

# A review of NMF, PLSA, LBA, EMA, and LCA with a focus on the identifiability issue

Qianqian Qi

Hangzhou Dianzi University, China

[q.qi@hdu.edu.cn](mailto:q.qi@hdu.edu.cn)

Peter G. M. van der Heijden

Utrecht University, the Netherlands and University of Southampton, UK

[p.g.m.vanderheijden@uu.nl](mailto:p.g.m.vanderheijden@uu.nl)

## Abstract

Across fields such as machine learning, social science, geography, considerable attention has been given to models that factorize a nonnegative matrix into the product of two or three matrices, subject to nonnegative or row-sum-to-1 constraints. Although these models are to a large extend similar or even equivalent, they are presented under different names, and their similarity is not well known. This paper highlights similarities among five popular models, latent budget analysis (LBA), latent class analysis (LCA), end-member analysis (EMA), probabilistic latent semantic analysis (PLSA), and nonnegative matrix factorization (NMF). We focus on an essential issue-identifiability of these models and prove that the solution of LBA, EMA, LCA, PLSA is unique if and only if the solution of NMF is unique. We also provide a brief review for algorithms of these models. We illustrate the models with a time budget dataset from social science, and end the paper with a discussion of closely related models such as archetypal analysis.

**Keywords:** Archetypal analysis; Blind hyperspectral unmixing; Blind sound separation; Uniqueness; Separability; Inner extreme solution; Minimum end member; Minimum volume

## 1 Introduction

In many settings in machine learning, geology, social science, observed data are nonnegative and can be considered a mixture of multiple latent or unobserved sources (Hayashi, Aksoy, Ballard, & Park, 2025; Jones et al., 2024; Lin et al., 2025). For these data we are

interested in the latent sources and the mixing weights. Approximating the observed non-negative matrix by a non-negative factorization of lower rank is a natural mathematical framework to extract these latent structures from the observed data.

In many applications it makes sense to transform the non-negative data into compositional data, where the elements in each row in a matrix add up to one. For the approximation of such data by latent compositions, a number of equivalent models exist in statistics and machine learning. The equivalence of these models is rarely discussed and it is one of the aims of this review to give attention to this equivalence.

- Latent budget analysis (LBA) is a statistical model particularly known in the social sciences designed for the analysis of compositional data (Clogg, 1981; De Leeuw & Van der Heijden, 1988; De Leeuw, Van der Heijden, & Verboon, 1990; Van der Heijden, 1994). The name "latent budgets" stems from the original proposal to approximate so-called observed time budgets of (groups of) individuals by a number of latent, i.e. unknown time budgets. The term budget reflects that the elements of a budget adds up to 1: comparable to a (traditional) financial budget, one can spend an amount of time only once.
- With a similar aim, in geology, the technique end-member analysis (EMA), having the same parametric representation as LBA, was proposed for deriving provenance, transport, and sedimentation processes from grain-size distribution datasets (Rennier, 1989, 1995; Weltje, 1997; Weltje & Prins, 2007). End members refer to sources, whose compositions may be identified with particular source materials. For example, an end member is one extreme in sedimentary rocks or fossils.
- In machine learning, the technique probabilistic latent semantic analysis (PLSA), also called probabilistic latent semantic indexing (PLSI), whose asymmetric form also has the same parametric representation as LBA, was proposed for automated document indexing (Hofmann, 1999, 2001). PLSA is inspired by latent semantic analysis (LSA), another document indexing technique (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Dumais, 1991; Dumais, Furnas, Landauer, Deerwester, & Harshman, 1988). Compared to LSA, PLSA defines a proper generative model for the data in terms of (conditional) probabilities.

- LBA, EMA and asymmetric PLSA are models for compositional data, where the row elements add up to 1. There are two mathematically equivalent models for non-negative matrices where all elements of the matrix add up to 1. These equivalences are well known: LBA is equivalent to latent class analysis (LCA) for two-way tables (Clogg, 1981, 1995; Evans, Gilula, & Guttman, 1989; Gilula, 1979, 1983; Goodman, 1974a, 1974b, 1987; Haberman, 1979; Lazarsfeld & Henry, 1968; Van der Heijden, Mooijaart, & De Leeuw, 1989) and asymmetric PLSA is equivalent to symmetric PLSA (Hofmann, 1999, 2001).

LBA, EMA, and asymmetric PLSA are special cases of non-negative matrix factorization (NMF), which is often used to learn parts of faces and semantic features of text (Lee & Seung, 1999; Paatero & Tapper, 1994). Lee and Seung (1999)'s paper published in *Nature* sparked much research in NMF. Now NMF is well recognized as a workhorse for data science (Fu, Huang, Sidiropoulos, & Ma, 2019; Gillis, 2020; Guo, Li, & Liang, 2024; Hobolth, Guo, Kousholt, & Jensen, 2020; Saberi-Movahed et al., 2025). Despite research on NMF, PLSA, EMA, and LBA, only few attempts have been made to establish the connections among them (De Leeuw & Van der Heijden, 1991; Ding, Li, & Peng, 2008; Gaussier & Goutte, 2005; Van der Heijden, 1994). As far as we are aware, our work appears to be the first linking LBA from social science and EMA from geology to PLSA and NMF from computing science.

One essential problem of LBA, EMA, PLSA, LCA for two-way tables, and NMF is model identifiability - under what conditions latent factors are unique (up to permutation and scaling) (Gillis, 2020; Goodman, 1987; Van der Ark, Van der Heijden, & Sikkel, 1999; Weltje, 1997). The identifiability issue is tantamount to the question of whether or not these latent factors are the only interpretation of the data, or alternatives exist. Theoretical analysis for the LBA, EMA, LCA for two-way tables, and PLSA solutions to be unique remains relatively limited. On the other hand, in NMF the identifiability issue has been thoroughly studied. Another aim of this paper is to provide a better understanding of the identifiability in LBA, PLSA, LCA for two-way tables, and EMA by leveraging the theoretical results established for NMF. While NMF can benefit from insights in LBA, EMA, LCA for two-way tables, and PLSA in turn, this paper focuses primarily on applying NMF theory to LBA, EMA, LCA for two-way tables, and PLSA.

The paper is organized as follows. Section 2 shows the relationship among LBA, EMA, asymmetric PLSA, and NMF, and the equivalent models LCA for two-way tables and symmetric PLSA. Section 3 provides motivating examples for the identifiability issue. Section 4 proves that the solution of NMF is unique if and only if the solution of LBA, EMA, and PLSA is unique. We review uniqueness theorems for LBA, EMA, PLSA, and NMF with two main goals: to synthesize existing theorems and to establish new sufficient conditions for the uniqueness of LBA, EMA, and PLSA solutions using NMF. Section 5 briefly introduces algorithms for the estimation of models. Section 6 applies LBA, EMA, and NMF to study a time budget dataset from social science. Section 7 introduces other related models, such as archetypal analysis. Finally, Section 8 concludes this paper.

## 2 NMF, LBA, EMA, PLSA, and LCA

Let  $\mathbf{X}$  be an observed non-negative matrix of size  $I \times J$  with element  $x_{ij}$  ( $i = 1, \dots, I; j = 1, \dots, J$ ). In general, given a matrix  $\mathbf{C}$ , let  $\mathbf{D}_C$  be a diagonal matrix whose diagonal entries are the row sums of  $\mathbf{C}$ . For example, let  $\mathbf{P} = \mathbf{D}_{\mathbf{X}}^{-1} \mathbf{X}$  with element  $p_{ij} = x_{ij} / \sum_j x_{ij}$ , where  $\mathbf{D}_{\mathbf{X}}$  has elements  $\sum_j x_{ij}$ .  $\mathbf{P}$  is also known as a matrix with compositional data, where the observed conditional row elements  $p_{ij}$  for each row  $i$  add up to 1.

### 2.1 NMF

NMF approximates  $\mathbf{X}$  by a lower rank matrix which is the product of two nonnegative matrices  $\mathbf{M}$  and  $\mathbf{H}$ . The theoretical matrix for this lower rank matrix is  $\Phi$ . Thus, NMF is

$$\Phi = \mathbf{M}\mathbf{H} \tag{1}$$

where  $\mathbf{M}$  and  $\mathbf{H}$  have sizes  $I \times K$  and  $K \times J$  respectively and both have rank  $K$  (Lee & Seung, 1999). In (1),  $K$  is the so-called non-negative rank of  $\Phi$  (Ang, Hamed, & De Sterck, 2025; Gillis, 2020; Huang, Sidiropoulos, & Swami, 2014). The term “non-negative rank” is used, instead of the term “rank”, as  $\Phi$ ,  $\mathbf{M}$ , and  $\mathbf{H}$  are all non-negative. The non-negative rank  $K \leq \min(I, J)$ , and if  $K < \min(I, J)$ , then  $\Phi$  is a matrix of lower rank. Due to the non-negativity constraints of  $\mathbf{M}$  and  $\mathbf{H}$ , the non-negative rank of  $\Phi$  may be larger than the rank of  $\Phi$ , but if  $\Phi$  has rank 2, then the non-negative rank is also 2 (De Leeuw & Van der

Heijden, 1991; Gillis, 2020). For convenience, we call  $\mathbf{M}$  the coefficient matrix and  $\mathbf{H}$  the basis matrix.

The identifiability issue for NMF is, shortly, that for some  $K \times K$  transformation matrix  $S$ ,

$$\Phi = \mathbf{M}\mathbf{H} = (\mathbf{M}\mathbf{S}^{-1})(\mathbf{S}\mathbf{H}) = \tilde{\mathbf{M}}\tilde{\mathbf{H}}. \quad (2)$$

The choice of  $S$  is restricted by the requirement that  $\tilde{\mathbf{M}}$  and  $\tilde{\mathbf{H}}$  should be non-negative as well, just like  $\mathbf{M}$  and  $\mathbf{H}$ . The identifiability issue will be discussed in more detail in sections below.

## 2.2 LBA, EMA and asymmetric PLSA

In LBA, EMA, and asymmetric LSA, the observed matrix  $\mathbf{P} = \mathbf{D}_X^{-1}\mathbf{X}$  with row-sum-to-1 constraints is approximated by a lower rank matrix which is the product of two nonnegative matrices with row-sum-to-1 constraints. The theoretical matrix for this lower rank matrix is  $\Pi$ . LBA, EMA, and asymmetric PLSA have the same representation (De Leeuw et al., 1990; Hofmann, 2001; Weltje, 1997), namely

$$\begin{aligned} \Pi &= \mathbf{W}\mathbf{G} \\ \text{subject to } \Pi\mathbf{1} &= \mathbf{1}, \mathbf{W}\mathbf{1} = \mathbf{1}, \mathbf{G}\mathbf{1} = \mathbf{1} \end{aligned} \quad (3)$$

where  $\mathbf{W}$  and  $\mathbf{G}$  are nonnegative matrices of sizes  $I \times K$  and  $K \times J$  respectively, both have rank  $K$ , and  $\mathbf{1}$  is the vector of all ones of appropriate dimension. The model decomposes the  $I$  theoretical compositions in  $\Pi$  into  $K$  latent compositions in  $\mathbf{G}$ , where  $\mathbf{W}$  provides the mixing parameters that yield the  $I$  theoretical compositions from the  $K$  latent compositions.

The identifiability issue for LBA, EMA and asymmetric PLSA is, shortly, that for some  $K \times K$  transformation matrix  $T$ ,

$$\Pi = \mathbf{W}\mathbf{G} = (\mathbf{W}\mathbf{T}^{-1})(\mathbf{T}\mathbf{G}) = \tilde{\mathbf{W}}\tilde{\mathbf{G}}. \quad (4)$$

The choice of  $T$  is restricted by the requirement that  $\tilde{\mathbf{W}}$  and  $\tilde{\mathbf{G}}$  should be non-negative and  $\tilde{\mathbf{W}}\mathbf{1} = \mathbf{1}, \tilde{\mathbf{G}}\mathbf{1} = \mathbf{1}$ , just like  $\mathbf{W}$  and  $\mathbf{G}$ . The sum-to-1 constraints  $\tilde{\mathbf{W}}\mathbf{1} = \mathbf{1}, \tilde{\mathbf{G}}\mathbf{1} = \mathbf{1}$  are implemented by restricting the elements of  $T$  by  $T\mathbf{1} = \mathbf{1}$  (De Leeuw et al., 1990). Thus

weighted averages of the  $K$  compositions in the  $K \times J$  matrix  $\mathbf{G}$ , where the weights are in  $\mathbf{T}$ , lead to the new latent compositions in the matrix  $\tilde{\mathbf{G}}$ , corresponding to the new weights for  $\boldsymbol{\Pi}$  being in  $\tilde{\mathbf{W}}$ . The identifiability issue will be discussed in more details in sections below.

### 2.3 LCA of a two-way table and symmetric PLSA

LCA of a two-way table and symmetric PLSA are equivalent tools that exist under different names, where LCA is more often used in statistics and symmetric PLSA in computer sciences (Fienberg, Hersh, Rinaldo, & Zhou, 2009; Goodman, 1987; Hofmann, 1999, 2001). Let  $\mathbf{Y}$  be an observed matrix with elements  $x_{ij}/\sum_{ij} x_{ij}$  such that the sum of all elements in  $\mathbf{Y}$  is 1. LCA of a two-way table and symmetric PLSA approximate  $\mathbf{Y}$  by a lower rank matrix which is the product of three non-negative matrices with sum-to-one constraints. If  $\boldsymbol{\Psi}$  is the theoretical matrix for this lower rank matrix, its elements can be interpreted as probabilities. Mathematically,

$$\begin{aligned}\boldsymbol{\Psi} &= \mathbf{A}\boldsymbol{\Theta}\mathbf{B} \\ \text{subject to } \mathbf{1}^T \boldsymbol{\Psi} \mathbf{1} &= 1, \mathbf{A}^T \mathbf{1} = \mathbf{1}, \mathbf{1}^T \boldsymbol{\Theta} \mathbf{1} = 1, \mathbf{B} \mathbf{1} = \mathbf{1}.\end{aligned}\tag{5}$$

where  $\boldsymbol{\Theta}$  is a diagonal matrix, and  $\mathbf{A}$ ,  $\boldsymbol{\Theta}$ , and  $\mathbf{B}$  are nonnegative matrices of sizes  $I \times K$ ,  $K \times K$ , and  $K \times J$ , respectively. Due to the sum-to-1 constraints the elements of these matrices can be interpreted as (conditional) probabilities.

### 2.4 Relations

We first discuss the equivalence of LBA, EMA and asymmetric PLSA with LCA of a two-way table and symmetric PLSA.

**Theorem 1.** *LCA of a two-way table and symmetric PLSA are equivalent to LBA, EMA, and asymmetric PLSA.*

*Proof.* From (3) we have  $\boldsymbol{\Pi} = \mathbf{W}\mathbf{G}$ , with  $\boldsymbol{\Pi}\mathbf{1} = \mathbf{1}$ ,  $\mathbf{W}\mathbf{1} = \mathbf{1}$ ,  $\mathbf{G}\mathbf{1} = \mathbf{1}$ . From (5) we have  $\boldsymbol{\Psi} = \mathbf{A}\boldsymbol{\Theta}\mathbf{B}$ , with  $\mathbf{1}^T \boldsymbol{\Psi} \mathbf{1} = 1$ ,  $\mathbf{A}^T \mathbf{1} = \mathbf{1}$ ,  $\mathbf{1}^T \boldsymbol{\Theta} \mathbf{1} = 1$ ,  $\mathbf{B} \mathbf{1} = \mathbf{1}$ . There is  $\mathbf{D}_{\boldsymbol{\Psi}}^{-1} \boldsymbol{\Psi} = \boldsymbol{\Pi}$  as  $\mathbf{D}_{\boldsymbol{\Psi}}^{-1} \boldsymbol{\Psi} \mathbf{1} = \mathbf{1}$ . Then  $\boldsymbol{\Pi} = \mathbf{D}_{\boldsymbol{\Psi}}^{-1} \mathbf{A}\boldsymbol{\Theta}\mathbf{B}$  where  $\mathbf{D}_{\boldsymbol{\Psi}}^{-1} \mathbf{A}\boldsymbol{\Theta} = \mathbf{W}$  and  $\mathbf{B} = \mathbf{G}$ . We obtain LBA, EMA, asymmetric PLSA of  $\boldsymbol{\Pi}$  as  $\mathbf{D}_{\boldsymbol{\Psi}}^{-1} \mathbf{A}\boldsymbol{\Theta} \mathbf{1} = \mathbf{D}_{\boldsymbol{\Psi}}^{-1} \mathbf{A}\boldsymbol{\Theta} \mathbf{B} \mathbf{1} = \mathbf{D}_{\boldsymbol{\Psi}}^{-1} \boldsymbol{\Psi} \mathbf{1} = \mathbf{1}$  and  $\mathbf{B} \mathbf{1} = \mathbf{1}$ .

On the other hand, suppose that we have LBA, EMA, asymmetric PLSA of  $\Pi$ :  $\Pi = WG$ . We have  $D_\Psi \Pi = \Psi$  as  $\mathbf{1}^T D_\Psi \Pi \mathbf{1} = 1$ . Then  $\Psi = D_\Psi W G$  where  $D_\Psi W (\text{diag}(\mathbf{1}^T D_\Psi W))^{-1} = A$ ,  $\text{diag}(\mathbf{1}^T D_\Psi W) = \Theta$ , and  $G = B$ . We obtain LCA of a two-way table and symmetric PLSA of  $\Psi$  as  $(D_\Psi W (\text{diag}(\mathbf{1}^T D_\Psi W))^{-1})^T \mathbf{1} = \mathbf{1}$ ,  $\mathbf{1}^T \text{diag}(\mathbf{1}^T D_\Psi W) \mathbf{1} = 1$ , and  $G\mathbf{1} = \mathbf{1}$ .  $\square$

In the following we will only compare NMF with LBA, EMA, and asymmetric PLSA because the results for the relations of LBA, EMA, and asymmetric PLSA to NMF, and identifiability issues, will also hold for LCA of a two-way table and symmetric PLSA.

NMF of  $\Phi$  is more general than LBA, EMA, and asymmetric PLSA of  $\Pi$ . In NMF the matrix  $\Phi$  to be decomposed has non-negative elements, but is further unrestricted, and the same holds for the two matrices  $M$  and  $H$  into which  $\Phi$  is decomposed. In LBA, EMA, and asymmetric PLSA there are sum-to-1 restrictions to each row of the matrix  $\Pi$  that is to be decomposed, and sum-to-1 restrictions to each row of the two matrices  $W$  and  $G$  into which  $\Pi$  is decomposed. Therefore we have

**Theorem 2.** *LBA, EMA, and asymmetric PLSA are a special case of NMF, as NMF is unrestricted, whereas LBA, EMA and asymmetric PLSA have additional constraints to the matrix that is decomposed in LBA, EMA and asymmetric PLSA, namely  $\Pi\mathbf{1} = \mathbf{1}$  and additional restrictions to the matrices into which it is decomposed, i.e.  $W\mathbf{1} = \mathbf{1}$  and  $G\mathbf{1} = \mathbf{1}$ .*

*Proof.* From (1) we have NMF:  $\Phi = MH$ . From (3) we have LBA, EMA, asymmetric PLSA:  $\Pi = WG$ , with  $\Pi\mathbf{1} = \mathbf{1}$ ,  $W\mathbf{1} = \mathbf{1}$ ,  $G\mathbf{1} = \mathbf{1}$ .  $\square$

Disregarding the objective function and algorithm, a decomposition by NMF and a decomposition by LBA, EMA and asymmetric PLSA are related as follows:

**Theorem 3.** *NMF of  $\Phi$  can be transformed as LBA, EMA, and asymmetric PLSA of  $\Pi$  and the other way around, using  $D_\Phi$ . LBA, EMA, and asymmetric PLSA of  $\Pi$  can be transformed into NMF of  $\Phi$ .*

*Proof.* Suppose that  $(M, H)$  is NMF of  $\Phi$ . We have  $D_\Phi^{-1} \Phi = \Pi$  as  $D_\Phi^{-1} \Phi \mathbf{1} = \mathbf{1}$ . Then  $\Pi = D_\Phi^{-1} M H$  where  $D_H^{-1} H = G$  and  $D_\Phi^{-1} M D_H = W$  as  $D_H^{-1} H \mathbf{1} = \mathbf{1}$  and  $D_\Phi^{-1} M D_H \mathbf{1} = D_\Phi^{-1} M D_H (D_H^{-1} H) \mathbf{1} = D_\Phi^{-1} M H \mathbf{1} = \mathbf{1}$ .

On the other hand, suppose we have LBA, EMA, asymmetric PLSA of  $\Pi$ :  $\Pi = \mathbf{WG}$ , and  $D_\Phi$  is known. We have  $D_\Phi\Pi = \Phi$ . Then  $\Phi = D_\Phi\mathbf{WG}$  where  $D_\Phi\mathbf{W} = \mathbf{M}$  and  $\mathbf{G} = \mathbf{H}$ . We obtain NMF of  $\Phi$  as  $D_\Phi\mathbf{W}$  and  $\mathbf{G}$  are nonnegative.  $\square$

### 3 Motivating examples for the identifiability issue

In the following, we will use results on convex cones and convex hulls (Avis & Bremner, 1995; Gillis, 2020; Sandgren, 1954). Given a matrix  $\mathbf{U}$  of size  $I \times J$ , the convex cone of  $\mathbf{U}$ , denoted by  $\text{cone}(\mathbf{U})$ , is the convex cone generated by the columns of  $\mathbf{U}$ :  $\mathbf{U}(:, j)$  ( $j = 1, \dots, J$ ). That is, for any nonnegative weights  $r_j \geq 0$  ( $j = 1, \dots, J$ ), the weighted sum  $\sum_j r_j \mathbf{U}(:, j) \in \text{cone}(\mathbf{U})$ . It can be expressed as follows:

$$\text{cone}(\mathbf{U}) = \{\mathbf{u} | \mathbf{u} = \sum_{j=1}^J r_j \mathbf{U}(:, j), r_j \geq 0, j = 1, 2, \dots, J\}. \quad (6)$$

The convex hull of  $\mathbf{U}$ , denoted by  $\text{conv}(\mathbf{U})$ , is the convex hull generated by the columns of  $\mathbf{U}$ . That is, for any nonnegative weights  $r_j \geq 0$  ( $j = 1, \dots, J$ ) with the sum of weights  $r_j$  being 1 (i.e.,  $\sum_j r_j = 1$ ), the weighted sum  $\sum_j r_j \mathbf{U}(:, j) \in \text{conv}(\mathbf{U})$ . It can be expressed as follows:

$$\text{conv}(\mathbf{U}) = \{\mathbf{u} | \mathbf{u} = \sum_{j=1}^J r_j \mathbf{U}(:, j), \sum_{j=1}^J r_j = 1, r_j \geq 0, j = 1, 2, \dots, J\} \quad (7)$$

The only difference between  $\text{conv}(\mathbf{U})$  and  $\text{cone}(\mathbf{U})$  is that  $\text{conv}(\mathbf{U})$  adds the sum-to-one constraint on weights  $r_j$  ( $j = 1, \dots, J$ ):  $\sum_{j=1}^J r_j = 1$ . If  $\mathbf{U}(:, 1), \mathbf{U}(:, 2), \dots, \mathbf{U}(:, J)$  are affine independent,  $\text{conv}(\mathbf{U})$  is the so-called  $(J - 1)$ -dimensional simplex and  $\mathbf{U}(:, 1), \mathbf{U}(:, 2), \dots, \mathbf{U}(:, J)$  are so-called vertices of the simplex.

Before we give a definition of identifiability, this section provides two toy examples to illustrate geometrically that LBA, EMA, asymmetric PLSA, and NMF are non-unique. Table 1a is a contingency table of size  $5 \times 2$ , where rows represents self-assessed health and columns represents two genders (Greenacre, 2017); Table 1b is a contingency table of size  $5 \times 3$  where rows represent education groups and columns represent readship class (Greenacre, 2017). Table 2a and Table 2b are normalized versions of Table 1a and Table 1b respectively, where elements in each row sum to 1.

Table 1: (a) Contingency table of self-assessed health with gender (Greenacre, 2017); (b) Contingency table of education group by readship class, where E1, E2, E3, E4, E5 are respective Some primary, Primary completed, Some secondary, Secondary completed, Some tertiary and C1, C2, C3 are respective Glance, Fairly thorough, Very thorough (Greenacre, 2017).

(a) Data source: Exhibit 16.1 Greenacre (2017)		(b) Data source: Exhibit 3.1 Greenacre (2017)		
	Male	Female	C1	C2
Very good	448	369	E1	5
Good	1789	1753	E2	18
Regular	636	859	E3	19
Bad	177	237	E4	12
Very bad	39	64	E5	3
			C3	7
				20
				39
				49
				16

Table 2: (a) Normalized Table 1a such that each row is sum-to-1; (b) Normalized Table 1b such that each row is sum-to-1

(a) Normalized Table 1a		(b) Normalized Table 1b		
	male	female	C1	C2
very good	0.548	0.452	E1	0.357
good	0.505	0.495	E2	0.214
regular	0.425	0.575	E3	0.218
bad	0.428	0.572	E4	0.119
very bad	0.379	0.621	E5	0.115
			C3	0.500
				0.143
				0.238
				0.448
				0.485
				0.615

The NMF decomposition  $\Phi = M\mathbf{H}$  means that each row of  $\Phi$  is a linear combination of the rows of  $\mathbf{H}$  weighted by the components of the corresponding row of  $M$  (Berman & Plemmons, 1994; Gillis, 2020). Namely,  $\phi_i = \mathbf{H}^T \mathbf{m}_i = m_{i1}\mathbf{h}_1 + m_{i2}\mathbf{h}_2 + \dots + m_{iK}\mathbf{h}_K$  where  $\phi_i$ ,  $\mathbf{m}_i$ , and  $\mathbf{h}_k$  are column vectors with elements corresponding to the  $i$ th row of  $\Phi$ ,  $i$ th row of  $M$ , and  $k$ th row of  $\mathbf{H}$ , respectively. This provides a nice geometric interpretation to NMF: the rows of  $\Phi$  are inside the convex cone generated by the rows of  $\mathbf{H}$ .

Rows of Table 1a and Table 1b are respectively presented in Figure 1a and Figure 2a as blue points. Both points in red and points in green could serve as two different basis matrices  $\mathbf{H}$ . Thus the NMF solution is not unique.

Likewise, in LBA, EMA, and asymmetric PLSA,  $\Pi = \mathbf{W}\mathbf{G}$  means that  $\pi_i = \mathbf{G}^T \mathbf{w}_i = w_{i1}\mathbf{g}_1 + w_{i2}\mathbf{g}_2 + \dots + w_{iK}\mathbf{g}_K$  (Berman & Plemmons, 1994; Gillis, 2020). Geometrically, this gives LBA, EMA, and asymmetric PLSA an even nicer interpretation due to the additional constraint  $\sum_{k=1}^K w_{ik} = 1$ : the rows of  $\Pi$  are contained in the convex hull generated by the rows of  $\mathbf{G}$ . This convex hull forms a simplex since  $\mathbf{g}_2 - \mathbf{g}_1, \dots, \mathbf{g}_K - \mathbf{g}_1$  are affine independent, which follows from the condition  $\text{rank}(\mathbf{G}) = K$  (Fu, Ma, Huang, & Sidiropoulos,

Figure 1: Geometric illustration that the solutions of NMF, LBA, EMA, and asymmetric PLSA are not unique using Table 1a and Table 2a where rows in Table 1a and Table 2a are in blue and two possible basis matrices are in green and in red.

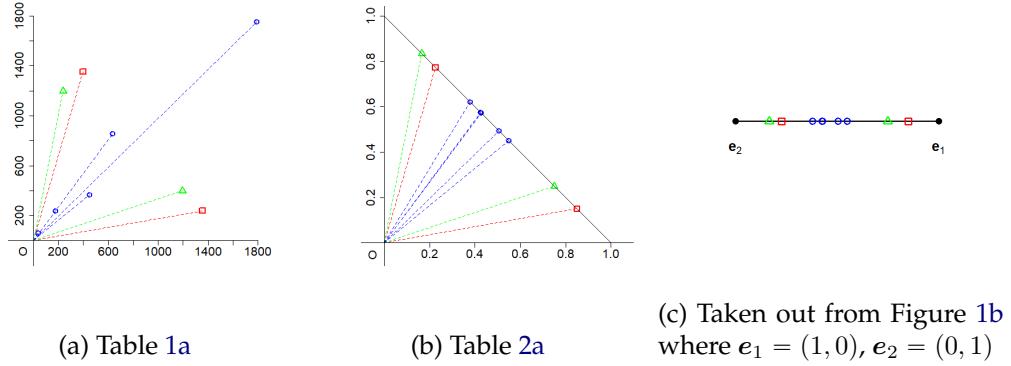
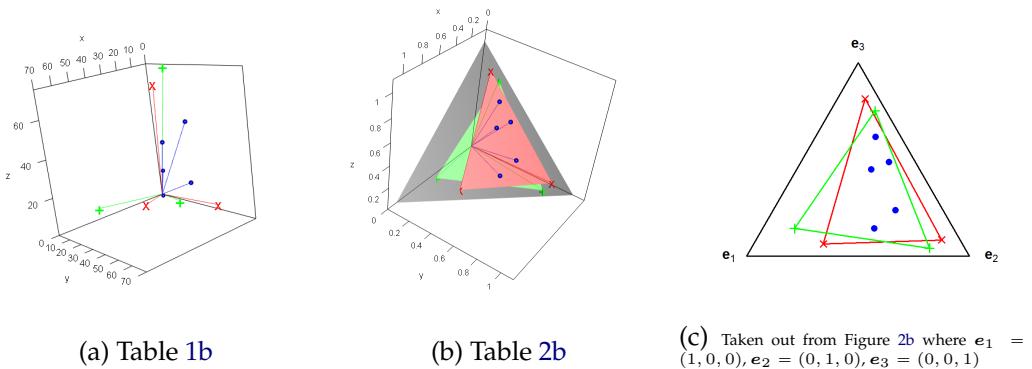


Figure 2: Geometric illustration that the solutions of NMF, LBA, EMA, and asymmetric PLSA are not unique using Table 1b and Table 2b where rows in Table 1b and Table 2b are in blue and two possible basis matrices are in green and in red.



2015). The  $I$  data points  $\pi_1, \dots, \pi_I \in \mathbb{R}_+^J$  lie in the simplex. The goal of LBA, EMA, and asymmetric PLSA is to identify the vertices  $g_1, \dots, g_K$  of a simplex and to estimate the coefficient matrix  $\mathbf{W}$  of the data points (Ge & Zou, 2015).

See Figure 1b and Figure 2b for Table 2a and Table 2b, respectively. Figure 1c and Figure 2c are the simplexes taken out from Figure 1b and Figure 2b respectively, which shows that the sum-to-one restriction for each row reduces the dimensionality with one. Both red points and green points are possible basis matrices  $G$ , demonstrating that the LBA, EMA, and asymmetric PLSA solutions are not unique.

## 4 Identifiability issue

The identifiability issue of NMF, LBA, EMA, and asymmetric PLSA is also known as the nonuniqueness issue (Gillis, 2020). The uniqueness of a NMF decomposition is formally defined as follows (Gillis, 2020; Huang et al., 2014).

**Definition 1.** *The NMF solution  $(\mathbf{M}, \mathbf{H})$  of  $\Phi$  is said to be unique if and only if any other NMF solution  $(\tilde{\mathbf{M}}, \tilde{\mathbf{H}})$  has the form*

$$\tilde{\mathbf{M}} = \mathbf{M}(\boldsymbol{\Gamma}\boldsymbol{\Sigma})^{-1} \text{ and } \tilde{\mathbf{H}} = (\boldsymbol{\Gamma}\boldsymbol{\Sigma})\mathbf{M} \quad (8)$$

where  $\boldsymbol{\Gamma}$  is a permutation matrix and  $\boldsymbol{\Sigma}$  is a diagonal matrix whose diagonal entries are positive.

The same definition is given for the LBA, EMA, and asymmetric PLSA solution  $(\mathbf{W}, \mathbf{G})$  to be unique, except that the diagonal matrix  $\boldsymbol{\Sigma}$  can be omitted, as mentioned in Section 2.2, because each row of the transformation matrix  $\mathbf{T}$  is constrained to sum to one (De Leeuw et al., 1990).

**Definition 2.** *The LBA, EMA, and asymmetric PLSA solution  $(\mathbf{W}, \mathbf{G})$  of  $\Pi$  is said to be unique if and only if any other LBA, EMA, and asymmetric PLSA solution  $(\tilde{\mathbf{W}}, \tilde{\mathbf{G}})$  has the form*

$$\tilde{\mathbf{W}} = \mathbf{W}\boldsymbol{\Gamma}^{-1} \text{ and } \tilde{\mathbf{G}} = \boldsymbol{\Gamma}\mathbf{G} \quad (9)$$

where  $\boldsymbol{\Gamma}$  is a permutation matrix.

It is worth noting that when rank  $K = 1$ , NMF, LBA, EMA, and asymmetric PLSA are unique because the transformation matrix  $\mathbf{S}$  in NMF is a scalar and  $\mathbf{T}$  in LBA, EMA, asymmetric PLSA is 1.

### 4.1 The solution of NMF is unique if and only if the solution of LBA, EMA, and asymmetric PLSA is unique.

We now discuss how uniqueness of NMF is related to uniqueness of LBA, EMA, and asymmetric PLSA. We provide the following theorem:

**Theorem 4.** *The solution of NMF of  $\Phi$  is not unique if and only if the solution of LBA, EMA, and asymmetric PLSA of  $\Pi$  is not unique.*

*Proof.* Suppose that the solution of NMF of  $\Phi$  is not unique. Thus  $S$  is not simply a permutation matrix, see Definition 1. We have  $\Phi = M\mathbf{H} = (MS^{-1})(SH) = \tilde{M}\tilde{\mathbf{H}}$ , as in Eq. (2). From Theorem 3, both  $(W, G)$  and  $(\tilde{W}, \tilde{G})$  are solutions of LBA, EMA, and asymmetric PLSA of  $\Pi$ , where  $W = D_{\Phi}^{-1}MD_H$ ,  $G = D_H^{-1}H$ ,  $\tilde{W} = D_{\Phi}^{-1}(MS^{-1})D_{(SH)}$ , and  $\tilde{G} = D_{(SH)}^{-1}(SH)$ . Thus, we have  $\tilde{W} = W(D_H^{-1}S^{-1}D_{(SH)})$  and  $\tilde{G} = (D_{(SH)}^{-1}SD_H)G$ . Because  $D_{(SH)}^{-1}SD_H$  is not a permutation matrix, the solution of LBA, EMA, and asymmetric PLSA of  $\Pi$  is not unique.

On the other hand, suppose that the solution of LBA, EMA, and asymmetric PLSA of  $\Pi$  is not unique. Thus  $T$  is not simply a permutation matrix, see Definition 2. We have  $\Pi = WG = (WT^{-1})(TG) = \tilde{W}\tilde{G}$ , as in Eq. (4). From Theorem 3, both  $(M, H)$  and  $(\tilde{M}, \tilde{H})$  are solutions of NMF of  $\Phi$ , where  $M = D_{\Phi}W$ ,  $H = G$ ,  $\tilde{M} = D_{\Phi}WT^{-1}$ ,  $\tilde{H} = TG$ . Because  $T$  is not a permutation matrix, the solution of NMF is not unique.  $\square$

A statement is true if and only if its contrapositive is true. Thus, from Theorem 4, we immediately have:

**Theorem 5.** *The solution of NMF is unique if and only if the solution of LBA, EMA, and asymmetric PLSA is unique.*

## 4.2 Proposals for identifiability of LBA, EMA, PLSA

We have illustrated that, in general, the solution of asymmetric PLSA is not unique. However, this is not always acknowledged in the PLSA literature, see, for instance, [Gaussier and Goutte \(2005\)](#).

In contrast, the identifiability issue in LBA is well recognized ([De Leeuw et al., 1990; Van der Ark et al., 1999; Van der Heijden, Mooijaart, & De Leeuw, 1992](#)). A classic proposal aiming at addressing the identifiability issue involves the, what they call, inner extreme solution and outer extreme solution, which are determined by the transformation matrix  $T$ :  $(WT^{-1}, TG)$ . The inner extreme solution selects  $T$  such that the basis vectors in the rows of  $TG$  are as similar as possible, while outer extreme solution selects  $T$  such that their difference is maximized.

For  $K = 2$ , the transformation matrix  $T$  is characterized by two parameters  $x, y$ :  $T = \begin{bmatrix} x & 1-x \\ y & 1-y \end{bmatrix}$  with feasible ranges for  $(x, y)$  governed by non-negative constraints in  $W$  and

$\mathbf{G}$  (De Leeuw et al., 1990; Moussaoui, Brie, & Idier, 2005). The following theorem is given:

**Theorem 6.** [(De Leeuw et al., 1990; Van der Ark et al., 1999)] When  $K = 2$ , both inner and outer extreme solutions are unique. The solution is inner extreme if and only if  $\mathbf{W}$  contains a permuted  $2 \times 2$  identity matrix, while the solution is outer extreme if and only if  $\mathbf{G}$  contains a permuted  $2 \times 2$  diagonal matrix.

This means that, for the inner extreme solution, two rows of  $\mathbf{G}$  are picked from two rows of  $\mathbf{\Pi}$  and thus the coefficient matrix  $\mathbf{W}$  includes a permuted identity matrix. As an example, in Figure 1c basis vectors of the inner extreme solution are the most extreme blue data point on the left and the most extreme blue data point on the right. In contrast, for the outer extreme solution, two rows of  $\mathbf{G}$  are  $e_1 = (1, 0)$  and  $e_2 = (0, 1)$ , i.e. the two black end points in Figure 1c.

For  $K > 2$ , the terms “as similar as possible” and “as different as possible” have to be explicitly defined. To quantify similarity and dissimilarity between basis vectors from the rows of  $\mathbf{G}$ , Van der Ark et al. (1999) proposed three criteria: (1) the sum of distances between the basis vectors, (2) the volume spanned by the basis vectors, (3) statistical dependence of basis vectors. Van der Ark et al. (1999) elaborated the first criterion. They propose a Monte Carlo method to find a transformation matrix  $\mathbf{T}$ , that has the property that the sum of chi-square distances between basis vectors

$$\sum_{k=1}^K \sum_{k'=k+1}^K \sqrt{\sum_{j=1}^J \frac{(g_{kj} - g_{k'j})^2}{\sum_i \phi_{ij}}} \quad (10)$$

is either minimized (“inner extreme solution”) or maximized (“outer extreme solution”), where  $g_{kj}$  is element  $(k, j)$  of  $\mathbf{G}$  and  $\phi_{ij}$  is element  $(i, j)$  of  $\Phi$ . The chi-square distance, playing a central role in methods like correspondence analysis (Qi, Hessen, Deoskar, & Van der Heijden, 2024), weights the squared difference between basis vectors elements  $(g_{kj} - g_{k'j})^2$  by the marginal proportion  $\sum_i \phi_{ij}$ . This adjustment corrects the difference  $(g_{kj} - g_{k'j})^2$  for the size of column  $j$  (Van der Ark et al., 1999).

For  $K > 2$ , inner or outer extreme solutions may be not unique. For example, for  $K = 3$ , there could be multiple triangles formed by the rows of different basis matrices  $\mathbf{G}$  that all achieve the same minimum/maximum total chi-square distance, rather than a single unique solution. Figure 2c illustrates this, assuming that the basis matrix  $\mathbf{G}$  for the

red points has the same total chi-square distance as the basis matrix  $\mathbf{G}$  for the blue points.

For EMA, minimum end members (Renner, 1993, 1995; Seidel & Hlawitschka, 2015; Weltje, 1997; Weltje & Prins, 2007) and outermost end members (Zhang, Wang, Xu, & Yang, 2020) are proposed to arrive at uniqueness. These are analogous to inner and outer extreme solutions respectively, but there is no sufficient condition provided for the EMA solution to be unique.

In addition to the inner extreme solution and the outer extreme solution in LBA, Van der Heijden et al. (1992) discussed obtaining uniqueness by constraining the parameters. They suggested three types of constraints for LBA: fixed value, equality, and multinomial logit constraints. However, sufficient conditions to guarantee the unique solution are not provided. In this paper we ignore the constraints, but note that for  $K = 2$  the inner extreme solution and outer extreme solution can be obtained by imposing zero-value constraints to  $\mathbf{W}$  and  $\mathbf{G}$ , respectively.

### 4.3 Proposals for identifiability of NMF

The sufficient conditions for NMF to have an unique solution are well studied, where separability and minimum volume are two main assumptions (Gillis, 2020). Note that many NMF work takes  $\mathbf{M}$  as basis matrix and  $\mathbf{H}$  as coefficient matrix. Here, to keep consistent with LBA, EMA, and asymmetric PLSA, we take  $\mathbf{M}$  as coefficient matrix and  $\mathbf{H}$  as basis matrix. This can be achieved by  $\Phi^T = \mathbf{H}^T \mathbf{M}^T$ .

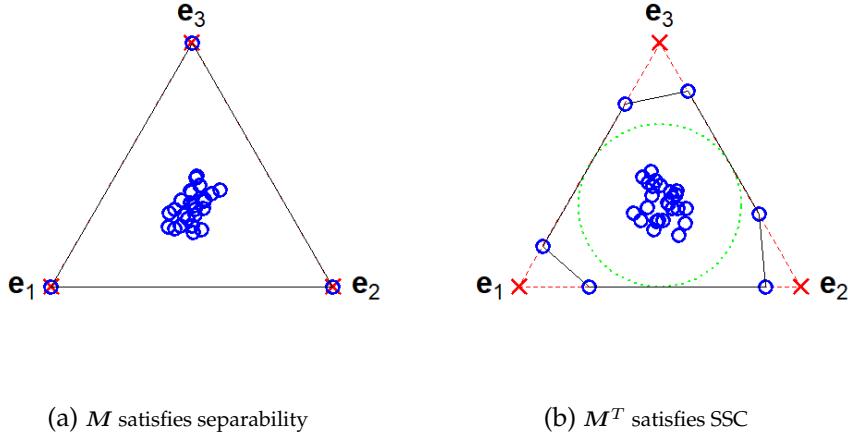
Separability is first introduced by Donoho and Stodden (2003). A matrix  $\mathbf{M}$  of size  $I \times K$  is said to be separable or have a separability property if it contains a  $K \times K$  submatrix that is a permutation of a diagonal matrix with positive diagonal entries. In LBA, due to  $\mathbf{W}\mathbf{1} = \mathbf{1}$ , this corresponds to  $\mathbf{W}$  containing an identity matrix of size  $K \times K$  up to permutation. The uniqueness theorem of NMF for separability can be described as follows:

**Theorem 7.** [(Chan, Chi, Huang, & Ma, 2009; Gillis, 2020; Zhou, Xie, Yang, Yang, & He, 2011)]  
*Suppose that  $\text{rank}(\Phi) = \text{rank}(\mathbf{M}) = \text{rank}(\mathbf{H})$  and the coefficient matrix  $\mathbf{M}$  is separable. Then the NMF solution  $(\mathbf{M}, \mathbf{H})$  of  $\Phi$  is unique.*

Separability on  $\mathbf{M}$  means that the rows of  $\mathbf{H}$  are themselves data points up to scaling. When  $K = 2$ , Theorem 6 for inner extreme solution of LBA coincides with Theorem 7, but Theorem 7 considers uniqueness for any dimensionality  $K$ .

See Figure 3a for an illustration that a coefficient matrix  $M$  satisfies the separability assumption for  $K = 3$  (Gillis, 2020).  $M$  is size of  $28 \times 3$ . That is,  $M$  has 28 data points and each data point is a three-dimensional vector. Assuming that viewer stands in the nonnegative orthant, faces the origin, and looks at the two-dimensional plane  $x_1 = 1$  (Gillis, 2020), we have Figure 3a. In the figure, the blue dots "o" are rows of  $M$  and the red crosses "X" are standard basis vectors  $e_1, e_2, e_3$  ( $e_k$  denoting a three-dimensional standard basis vector whose  $k$ th element is 1, otherwise 0). The triangle is nonnegative orthant cone( $e_1, e_2, e_3$ ) and the polygon formed by the dots is cone( $M^T$ ). From the figure, we can see that cone( $M^T$ ) coincides with cone( $e_1, e_2, e_3$ ). For every  $k \in \{1, 2, 3\}$ , there exists a row  $i \in \{1, \dots, I\}$ , such that  $m_i = \alpha_k e_k$  where  $\alpha_k$  is a scalar. Note that, despite theoretical appeal, separability is rarely satisfied in practice because of the fact that true basis vectors (the rows of  $H$ ) may not be data points (the rows of  $\Phi$ ) up to scaling.

Figure 3: Geometric illustrations of (a) separability of a matrix  $M$ , (b) SSC of a matrix  $M^T$  by assuming that viewer stands in the nonnegative orthant, faces the origin, and looks at the two-dimensional plane  $x_1 = 1$  (Gillis, 2020). The blue dots "o" are rows of  $M$ ; the red crosses "X" are standard basis vectors  $e_1, e_2, e_3$ .



The minimum volume assumption remains robust when the separability assumption is violated. Under the minimum volume assumption, the determinant

$$\det(HH^T) \tag{11}$$

is minimum (Fu et al., 2015; Lin, Ma, Li, Chi, & Ambikapathi, 2015). The rationale for using  $\det(HH^T)$  as the objective function in Equation (11) is that  $\sqrt{\det(HH^T)}/K!$  corresponds

to the volume of the simplex  $\text{conv}(\mathbf{h}_1, \dots, \mathbf{h}_K)$  and the origin where  $\mathbf{h}_k$  is row  $k$  from basis matrix  $\mathbf{H}$  (Leplat, Gillis, & Ang, 2020).

The minimum volume assumption in NMF is analogous to the inner extreme solution in LBA (Fu et al., 2015; Lin et al., 2015; Van der Ark et al., 1999): the former seeks basis vectors (i.e., rows of  $\mathbf{H}$ ) whose convex hull encloses all data points (i.e., rows of  $\Phi$ ) with minimum volume, whereas the latter seeks basis vectors (i.e., rows of  $\mathbf{G}$ ) whose convex hull encloses all data points (i.e., rows of  $\Pi$ ) and whose total chi-square distance from each other is minimum. Actually, Van der Ark et al. (1999) also proposed the volume criteria, as mentioned in above Section, but they did not investigate it.

Under the minimum volume assumption, several sufficient conditions have been proposed to ensure that  $\Phi$  has an unique NMF solution  $(\mathbf{M}, \mathbf{H})$  (Fu, Huang, & Sidiropoulos, 2018; Fu et al., 2015; Leplat et al., 2020; Lin et al., 2015). Fu et al. (2015) provide the following theorem based on the so-called sufficient scattered conditions (SSC).

**Theorem 8.** [Fu et al. (2015)] Suppose that  $\text{rank}(\Phi) = \text{rank}(\mathbf{M}) = \text{rank}(\mathbf{H})$  and the minimum volume assumption holds. If  $\mathbf{M}\mathbf{1} = \mathbf{1}$  and  $\mathbf{M}^T$  satisfies the so-called sufficient scattered conditions (SSC), the NMF solution  $(\mathbf{M}, \mathbf{H})$  of  $\Phi$  is unique.

A matrix  $\mathbf{M}^T$  is sufficient scattered if (1) SSC1:  $\mathbb{C} \subseteq \text{cone}(\mathbf{M}^T)$ , where  $\mathbb{C} = \{x \in \Re^K | \mathbf{1}^T x \geq \sqrt{K-1} \|x\|_2\}$  is second-order cone, (2) SSC2: there does not exist any orthogonal matrix  $\mathbf{Q} \in \Re^{K \times K}$  such that  $\text{cone}(\mathbf{M}^T) \subseteq \text{cone}(\mathbf{Q})$  except for permutation matrices. Lin et al. (2015) proposed another sufficient condition for NMF to be unique under minimum volume assumption. Fu, Huang, Yang, Ma, and Sidiropoulos (2016) later proved this condition equivalent to the one in Theorem 8.

Figure 3b serves as an illustration that a matrix  $\mathbf{M}^T$  satisfies SSC for  $K = 3$  (Gillis, 2020).  $\mathbf{M}$  has size  $31 \times 3$ . As in Figure 3a, by assuming that a viewer stands in the non-negative orthant, faces the origin, and looks at the two-dimensional plane  $x_1 = 1$  (Gillis, 2020), we have Figure 3b. The triangle is a nonnegative orthant cone( $e_1, e_2, e_3$ ), the polygon formed by the dots is  $\text{cone}(\mathbf{M}^T)$ , and the circle is the second-order cone  $\mathbb{C}$ . From the figure, we can see that the circle is contained in the polygon formed by the dots. Thus SSC1 holds.

Any permutation matrix simply reorders the standard basis vectors, and thus the cone generated by the columns of any permutation matrix is exactly the triangle in the figure.

Any orthogonal matrix is a rotated version of the triangle in the figure. We can see that no rotated version of this triangle (except for itself) can contain the polygon formed by the dots. Thus SSC2 hold.

Theorem 8 has the condition that the sum of each row of coefficient matrix  $M$  is 1:  $M\mathbf{1} = \mathbf{1}$ . Instead of  $M\mathbf{1} = \mathbf{1}$ , Fu et al. (2018) extended this result to the case where  $M^T\mathbf{1} = \mathbf{1}$ ; Leplat et al. (2020) proved the case where  $H\mathbf{1} = \mathbf{1}$ . Beyond separability and minimum volume approaches, other frameworks exist for ensuring NMF to be unique such as those in Abdolali, Barbarino, and Gillis (2024); Abdolali and Gillis (2021); Chan et al. (2009); Ge and Zou (2015); Gillis (2012); Gillis and Rajkó (2023); Huang et al. (2014); Javadi and Montanari (2020); Laurberg, Christensen, Plumbley, Hansen, and Jensen (2008); Lin, Wu, Ma, Chi, and Wang (2018). Necessary conditions have also been extensively studied such as those in Gillis (2020); Huang et al. (2014); Lin et al. (2015); Moussaoui et al. (2005). Based on Theorem 5, these theorems for the NMF solution to be unique provide conditions for the LBA, EMA, and PLSA solution to be unique.

#### 4.4 Conclusion

The identification of the solution by making basis vectors be as similar as possible, called inner extreme solution, minimum end members, and minimum volume in LBA, EMA, and NMF, respectively, is conceptually similar. However, the measures for the concept "as similar as possible" are different. The unambiguous volume measurement  $\det(H^T H)$  in NMF may be the reason that identifiability theorems have been developed so well in this research field.

Although identifiability theorems for NMF have been proposed, their sufficient conditions—such as separability and sufficient scatter—remain restrictive and may fail to hold in practice. Most theorems depend on the sparsity of either the basis or the coefficient matrix. Consequently, they are generally unsuitable for cases where the true basis and coefficient matrices lack this sparsity property.

Last, the assumption that basis vectors are as different as possible is explored in LBA and EMA under the name outer extreme solution and outermost end members, respectively. However, the identifiability theorem with respect to the assumption is not investigated much. As far as we know, the only relevant identifiability theorem is Theorem 6,

which is limited to dimensionality  $K = 2$ .

## 5 Algorithms

In NMF, the observed matrix  $\mathbf{X}$  is partitioned in a signal part  $\Phi = \mathbf{M}\mathbf{H}$  and error part  $\mathbf{E}_\mathbf{X}$  (Gillis, 2020):

$$\mathbf{X} = \mathbf{M}\mathbf{H} + \mathbf{E}_\mathbf{X} \quad (12)$$

where  $\mathbf{M}$  and  $\mathbf{H}$  are nonnegative matrices of size  $I \times K$  and  $K \times J$ , respectively, and  $\mathbf{E}_\mathbf{X}$  is a matrix of size  $I \times J$ . Similarly, in LBA, EMA, and asymmetric PLSA, the observed conditional proportion  $\mathbf{P}$  can be partitioned as the signal part  $\Pi = \mathbf{W}\mathbf{G}$  and the error part  $\mathbf{E}_\mathbf{P}$  (Dietze, Schulte, & Dietze, 2022; Van Hateren, Prins, & Van Balen, 2018; Van der Heijden, 1994; Zhang et al., 2020):

$$\begin{aligned} \mathbf{P} &= \mathbf{W}\mathbf{G} + \mathbf{E}_\mathbf{P} \\ \text{subject to } \mathbf{P}\mathbf{1} &= \mathbf{1}, \mathbf{W}\mathbf{1} = \mathbf{1}, \mathbf{G}\mathbf{1} = \mathbf{1} \end{aligned} \quad (13)$$

where  $\mathbf{W}$  and  $\mathbf{G}$  are nonnegative matrices of size  $I \times K$  and  $K \times J$ , respectively, and  $\mathbf{E}_\mathbf{P}$  is a matrix of size  $I \times J$ . Dimensionality  $K$  is a hyperparameter pre-defined by the researcher.

Given a choice for  $K$ , NMF (LBA, EMA, asymmetric PLSA) involves estimating both the coefficient matrix  $\mathbf{M}$  ( $\mathbf{W}$ ) and the basis matrix  $\mathbf{H}$  ( $\mathbf{G}$ ) from the observed matrix  $\mathbf{X}$  ( $\mathbf{P}$ ) under nonnegative constraints (nonnegative and sum-to-one constraints), and error matrix  $\mathbf{E}_\mathbf{X}$  ( $\mathbf{E}_\mathbf{P}$ ) can be determined by  $\mathbf{X} - \mathbf{M}\mathbf{H}$  ( $\mathbf{P} - \mathbf{W}\mathbf{G}$ ). The objective function for NMF (LBA, EMA, asymmetric PLSA) is closely related to the distributional assumptions of the data. For example, if the data are supposed to follow a Gaussian distribution, or do not have any distributional assumption, the Frobenius norm  $\|\mathbf{X} - \mathbf{M}\mathbf{H}\|_F^2$  ( $\|\mathbf{P} - \mathbf{W}\mathbf{G}\|_F^2$ ) can be used. If the data follow a Laplace distribution, the  $l_1$  norm  $\|\mathbf{F} - \mathbf{M}\mathbf{H}\|_1$  ( $\|\mathbf{P} - \mathbf{W}\mathbf{G}\|_1$ ) is more appropriate. Some objective functions can be found in Gillis (2020); Saberi-Movahed et al. (2025). Also, a regularizer can be added to the objective function, such as the determinant of  $\mathbf{H}\mathbf{H}^T$  and the sum of distances between basis vectors (Ang et al., 2025; Ang & Gillis, 2019; Zhang et al., 2020).

There is no closed solution for  $(\mathbf{M}, \mathbf{H})$  ( $(\mathbf{W}, \mathbf{G})$ ) in NMF (LBA, EMA, asymmetric PLSA) (Gillis, 2020; Paterson & Heslop, 2015). The objective function involving both the

coefficient matrix and basis matrix is generally non-convex. Thus maximizing or minimizing the objective function does not necessarily lead to the global optimum. Nevertheless, some algorithms can ensure convergence to a local optimum. Most algorithms employ an iterative alternating update scheme for the pair  $(M, H)$  or  $(W, G)$ : each time optimizing one matrix while keeping the other one fixed.

Some algorithms are one-stage, in which the pair  $(M, H)$  or  $(W, G)$  is obtained directly from  $X$  or  $P$ , respectively (Leplat, Ang, & Gillis, 2019; Zhang et al., 2020). Other algorithms are two-stage (Fu et al., 2018; Mooijaart, Van der Heijden, & Van der Ark, 1999; Weltje, 1997). I.e., first an initial lower rank approximation  $\hat{\Phi}^0$  or  $\hat{\Pi}^0$  is computed. Then, in the second stage, the pair  $(M, H)$  or  $(W, G)$  is derived that may, due to the non-negativity constraints of all matrices, not equal  $\hat{\Phi}^0$  or  $\hat{\Pi}^0$  respectively.

This section provides a brief outline about algorithms of LBA, EMA, PLSA, and NMF. The estimated approximations for  $\Phi$ ,  $\Pi$ ,  $W$ ,  $G$ ,  $M$ ,  $H$  are denoted by  $\hat{\Phi}$ ,  $\hat{\Pi}$ ,  $\hat{W}$ ,  $\hat{G}$ ,  $\hat{M}$ ,  $\hat{H}$ . We also briefly introduce algorithm for LCA of two-way tables.

## 5.1 LBA

Computing inner extreme solution and outer extreme solution is a two-stage algorithm. First, non-unique non-negative LBA parameters  $W$  and  $G$  are estimated. As a result a non-negative lower rank matrix  $\Phi$  is estimated. For this, two frameworks are proposed.

In the first framework, for the situation where the observed data are frequencies, one may assume that the observations follow a product-multinomial distribution. For this situation, De Leeuw et al. (1990) and Van der Heijden et al. (1992) proposed a maximum likelihood estimator (MLE). To compute the MLE, De Leeuw et al. (1990) employed the idea of majorization from convex analysis (Sun, Babu, & Palomar, 2017), while Van der Heijden et al. (1992) used the expectation maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). These two approaches lead to the same iterative update scheme. The iterative process increases the likelihood at every step and therefore converges to a maximum (De Leeuw et al., 1990; Van der Heijden et al., 1992). This maximum may be a local maximum, so different sets of starting values should be tried to have some confidence that the local maximum is also the global maximum. As the EM algorithm only has linear convergence, convergence may be slow. Also, when in the iteration process certain parameters

obtain a value of zero, these parameters cannot move away from this zero value, and it is possible that this leads to a local maximum of the likelihood that is not the global maximum. Nonnegative and sum-to-one constraints are preserved throughout the iterative process, provided that initial estimates for  $\hat{\mathbf{W}}$  and  $\hat{\mathbf{G}}$  satisfy these constraints.

In the second framework, one can make use of a constrained weighted least squares estimator (CWLSE), which has no distributional assumptions (Mooijaart et al., 1999; Van der Ark, 1999). For CWLSE, the active constraint method (ACM) is used to deal with the sum-to-1 and nonnegative constraints of the LBA parameters. Minimizing the CWLSE will lead to a local minimum, and also here, different sets of starting values should be tried to investigate if this local minimum is also the global minimum.

After finding a non-negative lower rank matrix and initial - i.e. not identified - estimates of  $\mathbf{W}$  and  $\mathbf{G}$ , the second stage is to find the transformation matrix  $\mathbf{T}$  by Metropolis algorithm, so that the sum of chi-squares distances between the rows of basis vectors  $\mathbf{T}\mathbf{G}$  is minimum or maximum (See Equation (10)).

Jelihovschi and Allaman (2018) provided R package *lba* for inner and outer extreme solutions.

## 5.2 Symmetric and asymmetric PLSA

PLSA, as its name indicates, was inspired by latent semantic analysis (LSA) (Hofmann, 2001; Hofmann, Puzicha, & Jordan, 1999). LSA is based on an SVD with an objective function using the Frobenius norm  $\|\mathbf{X} - \mathbf{MH}\|_F^2$ . The basis matrix and coefficient matrix of LSA are (normalized) orthogonal and may contain negative entries. LSA projects a document-term matrix or a word-context matrix into a lower-dimensional latent space.

In contrast to LSA, PLSA is based on the statistical model LCA (Goodman, 1974b, 1987; Hofmann, 1999, 2001; Hofmann et al., 1999). For LCA, the EM algorithm is a commonly used algorithm for maximum likelihood estimator (MLE) (Fienberg et al., 2009; Gilula, 1979, 1983; Vermunt & Magidson, 2016). Under identical initialization, the LCA solution obtained via the EM algorithm can be recovered through an appropriate transformation of the EM solutions from LBA (Clogg, 1981; Van der Heijden et al., 1989).

However, in many machine learning or text mining related experiments, MLE is often plagued by overfitting, yielding a higher rank for the lower rank approximation than

necessary. In order to avoid overfitting a generalization of MLE is proposed for PLSA (Hofmann, 1999, 2001) that makes use of a so-called temperature controlled version of the EM algorithm for model fitting. For this PLSA introduces a control parameter  $\beta$  to EM algorithm, where  $0 < \beta \leq 1$ . Note that  $\beta = 1$  provides the standard EM algorithm. When  $\beta = 1$ , the EM estimates of the asymmetric PLSA are identical to the EM estimates obtained from LBA, whereas the EM estimates of the symmetric PLSA are identical to those obtained from LCA.

The R package *svs* is available to implement symmetric and asymmetric PLSA, where there is a hyperparameter to adjust the control parameter (Plevoets, 2024).

### 5.3 EMA

For estimation, two main EMA approaches have been proposed: non-parametric EMA and parametric EMA (Paterson & Heslop, 2015; Peng, Wang, Zhao, & Dong, 2023; Renny et al., 2026; Van Hateren et al., 2018; Weltje & Prins, 2003, 2007; Zhang et al., 2025). Non-parametric EMA estimates basis vectors from the data set itself, whereas parametric EMA assumes different basis vectors following the same family of distributions, such as lognormal, with different parameters. However, parametric EMA may fail to reveal true basis vectors because true basis vectors may not conform to the chosen distributional form. This paper focuses on nonparametric EMA.

In end-member modeling, finding minimum end members, i.e., basis vectors that enclose data samples as tightly as possible, seems to be preferred (Dietze et al., 2022; Seidel & Hlawitschka, 2015; Van Hateren et al., 2018; Weltje, 1997; Weltje & Prins, 2007). The end-member modeling algorithm (EMMA), proposed by Weltje (1997), has been a benchmark algorithm for this purpose. EMMA consists of two stages and is described briefly as follows. In the first stage, an initial signal part  $\hat{\Pi}^0$  is estimated as  $\hat{\Pi}^0 = \mathbf{A}_K \mathbf{B}_K$ , where  $\mathbf{A}_K$  and  $\mathbf{B}_K$  are derived from the first  $K$  columns of a singular value decomposition (SVD). Since  $\hat{\Pi}^0$  may have negative values, a constrained least squares correction is applied to eliminate the negative values. In the second stage a simplex expansion algorithm estimates  $\hat{\mathbf{W}}$  and  $\hat{\mathbf{G}}$ . Basis matrix  $\mathbf{G}$  is initialized using fuzzy cluster centers of fuzzy clustering algorithm (FCM) rather than  $\mathbf{B}_k$  because  $\mathbf{B}_k$  may have negative values. Each basis vector is then updated via a transformation matrix  $\mathbf{T}$ , expanding the simplex and reducing negative

coefficients in  $\hat{W}$ . The procedure iterates until  $\hat{W}$  becomes nonnegative or the remaining negative values fall within tolerance. The remaining (trivial) negative values are removed by constrained least squares. We note that, due to the handling of the negative values in stages 1 and 2, the end result of this algorithm is not necessarily optimal in a least-squares sense (Van der Heijden, 1994).

Paterson and Heslop (2015) proposed an alternative EMA algorithm for minimum end members, developed from hierarchical alternating least squares based non-negative matrix factorization (HALS-NMF) (Chen & Guillaume, 2012), based on NMF instead of SVD. In this NMF algorithm, both the coefficient matrix and basis matrix are nonnegative, making it an appealing approach for EMA and eliminating the need for nonnegativity corrections. The alternative EMA algorithm is a one-stage algorithm. The row-sum-to-one constraint for the coefficient matrix and the minimum distance constraint for the basis matrix are parts of the objective function (Paterson & Heslop, 2015). After each iterative update, this new EMA algorithm normalizes basis vectors to be row-sum-to-one.

In the identification choice for minimum end members, because the basis vectors are as close as possible, the algorithms have trouble factorizing a highly mixed dataset where no single data point is near basis vector (Paterson & Heslop, 2015; Van Hateren et al., 2018; Zhang et al., 2020). For this, like the outer extreme solution in LBA, a basic end-member model algorithm (BasEMMA) is proposed by Zhang et al. (2020) to seek basis vectors as different as possible. BasEMMA has been adopted in subsequent studies, including Ferreira, Nunes, Goya, and Mahiques (2025); Lin et al. (2025); Zhang et al. (2025).

Seidel and Hlawitschka (2015) created R functions for EMMA proposed by Weltje (1997). Paterson and Heslop (2015) provides AnalySize for EMMA as well as the alternative EMA algorithm based on MATLAB. BasEMMA is embedded in the Microsoft Excel program using Visual Basic for Applications (VBA) programming language (Zhang et al., 2020). In addition, there are other algorithms and implements such as those in Dietze and Dietze (2019); Elisabeth et al. (2012); Yu, Colman, and Li (2016).

## 5.4 NMF

Since Lee and Seung (1999)'s publication in *Nature*, many algorithms have been developed for NMF, see Ang and Gillis (2019); Fu et al. (2019); Gillis (2020); Guo et al. (2024);

Hobolth et al. (2020) for a survey. Identifiability conditions of NMF have significant implication on its algorithms.

NMF under the separability assumption means that the input matrix  $\mathbf{X}$  has the form  $\mathbf{X} \approx \mathbf{M}\mathbf{H}$ , where  $\mathbf{M}$  is separable. This is equivalent to assuming that rows of basis matrix  $\mathbf{H}$  can be taken directly from the input matrix  $\mathbf{X}$ . Thus the estimation problem of NMF is to find the  $K$  rows from  $\mathbf{X}$  for which the data are sufficiently approximated by the model (Araújo et al., 2001; Barbarino & Gillis, 2025; Bittorf, Recht, Ré, & Tropp, 2012; Gillis, 2013, 2014, 2019; Nascimento & Dias, 2005). These  $K$  data points construct basis matrix  $\hat{\mathbf{H}}$ , and then, given  $\hat{\mathbf{H}}$ ,  $\hat{\mathbf{M}}$  can be estimated by solving a constrained least programming problem (Fu & Ma, 2016; Gillis, 2020). We do not describe algorithms in detail, but refer interested reader to Chapter 7 of the book by Gillis (2020), which describes many algorithms under the separability assumptions.

Under the minimum volume assumption, a major class of algorithms such as those in Andersen Ang and Gillis (2018); Ang and Gillis (2019); Fu et al. (2016); Gillis (2020); Leplat et al. (2019, 2020); Miao and Qi (2007); Zhou et al. (2011) considers to minimize a regularized objective function which consist of two parts: data fitting and volume regularizer. The objective function  $\|\mathbf{X} - \mathbf{M}\mathbf{H}\|_2^2 + \lambda \text{logdet}(\mathbf{H}\mathbf{H}^T + \delta\mathbf{I})$  is commonly used where logdet denotes the logarithm of the determinant of  $\mathbf{H}\mathbf{H}^T + \delta\mathbf{I}$  and  $\delta > 0$  is a small value to avoid that the term tends to negative infinity even when  $\mathbf{H}\mathbf{H}^T$  is singular (Fu et al., 2016). This has the advantage of making the algorithm robust against noise.  $\lambda > 0$  balances the data fitting and the volume regularizer term. This hyperparameter has a significant influence on results. The NMF problem can be handled by alternatively updating the parameters of  $\mathbf{M}$  and  $\mathbf{H}$ , holding the other one fixed. For each sub-question, projected gradient gradient, linear programming, quadratic programming, multiplicative updates are some of the most popular methods. There are other objective functions and related algorithms involving the minimum volume assumption such as those in Bioucas-Dias (2009); Chan et al. (2009); Craig (1994); Fu et al. (2018, 2015); Li and Bioucas-Dias (2008).

Algorithms beyond separability and minimum volume assumptions include distance regularized related algorithms such as those in Ang et al. (2025); Jia and Qian (2009); Mei and He (2011); Yu and Sun (2007), geometry related algorithms such as those in Abdolali and Gillis (2021); Ge and Zou (2015); Lin et al. (2018), and other related algorithms such as

those in [Abdolali et al. \(2024\)](#); [Javadi and Montanari \(2020\)](#).

Most algorithms mentioned here are implemented in MATLAB. Examples include the work by [Barbarino and Gillis \(2025\)](#); [Gillis \(2013\)](#); [Leplat et al. \(2020\)](#). Notable implementations in Python include the works of [Javadi and Montanari \(2020\)](#) and [Ang et al. \(2025\)](#). The book by [Gillis \(2020\)](#) provides extensive MATLAB implementations for many related methods. In addition, some NMF algorithms have been available in R packages such as *NMF* ([Gaujoux & Seoighe, 2010](#); [Gaujoux, Seoighe, & Sauwen, 2024](#)) and *vrnmf* ([Slep'yarskiy et al., 2021](#); [Soldatov, Kharchenko, Petukhov, & Biederstedt, 2021](#)), in Python libraries such as *Nimfa* ([Zitnik & Zupan, 2012](#)) and *scikit – learn* ([Pedregosa et al., 2011](#)), and in other languages ([Guo et al., 2024](#)).

## 6 An example

For comparative purposes, we consider Table 3, a dataset from the social sciences involving time allocation in the Netherlands for a week, in minutes ([Mooijaart et al., 1999](#)). A row is cross-classified by gender (2 levels), age (5 levels) and year (three surveys taken in 1975, 1980, 1985) ([Mooijaart et al., 1999](#)). The columns represents 18 main activities where respondents were asked to keep diaries: (1) paid work, (2) domestic work, (3) caring for members household, (4) shopping, (5) personal need, (6) eating and drinking, (7) sleeping and resting, (8) education, (9) participation in volunteer work, (10) social contacts, (11) going out, (12) sports, hobbies, games, (13) gardening, taking care of pets, (14) recreation outside, (15) tv, radio, audio (16) reading, (17) relaxing, and (18) others ([Mooijaart et al., 1999](#)). The table has been analysed earlier by [Mooijaart et al. \(1999\)](#) using LBA.

In this section, we identify the solutions by making the basis vectors as similar as possible in LBA, EMA, and NMF: known under names inner extreme solution, minimum end members, and minimum volume respectively. For LBA we use the R package *lba* ([Jelihovschi & Allaman, 2018](#)) and use CWLS with weights being 1. For EMA, we use EMMA ([Seidel & Hlawitschka, 2015](#); [Weltje, 1997](#)). For NMF, we use the objective function  $\{\|X - M\mathbf{H}\|_2^2 + \lambda \text{logdet}(\mathbf{H}\mathbf{H}^T + \delta\mathbf{I})\}$  with constraint  $M\mathbf{1} = \mathbf{1}$  ([Gillis, 2020](#); [Leplat et al., 2019](#)). For NMF, the basis matrix  $\mathbf{H}$  is transformed into  $\mathbf{G} = \mathbf{D}_H^{-1}\mathbf{H}$  such that the sum of each row is 1. As the sum of each row is constant, namely the number of minutes in seven days, the data analysed in LBA, EMA and NMF are identical, and the differences between

the models are due to the algorithms used.

The results for  $K = 1$  is shown in the second column of Table 4. For the whole group, a large part of the time is spent with sleeping (0.358), next to paid work (0.080), then domestic work and tv-radio (both 0.078). On the other hand, one spends little time on recreation outside (0.006) and on relaxing (0.007). These estimates are useful as they provide a point of reference for the interpretation of the estimates of the basis vectors when  $K > 1$ .

Table 3: Time allocation in the Netherlands, cross-classified by Gender, Age and Year (Mooijaart et al., 1999).

Gender	Age	Year	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Male	12-24	1975	901	87	33	120	289	508	3737	1447	128	515	490	419	111	48	752	272	78	146
	12-24	1980	769	157	28	138	294	528	3765	1455	101	505	396	436	102	41	815	256	56	240
	12-24	1985	707	155	15	127	316	527	3744	1537	92	449	441	485	100	64	860	188	73	200
	25-34	1975	2180	250	194	152	293	623	3380	124	129	609	382	269	173	69	700	366	64	124
	25-34	1980	1992	269	206	157	316	649	3403	245	126	649	321	279	213	35	671	318	58	172
	25-34	1985	1899	341	184	183	302	605	3397	208	143	599	391	271	231	67	812	243	57	148
	35-49	1975	1901	249	99	173	351	660	3463	56	195	671	360	206	259	88	785	316	59	188
	35-49	1980	2008	289	128	157	339	709	3445	90	156	593	240	280	238	45	804	343	44	170
	35-49	1985	2093	331	136	185	332	650	3347	85	148	479	336	291	268	64	812	319	58	146
	50-64	1975	1708	244	51	227	350	709	3560	18	122	603	237	209	256	116	921	468	79	203
	50-64	1980	1357	337	54	221	364	744	3569	58	207	704	279	299	288	76	862	413	57	190
	50-64	1985	1206	450	25	230	352	686	3533	46	272	554	264	316	309	112	1012	467	68	174
	> 65	1975	176	617	124	273	365	763	3801	10	159	811	213	297	366	86	1161	477	157	223
	> 65	1980	71	563	27	251	392	767	3871	43	192	671	220	403	312	117	1198	660	92	230
	> 65	1985	95	636	38	264	383	707	3694	54	214	619	274	476	308	178	1233	578	104	225
Female	12-24	1975	723	494	135	208	359	536	3744	1163	125	592	364	348	90	32	594	292	73	208
	12-24	1980	665	460	99	200	377	513	3777	1321	88	557	400	370	76	32	581	257	74	234
	12-24	1985	564	397	86	223	387	495	3821	1436	80	527	396	352	86	41	702	207	63	214
	25-34	1975	439	1342	635	347	311	593	3526	77	85	780	316	306	149	41	547	300	88	199
	25-34	1980	471	1338	673	336	339	607	3532	115	115	776	270	352	131	32	497	275	54	167
	25-34	1985	704	1147	651	336	337	572	3447	120	115	736	303	368	145	44	565	265	63	164
	35-49	1975	299	1567	296	372	325	664	3567	104	133	694	225	335	198	37	622	356	76	207
	35-49	1980	375	1605	309	347	346	633	3554	98	143	689	229	440	154	34	576	311	63	174
	35-49	1985	412	1529	308	373	351	656	3444	68	196	699	277	453	170	45	582	307	59	153
	50-64	1975	151	1600	83	376	367	601	3673	27	195	758	255	323	197	53	710	478	78	154
	50-64	1980	153	1558	84	335	368	613	3701	30	179	810	212	504	190	41	644	390	53	212
	50-64	1985	233	1487	82	352	385	595	3566	40	195	721	268	545	217	76	708	377	64	170
	> 65	1975	11	1319	78	384	372	635	3849	6	108	929	219	297	169	37	888	485	63	230
	> 65	1980	6	1409	154	292	453	665	3713	21	124	796	187	482	191	39	860	404	67	216
	> 65	1985	19	1318	44	320	366	615	3675	23	139	749	202	579	169	52	1076	460	69	204

Following Mooijaart et al. (1999), we illustrate the models for  $K = 3$ . The results are shown in Table 4 where LBA results are in columns 3-5, EMA in 6-8, NMF in 9-11. For reasons of space, labels are abbreviated. For example, M1275 denotes males aged 12-24 surveyed in 1975. Since LBA, EMA, and NMF have similar results, we discuss the estimates of NMF. The first basis vector is denoted by domestic work (0.147 as compared with 0.078 for  $K = 1$ ) and an absence of paid work and education (both 0.000). The second basis vector is characterized by paid work (0.218 vs 0.080 for  $K = 1$ ). The third basis vector is

characterized by education (0.169 vs 0.033 for  $K = 1$ ).

In Table 4 the estimates for gender, age and year sum to 1 for each of the levels, and this allows us to make a ternary plot. Thus all row points fall into a two-dimensional triangle with corner points  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$ . See Figure 4. The mark *average* in the figure is the average contribution of each dimension  $k = 1, 2, 3$ , which is  $(0.488, 0.310, 0.203)$ , denoted by  $z$ .

Table 4: LBA, EMA, NMF results about time budget dataset for dimensionality  $K = 3$ .

Model Coefficient	/	LBA			EMA			NMF		
	$k = 1$	$k = 1$	$k = 2$	$k = 3$	$k = 1$	$k = 2$	$k = 3$	$k = 1$	$k = 2$	$k = 3$
M1275	1.000	0.000	0.172	0.829	0.000	0.156	0.844	0.000	0.163	0.837
M1280	1.000	0.057	0.107	0.836	0.049	0.095	0.856	0.056	0.101	0.843
M1285	1.000	0.049	0.064	0.887	0.038	0.049	0.913	0.048	0.059	0.892
M2575	1.000	0.000	0.990	0.011	0.004	0.975	0.021	0.000	0.954	0.046
M2580	1.000	0.035	0.886	0.079	0.035	0.873	0.091	0.035	0.854	0.111
M2585	1.000	0.081	0.849	0.070	0.078	0.838	0.084	0.081	0.819	0.100
M3575	1.000	0.091	0.892	0.018	0.093	0.874	0.033	0.090	0.859	0.051
M3580	1.000	0.068	0.923	0.008	0.074	0.903	0.023	0.069	0.889	0.042
M3585	1.000	0.050	0.958	0.000	0.049	0.949	0.002	0.049	0.925	0.025
M5075	1.000	0.156	0.817	0.027	0.158	0.795	0.047	0.154	0.787	0.058
M5080	1.000	0.275	0.656	0.069	0.265	0.643	0.092	0.272	0.633	0.094
M5085	1.000	0.345	0.590	0.065	0.328	0.581	0.091	0.341	0.571	0.088
M6575	1.000	0.703	0.150	0.147	0.658	0.159	0.183	0.695	0.151	0.154
M6580	1.000	0.704	0.097	0.199	0.656	0.109	0.235	0.696	0.099	0.205
M6585	1.000	0.711	0.106	0.183	0.669	0.107	0.224	0.702	0.108	0.190
F1275	1.000	0.243	0.120	0.637	0.227	0.114	0.659	0.241	0.117	0.642
F1280	1.000	0.210	0.061	0.730	0.194	0.054	0.751	0.208	0.059	0.733
F1285	1.000	0.191	0.000	0.810	0.183	0.000	0.817	0.189	0.000	0.811
F2575	1.000	0.834	0.172	0.000	0.798	0.180	0.022	0.823	0.177	0.000
F2580	1.000	0.819	0.178	0.003	0.783	0.184	0.033	0.813	0.184	0.003
F2585	1.000	0.695	0.296	0.010	0.666	0.298	0.036	0.689	0.296	0.015
F3575	1.000	0.926	0.083	0.000	0.878	0.096	0.027	0.910	0.090	0.000
F3580	1.000	0.916	0.102	0.000	0.869	0.119	0.013	0.895	0.105	0.000
F3585	1.000	0.891	0.137	0.000	0.867	0.133	0.000	0.863	0.137	0.000
F5075	1.000	0.990	0.025	0.000	0.928	0.048	0.024	0.969	0.031	0.000
F5080	1.000	0.980	0.024	0.000	0.917	0.043	0.040	0.967	0.033	0.000
F5085	1.000	0.930	0.072	0.000	0.881	0.079	0.040	0.919	0.081	0.000
F6575	1.000	0.959	0.000	0.045	0.886	0.023	0.090	0.950	0.005	0.045
F6580	1.000	0.978	0.004	0.034	0.916	0.005	0.079	0.968	0.000	0.032
F6585	1.000	0.945	0.000	0.056	0.885	0.013	0.101	0.934	0.009	0.057
Basis	$k = 1$	$k = 1$	$k = 2$	$k = 3$	$k = 1$	$k = 2$	$k = 3$	$k = 1$	$k = 2$	$k = 3$
paidwork	0.080	0.002	0.212	0.063	0.000	0.215	0.064	0.000	0.218	0.063
dom.work	0.078	0.146	0.017	0.005	0.152	0.016	0.004	0.147	0.017	0.004
caring	0.017	0.024	0.016	0.005	0.025	0.016	0.000	0.024	0.016	0.000
shopping	0.025	0.036	0.016	0.013	0.037	0.016	0.013	0.036	0.017	0.013
per.need	0.035	0.037	0.032	0.033	0.037	0.032	0.033	0.037	0.032	0.033
eating	0.062	0.064	0.067	0.050	0.064	0.067	0.050	0.064	0.068	0.050
sleeping	0.358	0.361	0.336	0.382	0.358	0.336	0.381	0.363	0.335	0.381
educat.	0.033	0.000	0.006	0.171	0.000	0.004	0.167	0.000	0.000	0.169
particip	0.015	0.016	0.016	0.009	0.016	0.016	0.009	0.016	0.016	0.009
soc.cont	0.066	0.077	0.058	0.047	0.078	0.058	0.048	0.077	0.059	0.047
goingout	0.030	0.022	0.032	0.044	0.021	0.032	0.044	0.022	0.032	0.044
sports	0.036	0.043	0.024	0.042	0.042	0.023	0.042	0.043	0.023	0.042
gardening	0.019	0.020	0.025	0.008	0.020	0.025	0.009	0.020	0.026	0.008
outside	0.006	0.006	0.007	0.005	0.006	0.007	0.005	0.006	0.008	0.005
tv-radio	0.078	0.077	0.079	0.077	0.076	0.079	0.078	0.077	0.079	0.078
reading	0.036	0.042	0.034	0.023	0.042	0.034	0.024	0.042	0.034	0.024
relaxing	0.007	0.008	0.006	0.007	0.007	0.006	0.007	0.008	0.006	0.007
other	0.019	0.020	0.016	0.022	0.019	0.016	0.022	0.020	0.015	0.022

In Figure 4 women older than 24 and men older than 65 spent relatively more of their

time on the first dimension (mainly characterized by domestic work). Men aged between 25 and 64 spent relatively more time on the second dimension (mainly characterized by paid work), while respondents aged between 12 and 24 spent relatively more time on the third dimension (characterized by education). We also observe shifts of gender, age and year. For example, the points for males move from dimension 2 to dimension 1 when they get older, i.e. from 25-34, 35-49, 50-64 to  $> 65$ , showing that, the older they become, the more time they spent on the activities of the first dimension, that is dominated by domestic work. Also, girls aged 12-24 are closer to dimension 1 than boys aged 12-24, and last, in later years (i.e. from 1975 to 1985), both girls and boys spent more time on the third dimension, that is dominated by education.

We can make a similar triangle for columns (De Leeuw et al., 1990). Let the diagonal matrix of the average contribution  $z$  be  $D_z$ . The rescaled basis matrix  $G^{\text{res}} = (D_z G) (\text{diag} (1^T D_z G))^{-1}$ . I.e., if we multiply the three basis vectors in Table 4 with the averages  $z = (0.488, 0.310, 0.203)$ , we have estimates  $D_z G$  whose elements add up to 1 with rows being dimensions and columns being activities. If we then scale these estimated probabilities by dividing by the sizes of the activities, then we have for each activity three estimated conditional probabilities adding up to 1 that can be used to make a ternary plot of the activities. See Figure 5.

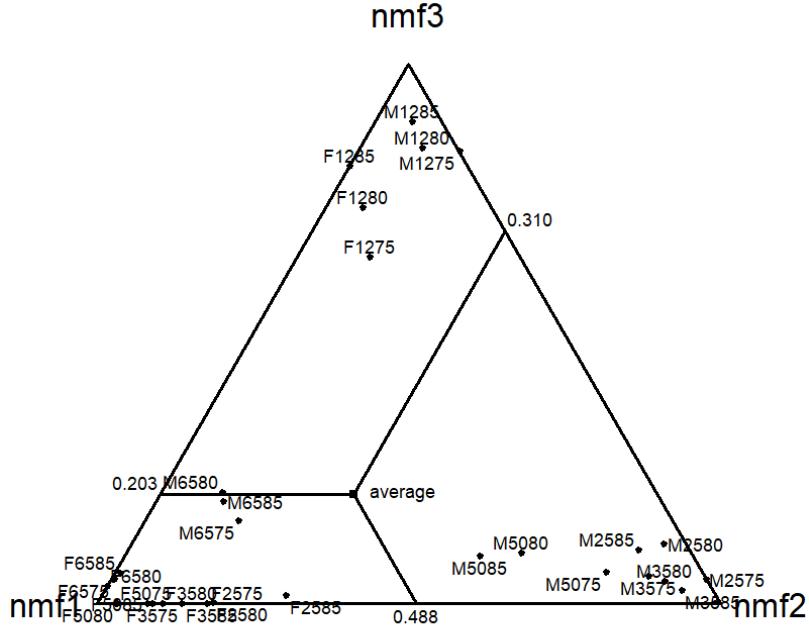
In Figure 5, the first, second, and third basis vectors are dominated by domestic work, paid work, and education, respectively. Domestic work is close to shopping, caring, and so on. Paid work is close to recreation outside, gardening, participant in volunteer work, and so on. Education is close to going out, sports, and so on.

## 7 Related models/methods

### 7.1 Archetypal analysis (AA)

Archetypal analysis (AA) is proposed by Cutler and Breiman (1994). In AA, basis vectors  $\mathbf{h}_1, \dots, \mathbf{h}_K$  are convex combinations of data points  $\mathbf{x}_1, \dots, \mathbf{x}_I$ , where  $\mathbf{h}_k$  is  $k$ th row of basis matrix  $\mathbf{H}$  and  $\mathbf{x}_i$  is  $i$ th row of observed data matrix  $\mathbf{X}$ . Specifically, each basis vector  $\mathbf{h}_k = \sum_{i=1}^I v_{ki} \mathbf{x}_i$ , where  $\sum_i v_{ki} = 1, v_{ki} \geq 0, i = 1, \dots, I$ . In matrix notation, it can be expressed as  $\mathbf{H} = \mathbf{V}\mathbf{X}$ . Furthermore, data points  $\mathbf{x}_1, \dots, \mathbf{x}_I$  are convex combination of

Figure 4: Simplex about time budget dataset for NMF under dimensionality  $K = 3$  about rows.



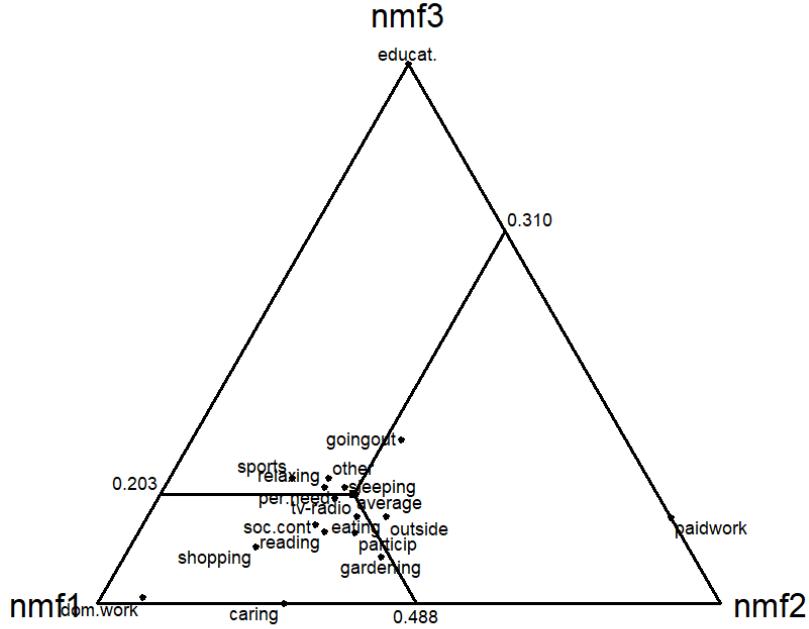
basis vectors  $\mathbf{h}_1, \dots, \mathbf{h}_K$ . Thus, AA can be expressed as follows:

$$\begin{aligned} \mathbf{X} &\approx \mathbf{U}\mathbf{V}\mathbf{X} \\ \text{subject to } \mathbf{U}\mathbf{1} &= \mathbf{1}, \quad \mathbf{V}\mathbf{1} = \mathbf{1} \end{aligned} \tag{14}$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are nonnegative matrices of size  $I \times K$  and  $K \times I$ , respectively. AA is general unique (Behdin & Mazumder, 2024; Cutler & Breiman, 1994; Javadi & Montanari, 2020; Mørup & Hansen, 2012).

When  $\mathbf{V}$  includes a diagonal matrix with positive elements up to permutation, AA can be taken as NMF under the separability assumption. Furthermore, AA in combination of NMF provides some algorithms (Behdin & Mazumder, 2024; De Handschutter, Gillis, Vandaele, & Siebert, 2020; Javadi & Montanari, 2020; Xu et al., 2022).

Figure 5: Simplex about time budget dataset for NMF under dimensionality  $K = 3$  about columns.



## 7.2 Blind hyperspectral unmixing (BHU) and blind sound separation (BSS)

Blind hyperspectral unmixing (BHU) and blind sound separation (BSS) are respectively from remote sensing and signal processing domains. The general form for BHU and BSS is ([Chan et al., 2009](#); [Fu et al., 2015](#); [Gillis, 2020](#)):

$$\begin{aligned} \mathbf{X} &\approx \mathbf{C}\mathbf{L} \\ \text{subject to } \mathbf{C}\mathbf{1} &= \mathbf{1} \end{aligned} \tag{15}$$

where  $\mathbf{C}$  is a nonnegative matrix of size  $I \times K$ ,  $\mathbf{L}$  is a matrix of size  $K \times J$ . Compared with NMF, the main difference is that BHU and BSS have no nonnegative constraint to basis matrix  $\mathbf{L}$  and have sum-to-1 constraint to coefficient matrix  $\mathbf{C}$ . Although BHU and BSS have no nonnegative constraint on the basis matrix  $\mathbf{L}$ , the nonnegative constraint seems to be automatically satisfied when the decomposed matrix  $\mathbf{X}$  is nonnegative.

The connection between BHU, BSS, and NMF are widely noticed. For example, Theorems 8 is from BSS community ([Fu et al., 2015](#)), and it is well-known in NMF ([Gillis, 2020](#)).

### 7.3 Other models

There are other models related to NMF, LBA, EMA and PLSA that we cannot cover in detail, such as latent Dirichlet allocation (Arora et al., 2013; Arora, Ge, & Moitra, 2012; Blei, Ng, & Jordan, 2003; Li, 2024), the hidden Markov model (Huang, Fu, & Sidiropoulos, 2018; Lakshminarayanan & Raich, 2010; Ma et al., 2026), and mixed membership stochastic blockmodel (Fu et al., 2019; Mao, Sarkar, & Chakrabarti, 2017).

## 8 Conclusion

This paper has three goals. First, we illustrate that (1) EMA from sedimentary geology, LBA from social science, asymmetric PLSA from machine learning have the same parametric form; (2) They are equivalent to symmetric PLSA from machine learning and LCA for two-way tables from social science; (3) They are a special case of NMF from machine learning. Second, we study an essential problem: the identifiability. We prove that the solution of NMF is unique if and only if the solution of LBA, LCA, EMA, and PLSA is unique. Existing results regarding uniqueness properties for NMF are worked out in much more detail than for LBA, LCA, EMA, and PLSA. Thus, these results for NMF are also useful for LBA, LCA, EMA, and PLSA. Third, we provide a brief review for algorithms of these models.

The code for this paper is on the GitHub website <https://github.com/qianqianqi28/review>.

## References

- Abdolali, M., Barbarino, G., & Gillis, N. (2024). Dual simplex volume maximization for simplex-structured matrix factorization. *SIAM J. Imaging Sci.*, 17(4), 2362-2391.
- Abdolali, M., & Gillis, N. (2021). Simplex-structured matrix factorization: Sparsity-based identifiability and provably correct algorithms. *SIAM J. Math. Data Sci.*, 3(2), 593-623.
- Andersen Ang, M., & Gillis, N. (2018). Volume regularized non-negative matrix factorizations. In *2018 9th workshop on hyperspectral image and signal processing: Evolution in remote sensing (whispers)* (p. 1-5).

- Ang, A., Hamed, W. B., & De Sterck, H. (2025). Sum-of-norms regularized nonnegative matrix factorization. *Neural Comput.*, 1–28.
- Ang, A. M. S., & Gillis, N. (2019). Algorithms and comparisons of nonnegative matrix factorizations with volume regularization for hyperspectral unmixing. *IEEE J. Sel. Top. Appl.*, 12(12), 4843–4853.
- Araújo, M. C. U., Saldanha, T. C. B., Galvao, R. K. H., Yoneyama, T., Chame, H. C., & Visani, V. (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemom. Intell. Lab. Syst.*, 57, 65–73.
- Arora, S., Ge, R., Halpern, Y., Mimno, D., Moitra, A., Sontag, D., ... Zhu, M. (2013). A practical algorithm for topic modeling with provable guarantees. In S. Dasgupta & D. McAllester (Eds.), *Proceedings of the 30th international conference on machine learning* (Vol. 28, pp. 280–288). Atlanta, Georgia, USA: PMLR.
- Arora, S., Ge, R., & Moitra, A. (2012). Learning topic models – going beyond svd. In *Proceedings of the 2012 ieee 53rd annual symposium on foundations of computer science* (p. 1–10). USA: IEEE Computer Society.
- Avis, D., & Bremner, D. (1995). How good are convex hull algorithms? In *Proceedings of the eleventh annual symposium on computational geometry* (p. 20–28). New York, NY, USA: Association for Computing Machinery.
- Barbarino, G., & Gillis, N. (2025). On the robustness of the successive projection algorithm. *SIAM J. Matrix Anal. Appl.*, 46(3), 2140–2170.
- Behdin, K., & Mazumder, R. (2024). Sparse NMF with archetypal regularization: Computational and robustness properties. *J. Mach. Learn. Res.*, 25(36), 1–62.
- Berman, A., & Plemmons, R. J. (1994). *Nonnegative matrices in the mathematical sciences*. Society for Industrial and Applied Mathematics.
- Bioucas-Dias, J. M. (2009). A variable splitting augmented lagrangian approach to linear spectral unmixing. In *2009 first workshop on hyperspectral image and signal processing: Evolution in remote sensing* (p. 1–4).
- Bittorf, V., Recht, B., Ré, C., & Tropp, J. A. (2012). Factoring nonnegative matrices with linear programs. In *Proceedings of the 26th international conference on neural information processing systems - volume 1* (p. 1214–1222). Red Hook, NY, USA: Curran Associates Inc.

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 3, 993–1022.
- Chan, T.-H., Chi, C.-Y., Huang, Y.-M., & Ma, W.-K. (2009). A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing. *IEEE Trans. Signal Process.*, 57(11), 4418-4432.
- Chen, W., & Guillaume, M. (2012). HALS-based NMF with flexible constraints for hyperspectral unmixing. *EURASIP J. Adv. Sig. Pr.*, 2012(1), 54.
- Clogg, C. C. (1981). Latent structure models of mobility. *Am. J. Sociol.*, 86(4), 836–868.
- Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311–359). Boston, MA: Springer US.
- Craig, M. (1994). Minimum-volume transforms for remotely sensed data. *IEEE T. Geosci. Remote*, 32(3), 542-552.
- Cutler, A., & Breiman, L. (1994). Archetypal analysis. *Technometrics*, 36(4), 338–347.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *J. Am. Soc. Inform. Sci.*, 41(6), 391-407.
- De Handschutter, P., Gillis, N., Vandaele, A., & Siebert, X. (2020). Near-convex archetypal analysis. *IEEE Signal Proc. Let.*, 27, 81-85.
- De Leeuw, J., & Van der Heijden, P. G. M. (1988). The analysis of time-budgets with a latent time-budget model. In *Data analysis and informatics II*. Amsterdam, Netherlands.
- De Leeuw, J., & Van der Heijden, P. G. M. (1991). Reduced rank models for contingency tables. *Biometrika*, 78(1), 229–232.
- De Leeuw, J., Van der Heijden, P. G. M., & Verboon, P. (1990). A latent time-budget model. *Stat. Neerl.*, 44(1), 1-22.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 39(1), 1-22.
- Dietze, E., & Dietze, M. (2019). Grain-size distribution unmixing using the R package emmageo. *E&G Quaternary Science Journal*, 68(1), 29–46.
- Dietze, M., Schulte, P., & Dietze, E. (2022). Application of end-member modelling to grain-size data: Constraints and limitations. *Sedimentology*, 69(2), 845-863.
- Ding, C., Li, T., & Peng, W. (2008). On the equivalence between non-negative matrix

- factorization and probabilistic latent semantic indexing. *Comput. Statist. Data Anal.*, 52(8), 3913-3927.
- Donoho, D., & Stodden, V. (2003). When does non-negative matrix factorization give a correct decomposition into parts? In S. Thrun, L. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems* (Vol. 16). MIT Press.
- Dumais, S. T. (1991). Improving the retrieval of information from external sources. *Behav. Res. Meth. Ins. C.*, 23(2), 229-236.
- Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve access to textual information. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 281–285). New York, NY, USA: Association for Computing Machinery.
- Elisabeth, D., Kai, H., Bernhard, D., Janneke, I., Frank, L., Stephan, O., ... Andreas, B. (2012). An end-member algorithm for deciphering modern detrital processes from lake sediments of lake donggi cona, ne tibetan plateau, china. *Sediment. Geol.*, 243-244, 169-180.
- Evans, M. J., Gilula, Z., & Guttman, I. (1989). Latent class analysis of two-way contingency tables by bayesian methods. *Biometrika*, 76(3), 557-563.
- Ferreira, P. A. d. L., Nunes, L. S., Goya, S. C. y., & Mahiques, M. M. d. (2025). Grain-size signatures reveal the complexity of seafloor processes in the santos bifurcation area (southwestern atlantic margin). *Ocean and Coastal Research*, 73, e25041.
- Fienberg, S. E., Hersh, P., Rinaldo, A., & Zhou, Y. (2009). Maximum likelihood estimation in latent class models for contingency table data. In P. Gibilisco, E. Riccomagno, M. P. Rogantin, & H. P. Wynn (Eds.), *Algebraic and geometric methods in statistics* (p. 27–62). Cambridge University Press.
- Fu, X., Huang, K., & Sidiropoulos, N. D. (2018). On identifiability of nonnegative matrix factorization. *IEEE Signal Proc. Let.*, 25(3), 328-332.
- Fu, X., Huang, K., Sidiropoulos, N. D., & Ma, W.-K. (2019). Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications. *IEEE Signal Proc. Mag.*, 36(2), 59-80.
- Fu, X., Huang, K., Yang, B., Ma, W.-K., & Sidiropoulos, N. D. (2016). Robust volume minimization-based matrix factorization for remote sensing and document cluster-

- ing. *IEEE Trans. Signal Process.*, 64(23), 6254-6268.
- Fu, X., & Ma, W.-K. (2016). Robustness analysis of structured matrix factorization via self-dictionary mixed-norm optimization. *IEEE Signal Proc. Let.*, 23(1), 60-64.
- Fu, X., Ma, W.-K., Huang, K., & Sidiropoulos, N. D. (2015). Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain. *IEEE Trans. Signal Process.*, 63(9), 2306-2320.
- Gaujoux, R., & Seoighe, C. (2010). A flexible r package for nonnegative matrix factorization. *BMC Bioinformatics*, 11(1), 367: 1-9.
- Gaujoux, R., Seoighe, C., & Sauwen, N. (2024). *The package nmf: manual pages*. CRAN. (R package version 0.28)
- Gaussier, E., & Goutte, C. (2005). Relation between PLSA and NMF and implications. In *Proceedings of the 28th annual international acm sigir conference on research and development in information retrieval* (p. 601–602). New York, NY, USA: Association for Computing Machinery.
- Ge, R., & Zou, J. (2015). Intersecting faces: non-negative matrix factorization with new guarantees. In *Proceedings of the 32nd international conference on international conference on machine learning - volume 37* (p. 2295–2303). JMLR.org.
- Gillis, N. (2012). Sparse and unique nonnegative matrix factorization through data preprocessing. *J. Mach. Learn. Res.*, 13(1), 3349–3386.
- Gillis, N. (2013). Robustness analysis of hottopixx, a linear programming model for factoring nonnegative matrices. *SIAM J. Matrix Anal. Appl.*, 34(3), 1189-1212.
- Gillis, N. (2014). Successive nonnegative projection algorithm for robust nonnegative blind source separation. *SIAM J. Imaging Sci.*, 7(2), 1420-1450.
- Gillis, N. (2019). Successive projection algorithm robust to outliers. In *2019 ieee 8th international workshop on computational advances in multi-sensor adaptive processing (camsap)* (p. 331-335).
- Gillis, N. (2020). *Nonnegative matrix factorization*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Gillis, N., & Rajkó, R. (2023). Partial identifiability for nonnegative matrix factorization. *SIAM J. Matrix Anal. Appl.*, 44(1), 27-52.
- Gilula, Z. (1979). Singular value decomposition of probability matrices: Probabilistic as-

- pects of latent dichotomous variables. *Biometrika*, 66(2), 339-344.
- Gilula, Z. (1983). Latent conditional independence in two-way contingency tables: A diagnostic approach. *British Journal of Mathematical and Statistical Psychology*, 36(1), 114-122.
- Goodman, L. A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable. part i-a modified latent structure approach. *Am. J. Sociol.*, 79(5), 1179-1259.
- Goodman, L. A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2), 215-231.
- Goodman, L. A. (1987). New methods for analyzing the intrinsic character of qualitative variables using cross-classified data. *Am. J. Sociol.*, 93(3), 529-583.
- Greenacre, M. (2017). *Correspondence analysis in practice*. New York: Chapman and Hall/CRC.
- Guo, Y.-T., Li, Q.-Q., & Liang, C.-S. (2024). The rise of nonnegative matrix factorization: Algorithms and applications. *Inform. Syst.*, 123, 102379.
- Haberman, S. J. (1979). *Analysis of qualitative data (vol 2)*. New York: Academic Press.
- Hayashi, K., Aksoy, S. G., Ballard, G., & Park, H. (2025). Randomized algorithms for symmetric nonnegative matrix factorization. *SIAM J. Matrix Anal. Appl.*, 46(1), 584-625.
- Hobolth, A., Guo, Q., Kousholt, A., & Jensen, J. L. (2020). A unifying framework and comparison of algorithms for non-negative matrix factorisation. *Int. Stat. Rev.*, 88(1), 29–53.
- Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international acm sigir conference on research and development in information retrieval* (p. 50–57). New York, NY, USA: Association for Computing Machinery.
- Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Mach. Learn.*, 42(1), 177-196.
- Hofmann, T., Puzicha, J., & Jordan, M. I. (1999). Unsupervised learning from dyadic data. In *Advances in neural information processing systems*.
- Huang, K., Fu, X., & Sidiropoulos, N. (2018). Learning hidden Markov models from pairwise co-occurrences with application to topic modeling. In J. Dy & A. Krause

- (Eds.), *Proceedings of the 35th international conference on machine learning* (Vol. 80, pp. 2068–2077). PMLR.
- Huang, K., Sidiropoulos, N. D., & Swami, A. (2014). Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. *IEEE Trans. Signal Process.*, 62(1), 211-224.
- Javadi, H., & Montanari, A. (2020). Nonnegative matrix factorization via archetypal analysis. *J. Amer. Statist. Assoc.*, 115(530), 896–907.
- Jelihovschi, E., & Allaman, I. B. (2018). lba: An r package for latent budget analysis. *The R Journal*, 10, 269-287.
- Jia, S., & Qian, Y. (2009). Constrained nonnegative matrix factorization for hyperspectral unmixing. *IEEE T. Geosci. Remote*, 47(1), 161-173.
- Jones, N., Tong, L., Pagdon, S., Ebuenyi, I. D., Harrow, M., Sharma, R. P., & Rosen, C. (2024). Using latent class analysis to investigate enduring effects of intersectional social disadvantage on long-term vocational and financial outcomes in the 20-year prospective chicago longitudinal study. *Psychol. Med.*, 54(10), 2444–2456.
- Lakshminarayanan, B., & Raich, R. (2010). Non-negative matrix factorization for parameter estimation in hidden markov models. In *2010 ieee international workshop on machine learning for signal processing* (p. 89-94).
- Laurberg, H., Christensen, M., Plumbley, M., Hansen, L., & Jensen, S. (2008). Theorems on positive data: On the uniqueness of nmf. *Comput. Intel. Neurosc.*, 2008, 764206.
- Lazarsfeld, P. F., & Henry, W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 788-791.
- Leplat, V., Ang, A. M., & Gillis, N. (2019). Minimum-volume rank-deficient nonnegative matrix factorizations. In *Icassp 2019 - 2019 ieee international conference on acoustics, speech and signal processing (icassp)* (p. 3402-3406).
- Leplat, V., Gillis, N., & Ang, A. M. (2020). Blind audio source separation with minimum-volume beta-divergence nmf. *IEEE Trans. Signal Process.*, 68, 3400-3410.
- Li, H. (2024). Latent dirichlet allocation. In *Machine learning methods* (pp. 439–471). Singapore: Springer Nature Singapore.
- Li, J., & Bioucas-Dias, J. M. (2008). Minimum volume simplex analysis: A fast algorithm to

- unmix hyperspectral data. In *Igarss 2008 - 2008 ieee international geoscience and remote sensing symposium* (Vol. 3, p. III - 250-III - 253).
- Lin, C.-H., Ma, W.-K., Li, W.-C., Chi, C.-Y., & Ambikapathi, A. (2015). Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case. *IEEE T. Geosci. Remote*, 53(10), 5530-5546.
- Lin, C.-H., Wu, R., Ma, W.-K., Chi, C.-Y., & Wang, Y. (2018). Maximum volume inscribed ellipsoid: A new simplex-structured matrix factorization framework via facet enumeration and convex optimization. *SIAM J. Imaging Sci.*, 11(2), 1651-1679.
- Lin, Z., Wan, Q., Zhang, F., Zhong, J., Zhou, Z., & Bao, K. (2025). Using end-member model algorithm to infer sedimentary processes from mangrove sediment grain-size in guangdong, south china. *Regional Studies in Marine Science*, 83, 104069.
- Ma, Y., Chen, H., Kang, J., Guo, X., Sun, C., Xu, J., ... Zhang, M. (2026). The hidden markov model and its applications in bioinformatics analysis. *Genes & Diseases*, 13(1), 101729.
- Mao, X., Sarkar, P., & Chakrabarti, D. (2017). On mixed memberships and symmetric nonnegative matrix factorizations. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th international conference on machine learning* (Vol. 70, pp. 2324–2333). PMLR.
- Mei, S., & He, M. (2011). Minimum endmember-wise distance constrained nonnegative matrix factorization for spectral mixture analysis of hyperspectral images. In *2011 ieee international geoscience and remote sensing symposium* (p. 1299-1302).
- Miao, L., & Qi, H. (2007). Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE T. Geosci. Remote*, 45(3), 765-777.
- Mooijaart, A., Van der Heijden, P. G., & Van der Ark, L. (1999). A least squares algorithm for a mixture model for compositional data. *Comput. Statist. Data Anal.*, 30(4), 359-379.
- Moussaoui, S., Brie, D., & Idier, J. (2005). Non-negative source separation: range of admissible solutions and conditions for the uniqueness of the solution. In *Proceedings. (icassp '05). ieee international conference on acoustics, speech, and signal processing, 2005*. (Vol. 5, p. v/289-v/292 Vol. 5).
- Mørup, M., & Hansen, L. K. (2012). Archetypal analysis for machine learning and data mining. *Neurocomputing*, 80, 54-63. (Special Issue on Machine Learning for Signal

Processing 2010)

- Nascimento, J., & Dias, J. (2005). Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE T. Geosci. Remote*, 43(4), 898-910.
- Paatero, P., & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2), 111-126.
- Paterson, G. A., & Heslop, D. (2015). New methods for unmixing sediment grain size data. *Geochem. Geophys. Geosy.*, 16(12), 4494-4506.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 12, 2825–2830.
- Peng, J., Wang, X., Zhao, H., & Dong, Z. (2023). Single-sample unmixing and parametric end-member modelling of grain-size distributions with transformed probability density functions and their performance comparison using aeolian sediments. *Sediment. Geol.*, 445, 106328.
- Plevoets, K. (2024). *svs: Tools for semantic vector spaces*. CRAN. (R package version 3.1.1)
- Qi, Q., Hessen, D. J., Deoskar, T., & Van der Heijden, P. G. M. (2024). A comparison of latent semantic analysis and correspondence analysis of document-term matrices. *Nat. Lang. Eng.*, 30(4), 722–752.
- Renner, R. M. (1989). *On the resolution of compositional datasets into convex combinations of extreme vectors* (PhD dissertation). Victoria University of Wellington.
- Renner, R. M. (1993). The resolution of a compositional data set into mixtures of fixed source compositions. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, 42(4), 615-631.
- Renner, R. M. (1995). The construction of extreme compositions. *Math. Geol.*, 27(4), 485-497.
- Renny, A., Kawsar, M., Manoj, M., Bikkina, S., Phartiyal, B., Kurian, P. J., ... Thakur, B. (2026). Decoding the sedimentary responses to the monsoon seasonality and ocean circulation in the southeast arabian sea during the last 50 ka. *Palaeogeogr. Palaeocl.*, 681, 113384.
- Saberi-Movahed, F., Berahmand, K., Sheikhpour, R., Li, Y., Pan, S., & Jalili, M. (2025). Non-negative matrix factorization in dimensionality reduction: A survey. *ACM Comput.*

*Surv.*, 58(5).

- Sandgren, L. (1954). On convex cones. *Math. Scand.*, 2, 19–28.
- Seidel, M., & Hlawitschka, M. (2015). An R-based function for modeling of end member compositions. *Math. Geosci.*, 47(8), 995-1007.
- Septyarskiy, V. B., Soldatov, R. A., Koch, E., McGinty, R. J., Goldmann, J. M., Hernandez, R. D., ... Sunyaev, S. (2021). Population sequencing data reveal a compendium of mutational processes in the human germ line. *Science*, 373(6558), 1030-1035.
- Soldatov, R., Kharchenko, P., Petukhov, V., & Biederstedt, E. (2021). *Volume-regularized structured matrix factorization*. CRAN. (R package version 1.0.2)
- Sun, Y., Babu, P., & Palomar, D. P. (2017). Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Trans. Signal Process.*, 65(3), 794-816.
- Van der Ark, L. A. (1999). *Contributions to latent budget analysis; a tool for the analysis of compositional data* (PhD dissertation). Utrecht University.
- Van der Heijden, P. G. M., Mooijaart, A., & De Leeuw, J. (1989). Latent budget analysis. In A. Decarli, B. Francis, R. Gilchrist, & G. Seeber (Eds.), *Statistical modelling. proceedings of glim 89 and the 4th international workshop on statistical modelling*. (pp. 301–313). Springer.
- Van Hateren, J., Prins, M., & Van Balen, R. (2018). On the genetically meaningful decomposition of grain-size distributions: A comparison of different end-member modelling algorithms. *Sediment. Geol.*, 375, 49-71. (Analysis of sediment properties)
- Van der Ark, L. A., Van der Heijden, P. G. M., & Sikkel, D. (1999). On the identifiability in the latent budget model. *J. Classif.*, 16(1), 117-137.
- Van der Heijden, P. G. M. (1994). End-member analysis and latent budget analysis. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, 43(3), 527-530.
- Van der Heijden, P. G. M., Mooijaart, A., & De Leeuw, J. (1992). Constrained latent budget analysis. *Sociol. Methodol.*, 22, 279–320.
- Vermunt, J., & Magidson, J. (2016). *Technical guide for latent gold 5.1: Basic, advanced, and syntax*. Belmont, MA: Statistical Innovations Inc.
- Weltje, G. J. (1997). End-member modeling of compositional data: Numerical-statistical algorithms for solving the explicit mixing problem. *Math. Geol.*, 29(4), 503-549.

- Weltje, G. J., & Prins, M. A. (2003). Muddled or mixed? inferring palaeoclimate from size distributions of deep-sea clastics. *Sediment. Geol.*, 162(1), 39-62. (Climate Impact on Sedimentary Systems)
- Weltje, G. J., & Prins, M. A. (2007). Genetically meaningful decomposition of grain-size distributions. *Sediment. Geol.*, 202(3), 409-424. (From Particle Size to Sediment Dynamics)
- Xu, M., Yang, Z., Ren, G., Sheng, H., Liu, S., Liu, W., & Ye, C. (2022). L<sub>1</sub> sparsity-constrained archetypal analysis algorithm for hyperspectral unmixing. *IEEE Geosci. Remote S.*, 19, 1-5.
- Yu, S.-Y., Colman, S. M., & Li, L. (2016). Bemma: A hierarchical bayesian end-member modeling analysis of sediment grain-size distributions. *Math. Geosci.*, 48(6), 723-741.
- Yu, Y., & Sun, W. (2007). Minimum distance constrained non-negative matrix factorization for the endmember extraction of hyperspectral images. In Y. Wang, B. Lei, J.-Y. Yang, J. Li, C. Wang, & L.-P. Zhang (Eds.), *MIPPR 2007: Remote sensing and gis data processing and applications; and innovative multispectral technology and applications* (Vol. 6790, p. 679015). SPIE.
- Zhang, X., Wang, H., Xu, S., & Yang, Z. (2020). A basic end-member model algorithm for grain-size data of marine sediments. *Estuar. Coast. Shelf S.*, 236, 106656.
- Zhang, X., Zhou, X., Tan, G., Huang, S., Shan, X., Lin, C., & Zhang, X. (2025). Sea-level and climatic evolution in the postglacial qiantang river incised-valley system: Insights from grain-size end-member modeling and geochemical characteristics. *Palaeogeogr. Palaeocl.*, 661, 112707.
- Zhou, G., Xie, S., Yang, Z., Yang, J.-M., & He, Z. (2011). Minimum-volume-constrained nonnegative matrix factorization: Enhanced ability of learning parts. *IEEE T. Neural Netw.*, 22(10), 1626-1637.
- Zitnik, M., & Zupan, B. (2012). Nimfa: A python library for nonnegative matrix factorization. *J. Mach. Learn. Res.*, 13, 849-853.