{D}\mathbf{\Pi}$, with a diagonal sign matrix $\mathbf{D}$ and a permutation matrix $\mathbf{\Pi}$, which further implies the identifiability of the generative model.

The "only if" part: Let $\mathbf{V}_\mathcal{P} \in \mathbb{R}^{r\times K}$ be a matrix that contains all $K$ vertices of $\mathcal{P}$ in its columns. Clearly, the choice $\mathbf{S}_g = \mathbf{V}_\mathcal{P}$ is a sufficiently scattered factor for $\mathcal{P}$. Suppose that $\text{ext}(\mathcal{P})$ is not a permutation-and/or-sign-only invariant set, then there exists an $\mathbf{A} \in \mathbb{R}^{r\times r}$, which is not a product of a diagonal and a permutation matrix, for which $\mathbf{A}\mathbf{V}_\mathcal{P} = \mathbf{V}_\mathcal{P}\mathbf{\Pi}$ for some permutation matrix $\mathbf{\Pi} \neq \mathbf{I}$. Since $|\det(\mathbf{A})| = 1$, $\mathbf{V}_\mathcal{P}\mathbf{\Pi}$ would be another solution for the Det-Max optimization problem in (7), violating the identifiability of all sufficiently scattered sets.

## REFERENCES


[1] 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*. Hoboken, NJ, USA: Wiley, 2009.

[2] X. Fu, K. Huang, N. D. Sidiropoulos, and W.-K. Ma, "Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 36, no. 2, pp. 59–80, Mar. 2019.

[3] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (Cat. No 03TH8684)*, 2003, pp. 177–180.

[4] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2003, pp. 267–273.

[5] C.-H. Lin, R. Wu, W.-K. Ma, C.-Y. Chi, and Y. Wang, "Maximum volume inscribed ellipsoid: A new simplex-structured matrix factorization framework via facet enumeration and convex optimization," *SIAM J. Imag. Sci.*, vol. 11, no. 2, pp. 1651–1679, 2018.

[6] A. T. Erdogan, "A class of bounded component analysis algorithms for the separation of both independent and dependent sources," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5730–5743, Aug. 2013.

[7] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 992–996, Jul. 2005.

[8] 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.

[9] 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, Oct. 1999.

[10] S. Cruces, "Bounded component analysis of linear mixtures: A criterion of minimum convex perimeter," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2141–2154, Jan. 2010.

[11] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Berlin, Germany: Springer Science & Business Media, 2010.

[12] E. Babatas and A. T. Erdogan, "An algorithmic framework for sparse bounded component analysis," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5194–5205, Aug. 2018.

[13] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Aug. 2011.

[14] K. Huang and X. Fu, "Detecting overlapping and correlated communities without pure nodes: Identifiability and algorithm," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2859–2868.

[15] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 550–563, Sep. 2009.

[16] V. Leplat, N. Gillis, and A. M. Ang, "Blind audio source separation with minimum-volume beta-divergence NMF," *IEEE Trans. Signal Process.*, vol. 68, no. 5, pp. 3400–3410, May 2020.

[17] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?," *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, Dec. 1997.

[18] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 801–808.

[19] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?," *Adv. Neural Inf. Process. Syst.*, vol. 16, pp. 1141–1148, Dec. 2003.

[20] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5530–5546, May 2015.

[21] G. Tatli and A. T. Erdogan, "Generalized polytopic matrix factorization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 3235–3239.

[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, Jul. 2011.

[23] 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, pp. 1–9, Mar. 2008.

[24] S. Arora, R. Ge, R. Kannan, and A. Moitra, "Computing a nonnegative matrix factorization-provably," *SIAM J. Comput.*, vol. 45, no. 4, pp. 1582–1611, 2016.

[25] 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, no. 1, pp. 211–224, Oct. 2013.

[26] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos, "Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2306–2320, Feb. 2015.

[27] X. Fu, K. Huang, and N. D. Sidiropoulos, "On identifiability of nonnegative matrix factorization," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 328–332, Jan. 2018.

[28] D. Bremner, M. D. Sikiric, and A. Schürmann, "Polyhedral representation conversion up to symmetries," in *CRM Proc.*, vol. 48, 2009, pp. 45–72.

[29] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, May 2011.

[30] H. A. Inan and A. T. Erdogan, "A convolutive bounded component analysis framework for potentially nonstationary independent and/or dependent sources," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 18–30, Nov. 2014.

[31] A. Sengupta, "Generalized variance," in *Encyclopedia of Statistical Sciences*, S. Kotz, C. B. Read, N. Balakrishnan, and B. Vidakovic, Eds. Wiley Online Library, 2004.

[32] 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, no. 23, pp. 6254–6268, Aug. 2016.

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

[34] K. Ball, "Ellipsoids of maximal volume in convex bodies," *Geometriae Dedicata*, vol. 41, no. 2, pp. 241–250, Feb. 1992.

[35] J.-J. Fuchs, "Spread representations," in *Proc. Conf. Rec. 45th Asilomar Conf. Signals, Syst. Comput.*, 2011, pp. 814–817.

[36] H. Jégou, T. Furon, and J.-J. Fuchs, "Anti-sparse coding for approximate nearest neighbor search," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 2029–2032.

[37] C. Studer, T. Goldstein, W. Yin, and R. G. Baraniuk, "Democratic representations," 2014, *arXiv:1401.3420*.

[38] D. L. Donoho, "For most large underdetermined systems of equations, the minimal $\ell_1$-norm near-solution approximates the sparsest near-solution," *Commun. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, Jul. 2006.

[39] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[40] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. New York, NY, USA: Academic, 2010.

[41] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the $\ell_1$-ball for learning in high dimensions," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 272–279.

[42] P. Virtanen *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in python," *Nature Methods*, vol. 17, pp. 261–272, Mar. 2020.
[43] S. Caron, "Python module for polyhedral manipulations - pypoman. version 1.0," 2020. [Online]. Available: https://scaron.info/doc/pypoman/
[44] A. Brondsted, *An Introduction to Convex Polytopes*. Berlin, Germany: Springer Science & Business Media, 2012.