

LA-UR-21-30686

Accepted Manuscript

Factorization of Binary Matrices: Rank Relations, Uniqueness and Model Selection of Boolean Decomposition

Alexandrov, Boian
DeSantis, Derek Freeman
Truong, Phan Minh Duc
Skau, Erik West

Provided by the author(s) and the Los Alamos National Laboratory (2024-02-12).

To be published in: ACM Transactions on Knowledge Discovery from Data

DOI to publisher's version: 10.1145/3522594

Permalink to record:

<https://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-21-30686>



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Factorization of Binary Matrices: Rank Relations, Uniqueness and Model Selection of Boolean Decomposition

Derek DeSantis*, Erik Skau†, Duc P. Truong‡, Boian Alexandrov§

November 3, 2021

Abstract

The application of binary matrices are numerous. Representing a matrix as a mixture of a small collection of latent vectors via low-rank decomposition is often seen as an advantageous method to interpret and analyze data. In this work, we examine the factorizations of binary matrices using standard arithmetic (real and nonnegative) and logical operations (Boolean and \mathbb{Z}_2). We examine the relationships between the different ranks, and discuss when factorization is unique. In particular, we characterize when a Boolean factorization $X = W \wedge H$ has a unique W , a unique H (for a fixed W), and when both W and H are unique, given a rank constraint. We introduce a method for robust Boolean model selection, called $\text{BMF}k$, and show on numerical examples that $\text{BMF}k$ not only accurately determines the correct number of Boolean latent features but reconstruct the pre-determined factors accurately.

Keywords— Boolean matrix factorization, nonnegative matrix factorization, \mathbb{Z}_2 matrix factorization, unique factorization, rank, model determination

1 Introduction

The rank of a real valued $N \times M$ matrix $X \in \mathbb{R}^{N,M}$ is the dimensionality of the vector space spanned by its columns. If the rank of X is R , then the matrix X can be written as a product of two matrices $X = WH$ where $W \in \mathbb{R}^{N,R}$ and $H \in \mathbb{R}^{R,M}$. In this fashion, the columns of X can be seen as a mixing of latent features w_1, \dots, w_R , the columns of W , according to their weights, the columns of H . By imposing various constraints, one obtains different factorizations. Perhaps the best-known factorization is the Singular Value Decomposition [35], where the factor matrices are restricted to be orthogonal. When the elements of the matrix X admit specific properties, these properties often suggest appropriate constraints on the factor matrices. For example, when X is nonnegative, it is natural to impose a nonnegative constraint, and the nonnegative rank is defined as the smallest number R such that $X = WH$ for nonnegative matrices $W \in \mathbb{R}_+^{N,R}$ and $H \in \mathbb{R}_+^{R,M}$ [9]. Similarly, binary rank is the smallest number R for which a binary matrix can be decomposed into a product of binary matrices [41]. Examples of applications of binary decompositions include social networks, market-basket data, DNA transcription profiles, and many others [22].

Instead of applying constraints, one can choose to change the underlying arithmetic to arrive at new types of decompositions and ranks. If one changes the arithmetic operations of “plus” and “times” to logical operations of “or” with “and”, this results in Boolean rank and Boolean

*Theoretical Division - CNLS, Los Alamos National Laboratory, USA ddesantis@lanl.gov

†CCS Division, Los Alamos National Laboratory, USA ewskau@lanl.gov

‡CCS Division, Los Alamos National Laboratory, USA dptruong@lanl.gov

§Theoretical Division, Los Alamos National Laboratory, USA boian@lanl.gov

matrix factorization (BMF) [26]. Boolean decompositions solve the tiling problem [13] that determines how to cover the 1's of a binary matrix by a minimum number of subitems, which is equivalent to the NP-hard bi-clique cover problem [10, 30]. One can also choose the logical operations of “xor” with “and”, which results in the Galois field decompositions [16, 40].

In all the cases discussed above, matrix factorization allows one to learn latent factors from a complex data subject to various constraints and relations between the elements of the data. Many applications in machine learning and data mining, e.g., document classification, recommendation systems, community detection, cryptography and others, involve data with binary values $\{0, 1\}$ [27]. As such, matrices consisting of 0's and 1's arise in various domains of applicability, and discovering their latent structure is critical for doing any fundamental analysis.

In this work, we present the mathematical theory allowing for comparisons between 1) real, nonnegative, Boolean, binary and \mathbb{Z}_2 ranks and 2) the uniqueness of the corresponding factorizations in these different contexts. The primary focus is to build a framework on which to extract not only the correct number of hidden features, but the correct ones when uniqueness is present. Unfortunately, exact factorizations aren't always possible and even single bit flips can destroy the uniqueness (see Example 9). Practically this leads one to naturally desire a model selection algorithm which 1) discovers a stable rank such that 2) uniqueness is approximately recovered.

In Section 2, we define the notations and standardize the common definitions used throughout the text. In Section 3, we provide the relationships between the five ranks and examples which demonstrate that no other relationships can exist. In Section 4, we move on to discuss the uniqueness of these types of decompositions, which contains our main results. We begin by redefining a familiar geometric framework for nonnegative factorizations to also include real, \mathbb{Z}_2 , and Boolean factorizations. We remark that while the real and \mathbb{Z}_2 factorizations are always (essentially) unique, nonnegative and Boolean need not to be. Indeed, given a Boolean factorization, $X = W \cdot H$, of a 0 – 1 matrix X , we note that uniqueness can be achieved in the patterns W without unique feature weights, H , and vice versa. Hence, we investigate criterion for which W , H or W and H are unique. For non-negative factorizations, these results either follow naturally from linear algebra or are well known in the literature. However in Boolean, these results are new.

Proposition 4 states that W is unique if and only if there exists a unique cone, an additive set analogous to a subspace, which contains the data X . We define a property called *freeness* analogous to linear independence, and show in Theorem 3 that the columns of X satisfy a freeness property if and only if the feature weights H are unique. Freeness leads naturally to the definition of a free rank for Boolean matrices. We then show that if the free rank is equal to the Boolean rank, that uniqueness is guaranteed.

In Section 5, we introduce a method for robust Boolean model selection - *BMFk*. This is a Boolean analog of the nonnegative matrix model selection algorithm *NMFk* [2, 39]. We compare *BMFk* to *NMFk* and demonstrate that for a matrix X with a unique Boolean factorization and $\text{rk}_{\mathbb{B}}(X) \neq \text{rk}_{\mathbb{R}_+}(X)$, *BMFk* discovers the correct hidden patterns while *NMFk* does not. Using theory developed in Section 4, we construct a set of Boolean matrices with unique Boolean factorizations and show that not only does *BMFk* correctly identify the latent dimension, but it also accurately extract the predetermined latent features. Adding noise to this set, we find that *BMFk* finds features that are highly correlated with the unique “true” features, which is important for a practical Boolean model selection and extraction technique .

2 Definitions

Throughout, we let $\mathbb{B} := \{0, 1\}$. When dealing with a $N \times M$ binary matrix $X \in \mathbb{B}^{M,N}$ one can consider different decompositions $X = WH$ where the pattern matrix W and weight matrix H either belong to different sets, such as the reals \mathbb{R} , the non-negatives \mathbb{R}_+ , or the binary set \mathbb{B} . Or alternatively one may want to consider decompositions employing different algebraic operations in the matrix multiplication of W and H . Three natural "addition" operators are

the addition, $+$, exclusive or \oplus , and logical or \vee :

$+$	0	1
0	0	1
1	1	2

\oplus	0	1
0	0	1
1	1	0

\vee	0	1
0	0	1
1	1	1

These addition operators are typically paired with multiplication \times , logical and \otimes , and logical and \wedge respectively:

\times	0	1
0	0	0
1	0	1

\otimes	0	1
0	0	0
1	0	1

\wedge	0	1
0	0	0
1	0	1

which are all identical on the binary set \mathbb{B} . Consequently, there are several different types of arithmetic structures one can consider on binary vectors and matrices. Throughout, we adopt the following notation:

- The *real numbers* \mathbb{R} is equipped with the the operations $(+, \times)$
- The *nonnegative real numbers* $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ is equipped with the the operations $(+, \times)$
- The *Galois field* $\mathbb{Z}_2 = \mathbb{B}$ is equipped with the the operations (\oplus, \otimes)
- The *Booleans* $\mathcal{B} = \mathbb{B}$ is equipped with the the operations (\vee, \wedge)

In what follows, the notation $\mathbb{R}, \mathbb{R}_+, \mathbb{Z}_2$ and \mathcal{B} will always imply usage of the associated operations. We will also make limited use of the restriction of \mathbb{R}_+ to \mathbb{B} , though this is not closed under addition.

Each of the three pairs of arithmetic operations $(+, \times), (\oplus, \otimes)$ and (\vee, \wedge) naturally defines matrix multiplication. Given two matrices $W \in \mathbb{B}^{N,R}$ and $H \in \mathbb{B}^{R,M}$, we use $WH, W \otimes H,$ and $W \wedge H$ to denote the real, \mathbb{Z}_2 , and Boolean matrix multiplications respectively. When referring to more than one potential operation, we will utilize the notation $W \cdot H$. These different constraints and algebraic operators lead to five natural factorizations of 0 – 1 matrices:

Definition 1. Let $X \in \mathbb{B}^{N,M}$.

- The *real rank* is $\text{rk}_{\mathbb{R}}(X) := \min\{R | X = WH, W \in \mathbb{R}^{N,R}, H \in \mathbb{R}^{R,M}\}$.
- The *nonnegative rank* is $\text{rk}_{\mathbb{R}_+}(X) := \min\{R | X = WH, W \in \mathbb{R}_+^{N,R}, H \in \mathbb{R}_+^{R,M}\}$.
- The *binary rank* is $\text{rk}_{\mathbb{B}}(X) := \min\{R | X = WH, W \in \mathbb{B}^{N,R}, H \in \mathbb{B}^{R,M}\}$.
- The \mathbb{Z}_2 *rank* is $\text{rk}_{\mathbb{Z}_2}(X) := \min\{R | X = W \otimes H, W \in \mathbb{Z}_2^{N,R}, H \in \mathbb{Z}_2^{R,M}\}$.
- The *boolean rank* is $\text{rk}_{\mathcal{B}}(X) := \min\{R | X = W \wedge H, W \in \mathcal{B}^{N,R}, H \in \mathcal{B}^{R,M}\}$.

For each defined rank ¹, a factorization is said to be a *rank factorization* or *rank revealing* if W and H correspond to a minimum R factorization.

Recall that a field is a set wherein addition, subtraction, multiplication and division are well defined, and a semiring is one where only addition and multiplication are well defined. We note that \mathbb{R} and \mathbb{Z}_2 are fields, while \mathbb{R}_+ and \mathcal{B} are semirings which are not fields. Hence, \mathbb{R}^N and \mathbb{Z}_2^N are vector spaces over their respective fields, and \mathbb{R}_+^N and \mathcal{B}^N are semimodules over their respective semirings [15, 33]. Therefore, they are afforded their respective properties, which we will infrequently leverage. Binary factorizations are not a closed arithmetic, and therefore can only be viewed as the restriction of \mathbb{R}_+ to $\{0, 1\}$.

Our goal is to discuss rank relations as well as uniqueness results for 0 – 1 matrices. For historical reasons, the literature has generally compared factorizations against the reals. However these comparisons are sometimes artificial, so we take the more mathematically consistent approach of considering the two cases of field factorizations and semiring factorizations.

¹The definitions of rank given here correspond to what is often referred to as the *factor* or *Schein* rank.

3 Rank relations

In this section, we briefly discuss how the five different ranks for 0 – 1 matrices are related. In some cases, two ranks are incomparable in the sense that one is not guaranteed to be larger than the other. However, there are four guaranteed relations between these five ranks which are summarized in Figure 1:

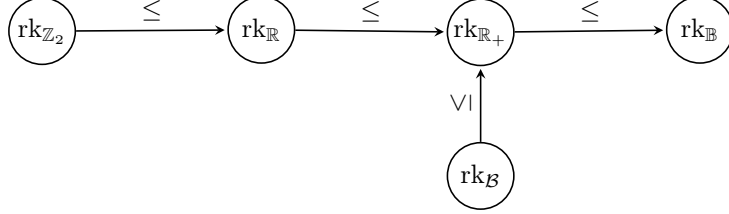


Figure 1: Flowchart of the five rank inequalities.

Since nonnegative factorization is a restriction of real, and binary is a restriction of nonnegative, $\text{rk}_{\mathbb{R}}(X) \leq \text{rk}_{\mathbb{R}_+}(X) \leq \text{rk}_{\mathbb{B}}(X)$ follow by definition. The relationship $\text{rk}_{\mathbb{B}}(X) \leq \text{rk}_{\mathbb{R}_+}(X)$ was previously known, as mentioned in [27], and $\text{rk}_{\mathbb{Z}_2}(X) \leq \text{rk}_{\mathbb{R}}(X)$ appears to be a new result, which is easy to prove by inspecting determinants. Since we could not find a detailed proof of either of these results, we state these results and provide the proofs in the appendix for completeness.

To prove $\text{rk}_{\mathbb{B}}(X) \leq \text{rk}_{\mathbb{R}_+}(X)$ we first define some notations and relations. Given a nonnegative matrix $X = (X_{i,j})$, we can form a binary matrix \hat{X} , via: $\hat{X}_{i,j} = 1$ if $X_{i,j} > 0$, and $\hat{X}_{i,j} = 0$ otherwise. This technique is often referred to as *thresholding* [41]. The following proposition states that thresholding a nonnegative decomposition reveals a Boolean decomposition:

Proposition 1. *Let $X \in \mathbb{B}^{N,M}$. If $X = WH$ is a nonnegative factorization of X , then $\hat{W} \wedge \hat{H}$ is a Boolean factorization of X .*

Proposition 1 states that the process of converting the nonnegative factors W, H of a binary matrix X into Boolean matrices \hat{H}, \hat{W} , via thresholding produces a Boolean factorization. However, we remark that this factorization may not be a minimal rank Boolean factorization. This is the contents of the following theorem:

Theorem 1. *Let $X \in \mathbb{B}^{N,M}$. Then $\text{rk}_{\mathbb{B}}(X) \leq \text{rk}_{\mathbb{R}_+}(X)$.*

Proof. By Proposition 1, each nonnegative factorization yields a Boolean factorization. In particular if $K = \text{rk}_{\mathbb{R}_+}(X)$, and $X = WH$ is a nonnegative rank K factorization of X , then $\hat{W} \wedge \hat{H}$ is a rank K Boolean factorization of X . Thus, $\text{rk}_{\mathbb{B}}(X) \leq \text{rk}_{\mathbb{R}_+}(X)$. \square

Next, we state the connection between the \mathbb{Z}_2 rank and the real rank. Its proof can be found in the Appendix.

Theorem 2. *Let $X \in \mathbb{B}^{N,M}$. Then $\text{rk}_{\mathbb{Z}_2}(X) \leq \text{rk}_{\mathbb{R}}(X)$.*

Given $X \in \mathbb{B}^{N,M}$, two entries $X_{i,j}$ and $X_{k,l}$ are called *isolated ones* if $X_{i,j} = X_{k,l} = 1$ and $X_{i,l}X_{k,j} = 0$ [4]. For example, if

$$X = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

then $\{X_{1,1}, X_{2,2}, X_{3,3}\}$ are isolated ones. The *isolation number* $\iota(X)$ is the maximum size of all sets of isolated ones [4]. The following useful proposition provides a lower bound on some the ranks of a matrix in terms of the isolation number.

Proposition 2 ([4]). *Let $X \in \mathbb{B}^{N,M}$. Then*

$$\iota(X) \leq \text{rk}_{\mathbb{B}}(X) \leq \text{rk}_{\mathbb{R}_+}(X) \leq \text{rk}_{\mathbb{B}}(X).$$

3.1 Rank examples

We will show that the inequalities in Figure 1 are strict, and that no other possible rank comparisons hold. For this, we will utilize the following series of examples:

Example 1. Let

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

It is easy to see that $\text{rk}_{\mathbb{Z}_2}(A) = 2$ and $\text{rk}_{\mathbb{R}}(A) = 3$. On the other hand, A has three isolated ones, $\{A_{1,2}, A_{2,3}, A_{3,1}\}$, so by Proposition 2 the Boolean rank is 3.

Example 2. Let

$$B = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Then each column of B are \mathbb{Z}_2 independent. Moreover, since none of the columns sum to another, we have that $\text{rk}_{\mathbb{Z}_2}(B) = 3$. However, $(0, 1, 1) \vee (1, 0, 1) = (1, 1, 1)$. Thus, $\text{rk}_{\mathcal{B}}(B) \leq 2$. Since $B_{1,2}$ and $B_{2,1}$ are isolated ones, we have that $\text{rk}_{\mathcal{B}}(B) = 2$. Furthermore, $\text{rk}_{\mathbb{R}_+}(B) = \text{rk}_{\mathbb{R}}(B) = 3$.

Example 3 ([9]). Let

$$C = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Then the entries $\{C_{2,1}, C_{1,2}, C_{3,4}, C_{4,3}\}$ are isolated ones. Hence by Proposition 2 $\text{rk}_{\mathcal{B}}(C) = 4$, so that $\text{rk}_{\mathbb{R}_+}(C) = 4$, while, $\text{rk}_{\mathbb{R}}(C) = 3$.

Example 4. Let

$$D = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

then $4 = \text{rk}_{\mathbb{R}_+}(D) < \text{rk}_{\mathbb{B}}(D) = 5$. To see this, note that D contains the matrix C from Example 3 as a submatrix in the bottom left corner. Thus, $\text{rk}_{\mathbb{R}_+}(D) \geq 4$. Furthermore, one has the decomposition

$$D = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

Thus $\text{rk}_{\mathbb{R}_+}(D) \leq 4$. One can also observe that $\text{rk}_{\mathbb{R}}(D) = 4$, so that Theorem 6 of [14] may be applied to show the above decomposition is unique. Since binary factorizations are a restriction of nonnegative, one concludes that $\text{rk}_{\mathbb{B}}(D) > 4$, and therefore $\text{rk}_{\mathbb{B}}(D) = 5$.

Table 1 summarizes the relevant information in Examples 1-4. Example matrices A , B , C and D show that the inequalities in Figure 1 are strict. Moreover, example matrices A and B show that $\text{rk}_{\mathbb{Z}_2}$ and $\text{rk}_{\mathcal{B}}$ are not comparable, while example matrices B and C show that $\text{rk}_{\mathbb{R}}$ and $\text{rk}_{\mathcal{B}}$ are not comparable.

4 Uniqueness of rank factorizations and geometry

Let $X \in \mathbb{B}^{N,M}$, and $X = W \cdot H$ be any of the five rank factorizations. If P is an $R \times R$ permutation matrix, then $W' = W \cdot P$ and $H' = P^{-1} \cdot H$ is another rank factorization for X of the same type. More generally, one could use any invertible matrix whose inverse is within the algebra of interest:

Example	$\text{rk}_{\mathbb{R}}$	$\text{rk}_{\mathbb{Z}_2}$	$\text{rk}_{\mathbb{R}_+}$	$\text{rk}_{\mathcal{B}}$	$\text{rk}_{\mathbb{B}}$
A	3	2	-	3	-
B	3	3	3	2	-
C	3	-	4	4	-
D	-	-	4	-	5

Table 1: Ranks from Examples 1-4.

Definition 2. Given a semiring \mathcal{S} (which may be a field), a matrix $U \in \mathcal{S}^{R,R}$ is said to be *invertible* if there exists another matrix $U^{-1} \in \mathcal{S}^{R,R}$ such that $UU^{-1} = U^{-1}U = I$, the identity matrix.

If $X \in \mathbb{B}^{N,M}$, and $X = W \cdot H$ is a rank factorization, and U is invertible over the respective algebra, then $W' = W \cdot U$ and $H' = U^{-1} \cdot H$ is another rank factorization for X of the same type. We note that if $U \in \mathbb{R}_+^{R,R}$ is an invertible matrix such that $U^{-1} \in \mathbb{R}_+^{R,R}$, then U must be of the form DP for some positive diagonal matrix D and permutation matrix P [6]. Similarly, if $U \in \mathcal{B}^{R,R}$ is invertible, then it must be the case that $U = P$ for some permutation matrix [20]. In either case, multiplication by an invertible matrix encodes a trivial change. For this reason, we make the following definition for uniqueness:

Definition 3. Let $X \in \mathbb{B}^{N,M}$. A factorization $X = W \cdot H$ is said to be *unique* if given another factorization $X = W' \cdot H'$, one has that $W' = W \cdot U$ and $H' = U^{-1} \cdot H$ for some matrix U that is invertible over the respective algebra.

It is easy to see that rank factorizations over fields are always unique. Recall that given a field \mathbb{F} and a positive integer N , the space \mathbb{F}^N defines a vector space with scalars drawn from \mathbb{F} . Any linear map $\mathbb{F}^N \rightarrow \mathbb{F}^M$ defines an $N \times M$ matrix with entries from \mathbb{F} , and vice versa. The appropriate geometric notion to describe rank factorizations of matrices over a field is a subspace. Considering the matrix X as a linear map from $\mathbb{F}^N \rightarrow \mathbb{F}^M$, the rank $\text{rk}_{\mathbb{F}}(X)$ is equal to the maximal number of \mathbb{F} -linearly independent vectors in the range of X , denoted $\text{rng}(X)$. Therefore if $R = \text{rk}_{\mathbb{F}}(X)$, each factorization $X = W \cdot H$ for $W \in \mathbb{F}^{N,R}$, $H \in \mathbb{F}^{R,M}$ is in 1-1 correspondence with a choice w_1, \dots, w_R of basis for $\text{rng}(X)$. Hence, any two rank factorizations differ by a change of basis, which is encoded in an invertible matrix.

The story for semiring factorizations is similar, but comes with some caveats. We will show that finding a rank factorization in these cases is equivalent to finding a minimal type of generating set of a *cone* rather than a subspace. While there is a unique minimal subspace that contains the data in the field case (namely the range of the matrix), there may not be a unique minimal cone that contains the data in the semiring cases. We begin with the definition of a cone:

Definition 4. Let \mathcal{S} be a semiring ($\mathcal{S} = (\mathbb{R}_+, +, \times)$ or $(\mathcal{B}, \vee, \wedge)$) and $N \in \mathbb{N}$. A *cone* is a subset $\mathcal{C} \subset \mathcal{S}^N$ that is closed under addition from \mathcal{S}^N and scalar multiplication from \mathcal{S} ². Given a collection $\{W_i\}_{i=1}^R \subset \mathcal{S}^N$, we define the *span* of $\{W_i\}_{i=1}^R$ as

$$\text{span}\{W_i\}_{i=1}^R := \left\{ \sum_{i=1}^R h_i \cdot W_i : W_i \in \mathcal{S}^N, h_i \in \mathcal{S} \right\}$$

where scalar multiplication and addition utilize the element-wise operations from \mathcal{S} . Given a matrix $W \in \mathcal{S}^{N,R}$, we define the *cone of the matrix* W to be

$$\text{cone}(W) = \{W \cdot h : h \in \mathcal{S}^R\} \subset \mathcal{S}^N.$$

A cone \mathcal{C} is said to have a *generating set* $G \subset \mathcal{C} \subset \mathcal{S}^N$ if $\text{span}(G) = \mathcal{C}$. The *order* of the cone $\mathcal{C} \subset \mathcal{S}^N$, denoted $\mathcal{O}(\mathcal{C})$ is the size of a minimal generating set. A cone is *finitely generated* if $\mathcal{O}(\mathcal{C}) < \infty$.

²Those familiar with semi-ring theory might note that the definition of a cone is the same as an \mathcal{S} sub-semimodule over the semiring \mathcal{S}^N . We have chosen to use the term cone rather than sub-semimodule due to its ubiquity in nonnegative matrix factorizations.

Every finitely generated cone $\mathcal{C} \subset \mathcal{S}^N$ is easily seen to be the span of some collection of vectors and vice versa. Furthermore, $\text{cone}(W)$ is a finitely generated cone, and every finitely generated cone is $\text{cone}(W)$ for some matrix W . We will only be interested in finitely generated cones, so we drop the text *finitely generated* in what follows. We also remark that for $\mathcal{S} = (\mathbb{R}_+, +, \times)$, a cone \mathcal{C} is equivalently the intersection of half spaces, though we don't use this in our analysis.

A cone $\mathcal{C} \subset \mathcal{S}^N$ may have many different generating sets. For example, consider the order 2 cone $\mathcal{C} \subset \mathcal{B}^2$ given by

$$\mathcal{C} = \text{cone} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} = \text{cone} \left(\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \right).$$

Any compact convex subset of \mathbb{R}_+^N is equal to the convex hull of the extreme rays by the Krien-Milman theorem. Consequently a minimal generating set for a cone $\mathcal{C} \subset \mathbb{R}_+^N$ can be uniquely described up to positive scalar multiplication by the extreme rays. It turns out that for any cone \mathcal{C} over \mathcal{B} , the minimal generating set is also unique. Thus in our two semiring cases of interest, \mathbb{R}_+ and \mathcal{B} , we can unambiguously talk about *the* minimal generating set for the cone \mathcal{C} . See the appendix for details.

As mentioned above, in fields each factorization $X = W \cdot H$ for $W \in \mathbb{F}^{N,R}$, $H \in \mathbb{F}^{R,M}$ is in 1-1 correspondence with a basis for $\text{rng}(X)$. The same is true for matrix factorizations over a semirings \mathbb{R}_+ and \mathcal{B} . The following proposition is a consequence of the definition:

Proposition 3. *Let $X \in \mathcal{S}^{N,M}$ for $\mathcal{S} = (\mathbb{R}_+, +, \times)$ or $(\mathcal{B}, \vee, \wedge)$. Then X has a factorization $X = W \cdot H$ for some $W \in \mathcal{S}^{N,R}$ and $H \in \mathcal{S}^{R,M}$ if and only if $\text{cone}(X) \subset \text{cone}(W)$. Moreover,*

$$\text{rk}_{\mathcal{S}}(X) = \min\{\mathcal{O}(\mathcal{C}) : \mathcal{C} \subset \mathcal{S}^N \text{ is a cone with } \text{cone}(X) \subset \mathcal{C}\}.$$

Proposition 3 states that finding a rank factorization $X = W \cdot H$ is equivalent to finding a cone \mathcal{C} of minimal order that contains the data $\text{cone}(X)$. In general, there can be more than one such minimal cone \mathcal{C} , as the following example demonstrates:

Example 5. Consider nonnegative factorizations of the matrix

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

It is well known that X has two distinct nonnegative rank 4 NMF factorizations: $X = IX$ and $X = XI$, where I is the 4×4 identity. As such, there are two separate sets of extreme rays that can be used to create the data X . From Example 3 we know that $\text{rk}_{\mathcal{B}}(X) = 4$, so this fails to have a unique Boolean rank factorization as well.

Another issue that can arise in the semi-ring cases is non-uniqueness of the feature weights, H :

Example 6. Consider Boolean decompositions of the Boolean matrix

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}.$$

If x_i is the i 'th column of X , then $W = [x_1 \ x_2 \ x_3]$ is the unique minimal generating set for $\text{cone}(X)$. However, since $x_1 \vee x_3 = x_2 \vee x_3 = x_4$, we see that the weights H cannot be unique.

While field factorizations are always unique, the above examples demonstrate that semiring factorizations can fail to be unique because either W , H or both W and H can be not unique. We now investigate criterion that guarantee uniqueness of one of the factors, or both of the factors.

4.1 Uniqueness of the patterns W

We begin by discussing how to achieve uniqueness of the factor W . From Proposition 3, one immediately achieves the following uniqueness statement for the patterns, W in terms of geometry of the cones:

Proposition 4. *Let $X \in \mathcal{S}^{N,M}$ have $\text{rk}_{\mathcal{S}}(X) = R$ for $\mathcal{S} = \mathbb{R}_+$ or \mathcal{B} . Then $X = W \cdot H$ for $W \in \mathcal{S}^{N,R}$, $H \in \mathcal{S}^{R,M}$ has a unique W if and only if there exists a unique order R cone $\mathcal{C} \subset \mathcal{S}^N$ such that $\text{cone}(X) \subset \mathcal{C}$.*

While Proposition 4 characterizes the uniqueness of cones, it isn't always the most practical to implement. Given a W , it would be nice to know if it is indeed unique. In [19], the authors defined a condition called boundary close which was a necessary condition for uniqueness of W in NMF. With very mild changes, these conditions also hold in Boolean case.

Definition 5. A subset $T \subset \mathbb{B}^R$ is *boundary close* if for each $i, j \in \{1, \dots, R\}$ with $i \neq j$, there exists a $t \in T$ such that $t_i = 1$ and $t_j = 0$.

Notice that the row vectors for W in Example 6 are boundary close. This happens to always be the case when there is a unique minimal cone containing the data X :

Proposition 5. *Let $X = W \cdot H$ be a nonnegative or Boolean factorization. Then if $\text{cone}(W)$ is the unique order R cone which contains $\text{cone}(X)$, then the set of row vectors for W are boundary close.*

The proof strategy for NMF implemented in [19] holds for Boolean with very mild changes, so we will not reproduce it here. While boundary close is necessary for uniqueness, it is easily seen to not be sufficient, e.g., see Example 5.

4.2 Uniqueness of the feature weights H

Suppose that $\text{cone}(X) \subset \text{cone}(W)$ for some matrix W . Then by Proposition 3, $X = W \cdot H$ for some feature weights H . This matrix may not be unique even if W is. We now discuss how to achieve uniqueness of the weights matrix H given a fixed pattern matrix W . We will show that when each column of X satisfies a particular property in terms of W , then H must be unique. For this we investigate NMF and Boolean separately, as the conditions are stated slightly different.

4.2.1 NMF

As $\mathbb{R}_+^N \subset \mathbb{R}^N$, one can simply utilize linear independence to codify the uniqueness of a weight matrix H . Recall that the solution space of the real linear system $x = Wh$ is given by $h = W^\dagger x + q$, where W^\dagger is the Moore-Penrose inverse of W and q is any element in the null space of W . The set $\{W^\dagger x + q : q \in \text{null}(W)\}$ intersects the nonnegative orthant at a singleton if and only if $x = Wh$ has a unique nonnegative solution. This is summarized in the following proposition:

Proposition 6. *Suppose $X \in \mathbb{R}_+^{N,M}$, $W \in \mathbb{R}_+^{N,R}$ and $\text{cone}(X) \subset \text{cone}(W)$. Then $X = WH$ has a unique $H \in \mathbb{R}_+^{R,M}$ for the fixed W if and only if $\{W^\dagger x_i + q : q \in \text{null}(W)\} \cap \mathbb{R}_+^R$ is a singleton for each column x_i .*

If $\text{cone}(X) \subset \text{cone}(W)$ and W has full rank, then the $\text{null}(W) = 0$. Thus, the solution space is a singleton which is nonnegative by assumption. We record this in the following corollary:

Corollary 1. *Suppose $X \in \mathbb{R}_+^{N,M}$, $W \in \mathbb{R}_+^{N,R}$ and $\text{cone}(X) \subset \text{cone}(W)$. If $\text{rk}_{\mathbb{R}}(W) = R$, then $X = WH$ has a unique H for the fixed W .*

The converse of Corollary 1 does not hold. Indeed, if X is as in Example 5 with $W = X$, then $H = I$ is unique. However, $\text{rk}_{\mathbb{R}}(W) = 3 < \text{rk}_{\mathbb{R}_+}(W) = 4$. We remark that in the case when $\text{rk}_{\mathbb{R}}(X) = \text{rk}_{\mathbb{R}_+}(X)$, uniqueness of W guarantees uniqueness of H . This is the contents of Theorem 4 below.

4.2.2 Boolean

\mathcal{B}^N does not naturally reside inside a field \mathbb{F}^N in the same way that \mathbb{R}_+ sits inside of \mathbb{R} ³. Consequently, concepts such as linear independence no longer hold. While there is a concept of *independence* for a generating set of a semiring, we will require a slightly stronger criterion to investigate the uniqueness.

Definition 6. A collection of Boolean vectors $\{x_i\}_{i=1}^K \subset \mathcal{B}^N$ is said to be *free* if given any two non-empty subsets $I_1, I_2 \subset \{1, \dots, K\}$ with $I_1 \neq I_2$, we have that

$$\bigvee_{I_1} x_i \neq \bigvee_{I_2} x_i.$$

We note that freeness is stronger than (linear) independence. Indeed, freeness is a statement about the uniqueness of mixing. Vectors $\{x_i\}_{i=1}^K \subset \mathcal{B}^N$ are free if and only if each element in $\text{cone}(\{x_i\})$ can be expressed as a linear combination in at most one way.

The next result states the exact criterion for unique H in a Boolean factorization. For that result, we require the following notation.

Definition 7. For $x, y \in \mathcal{B}^N$, we say that x is *dominated by* y , written $x \leq y$, if for each $x_i = 1$ we have that $y_i = 1$. We say x is *strictly dominated by* y , written $x < y$ if $x \leq y$ and $x \neq y$. Given a vector $x \in \mathcal{B}^N$, and a Boolean matrix $W \in \mathcal{B}^{N,R}$, we set

$$P(x, W) := \{w_i : w_i \leq x\}.$$

If $X = W \wedge H$, and x is a column of X , then $P(x, W)$ are the vectors in W that are dominated by x . That is, the vectors that when added *may* yield x .

Theorem 3. Suppose $X \in \mathcal{B}^{N,M}$, $W \in \mathcal{B}^{N,R}$ and $\text{cone}(X) \subset \text{cone}(W)$. Then $X = W \wedge H$ has a unique H for the fixed W if and only if the sets $P(x_i, W)$ are free for each column x_i .

Proof. Suppose that each $P(x_i, W)$ are free. Then any combination of the columns of W from $P(x_i, W)$ yields a unique representation for each i . In particular since $X = W \wedge H$, for each $i = 1, \dots, M$,

$$x_i = \bigvee_{w \in P(x_i, W)} w$$

is the unique mixing for x_i .

Now suppose that $X = W \wedge H$ has a unique H for W . Suppose to the contrary that there exists some i such that $P(x_i, W)$ is not free. So, there exists $I_1, I_2 \subset P(x_i, W)$ with $I_1 \neq I_2$ non-empty such that

$$\bigvee_{w \in I_1} w = \bigvee_{w \in I_2} w.$$

This yields a contradiction depending on one of two possible cases.

Suppose that I_1 is a proper subset of I_2 . Then since $\bigvee_{w \in I_1} w = \bigvee_{w \in I_2} w$,

$$\begin{aligned} x_i &= \bigvee_{w \in P(x_i, W)} w \\ &= \bigvee_{w \in P(x_i, W) \setminus I_2} w \bigvee_{w \in I_2} w \\ &= \bigvee_{w \in P(x_i, W) \setminus I_2} w \bigvee_{w \in I_1} w. \end{aligned}$$

Since I_1 is a proper subset of I_2 , this shows that x_i has two different decompositions, contradicting uniqueness. Similarly, if I_1 is not a subset of I_2 , then

$$x_i = \bigvee_{w \in P(x_i, W)} w = \bigvee_{w \in P(x_i, W) \setminus I_1} w \bigvee_{w \in I_2} w$$

is two different decompositions of X_i , also contradicting uniqueness. \square

³If it did, then since addition over a field forms an invertible group, each vector must have an additive inverse. But each Boolean vector is idempotent, and the only invertible idempotent in a group zero element.

Corollary 2. *Suppose $X \in \mathcal{B}^{N,M}$, $W \in \mathcal{B}^{N,R}$ and $\text{cone}(X) \subset \text{cone}(W)$. If the columns of W are free, then $X = W \wedge H$ has a unique H for the fixed W .*

Proof. If the columns of W are free, then $P(X_i, W)$ are free for each i . Indeed, given two non-empty subsets $I_1, I_2 \subset P(x_i, W)$ with $I_1 \neq I_2$, we have that $I_1, I_2 \subset \{w_j\}$. Hence, $\vee_{I_1} w_j \neq \vee_{I_2} w_j$. \square

As in the nonnegative case, the converse of Corollary 2 does not hold, as the following example illustrates:

Example 7. Let

$$X = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

and let $W = X$. Then the unique H such that $X = W \wedge H$ is $H = I$, however the columns of W are not free since $x_1 \vee x_2 = x_2 \vee x_3$.

4.3 Uniqueness of the decomposition

It follows from the preceding that in NMF (Boolean factorization), if both Propositions 4 and 6 (Propositions 4 and Theorem 3) hold, then both W and H are unique. Hence, the corresponding decomposition is unique. However, under an additional rank constraint, the more complicated unique mixing criterion can be dropped. This is what we will explore in this subsection.

4.3.1 Uniqueness of NMF

If $\text{rk}_{\mathbb{R}}(X) < \text{rk}_{\mathbb{R}_+}(X)$, then NMF is almost never unique. This occurs because any NMF $X = WH$ solution necessarily utilizes more equations than is needed to span the data X . Consequently, almost every point in the interior of $\text{cone}(X)$ cannot have a unique representation in terms of the extreme rays [14]. On the other hand, if $\text{rk}_{\mathbb{R}_+}(X) = \text{rk}(X) = R$ then the subspace spanned by the columns of X has R linearly independent columns. This forces columns of W to be linearly independent, so that the feature weights must be unique. In [36], Tam proves the following:

Theorem 4 ([36] - Theorem 4.1). *Let $X \in \mathbb{R}_+^{N,M}$ with $\text{rk}_{\mathbb{R}}(X) = \text{rk}_{\mathbb{R}_+}(X) = R$. Then the nonnegative rank factorization $X = WH$, $W \in \mathbb{R}_+^{N,R}$, $H \in \mathbb{R}_+^{R,M}$ is unique if and only if there exists a unique simplicial cone \mathcal{C} such that $\text{cone}(X) \subset \mathcal{C} \subset \text{range}(X) \cap \mathbb{R}_+^N$.*

4.3.2 Uniqueness of Boolean factorization

In NMF, the rank statement $\text{rk}_{\mathbb{R}}(X) = \text{rk}_{\mathbb{R}_+}(X)$ encodes a linear independence of the columns of W . As commented on above, the notion of linear independence does not hold in \mathcal{B}^N . Hence, the concept of rank is does not have an exact translation. However, we have a substitution in \mathcal{B}^N for linear independence, namely freeness. This stronger notion of linear independence naturally leads to a definition of type of column rank for Boolean matrices.

Definition 8. Given a matrix $X \in \mathcal{B}^{N,M}$, the *free (Boolean) rank* of X denoted $\text{frk}_{\mathcal{B}}(X)$ is the size of the largest free Boolean subset of $\text{cone}(X) \subset \mathcal{B}^N$.

Example 8. Consider once again the matrix

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

From Example 3, we know that $\text{rk}_{\mathcal{B}}(X) = 4$. However, the largest size of a set of columns one can find that are free is two, so $\text{frk}_{\mathcal{B}}(X) = 2$. We once again also note that $X = IX = XI$, so that X fails to have a unique factorization.

The column rank of a real matrix X is the size of a largest linearly independent subset of $\text{span}(X)$. Similarly, the row rank of X is the size of a largest linearly independent subset of $\text{span}(X^T)$. It is well known from elementary linear algebra that the column rank and row rank for real valued matrices are equal to the real rank. While we have defined the free rank of X to correspond to the column rank, we could equally have defined the free rank for the rows of X . The next useful result establishes the equality of the free row and column definitions to injectivity:

Proposition 7. *Let $X \in \mathcal{B}^{N,M}$. Then the following are equivalent:*

1. $\text{frk}_{\mathcal{B}}(X) = R$,
2. The size of the largest subset $S \subset \mathcal{B}^M$ on which X is injective is R ,
3. X contains a $R \times R$ permutation submatrix.

Proof. First we recall that X is injective on a set $S \subset \mathcal{B}^M$ if and only if for any pair $z_1, z_2 \in S$, we have

$$\bigvee_{i=1}^M z_{i,1} X_i = X \wedge z_1 \neq X \wedge z_2 = \bigvee_{i=1}^M z_{i,2} X_i.$$

Note the correspondence between $z \in S$ and elements in $\text{cone}(X)$ given by the multiplication $X \wedge z$. We then see that injectivity on S is equivalent to $\{X \wedge z : z \in S\}$ is free. Hence, (1) and (2) are equivalent. The equivalence of (2) and (3) follows from Lemma 7.2.6 in [12]. \square

Corollary 3. *Given a matrix $X \in \mathcal{B}^{N,M}$, $\text{frk}_{\mathcal{B}}(X) \leq \text{rk}_{\mathcal{B}}(X)$.*

Proof. By Proposition 7, we have that $\text{frk}_{\mathcal{B}}(X) = R$ if and only if X contains a $R \times R$ permutation matrix. If X has such a $R \times R$ submatrix, then $\text{rk}_{\mathcal{B}}(X) \geq R$ so that $\text{rk}_{\mathcal{B}}(X) \geq R = \text{frk}_{\mathcal{B}}(X)$. \square

Proposition 7 accomplishes several things. It connects free rank to the familiar concept of injectivity of a matrix on a set. For Boolean matrices, the largest sets for injectivity correspond to columns that contain the largest permutation matrices, which in turn, bound the free ranks by the Boolean rank. In the case of nonnegativity, we noted that when the matrix failed to be injective (full real rank) one was unlikely to achieve uniqueness of the factorization. The free column rank will play a role similar to real rank in NMF in Theorem 4. When the free rank is as large as possible, one achieves a unique Boolean matrix factorization:

Theorem 5. *Let $X \in \mathcal{B}^{N,M}$ and $R = \text{rk}_{\mathcal{B}}(X)$. Then $\text{frk}_{\mathcal{B}}(X) = R$ if and only if every rank R factorization $X = WH$ is permutation equivalent to*

$$W = \begin{pmatrix} I \\ W' \end{pmatrix} \quad H = (I \quad H'),$$

where I is the $R \times R$ identity matrix and $H' \in \mathcal{B}^{R, M-R}$ and $W' \in \mathcal{B}^{N-R, R}$ are some fixed matrices. In particular, $X = WH$ has a unique factorization.

Proof. Suppose that every rank R factorization $X = WH$ is permutation equivalent to

$$W = \begin{pmatrix} I \\ W' \end{pmatrix} \quad H = (I \quad H'),$$

where I is the $R \times R$ identity matrix, $H' \in \mathcal{B}^{R, M-R}$ and $W' \in \mathcal{B}^{N-R, R}$. Multiplying such W and H , we see that X is permutation equivalent to

$$WH = \begin{pmatrix} I & H' \\ W' & W'H' \end{pmatrix}.$$

In particular, X contains an $R \times R$ permutation submatrix. By Proposition 7, the free rank $\text{frk}_{\mathcal{B}}(X) = R$.

Conversely, by Proposition 7, $\text{frk}_{\mathcal{B}}(X) = R$ if and only if X similar to

$$X' = \begin{pmatrix} I & X_{1,2} \\ X_{2,1} & X_{2,2} \end{pmatrix}.$$

where I is the $R \times R$ identity matrix, $X_{1,2} \in \mathcal{B}^{R, M-R}$, $X_{2,1} \in \mathcal{B}^{N-R, R}$, and $X_{2,2} \in \mathcal{B}^{N-R, M-R}$. Since $\text{rk}_{\mathcal{B}}(X)$ is also equal to R , we can write $X' = WH$ with

$$W = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \quad H = (H_1 \quad H_2),$$

and $W_1, H_1 \in \mathcal{B}^{R, R}$, $W_2 \in \mathcal{B}^{N-R, R}$, and $H_2 \in \mathcal{B}^{R, M-R}$. Thus we have

$$\begin{pmatrix} I & X_{1,2} \\ X_{2,1} & X_{2,2} \end{pmatrix} = X' = WH = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} (H_1 \quad H_2) = \begin{pmatrix} W_1 H_1 & W_1 H_2 \\ W_2 H_1 & W_2 H_2 \end{pmatrix}.$$

Equating blocks, we see $I = W_1 H_1$. A Boolean matrix is left invertible if and only if it is right invertible if and only if it is invertible [31]. However, every invertible Boolean matrix must be a permutation matrix [20]. Therefore, $W_1 = P$ and $H_1 = P^{-1}$ for some permutation matrix. It then follows that $H_2 = P^{-1} X_{1,2}$ and $W_2 = X_{2,1} P$. That is,

$$W = \begin{pmatrix} P \\ X_{2,1} P \end{pmatrix} = \begin{pmatrix} I \\ X_{1,2} \end{pmatrix} P \quad \text{and} \quad H = (P^{-1} \quad P^{-1} X_{1,2}) = P^{-1} (I \quad X_{1,2}).$$

Thus up to permutation, every factorization of $X = WH$ must have the form

$$W = \begin{pmatrix} I \\ W' \end{pmatrix} \quad H = (I \quad H'),$$

□

The recent paper [27] appears to be the only other work which contains uniqueness results for Boolean factorization. They characterize uniqueness via a property they called partial uniqueness, namely, given a fixed rank decomposition of $X = \vee_{j=1}^R X_j$, a rank one matrix X_i in the rank decomposition is *partially unique (with respect to the decomposition)* if the only matrix Y that satisfies $X = \vee_{j \neq i} X_j \vee Y$ is X_i . The authors of [27] show that the rank decomposition is unique if and only if each rank one matrix in the decomposition is partially unique. Partial uniqueness, and hence their characterization of uniqueness, differs from our work, in the sense that it places a simultaneous constraint on the row and columns of the decomposition $W \wedge H$. Our consideration for uniqueness are statements about W (uniqueness of cones) and H given a W (freeness of $P(X_i, W)$). In particular, Theorem 3 does not require the user to have acquired the feature weights H . The experimental sections that appear below with regards to stability demonstrate the usefulness of our approach.

5 Boolean Matrix Factorization with Automatic Model Selection

An important problem for all factorization methods is the model selection, that is, the estimation of the (usually unknown) number of latent features. Various heuristics to solve this problem have been proposed, including, Akaike's information criterion (AIC) [1], Bayesian information criterion (BIC) [34], minimum description length (MDL) [32], L-curve method [17], and stability method [7]. Automatic Relevance Determination (ARD) method, introduced for neural networks by MakCay [23], and applied later for PCA by Bishop [5], and for NMF by Fevotte and Tan [11] and Morup and Kai [28] is also popular. Some of the heuristics applied to NMF model selection have been also applied to BMF [26, 24]. However, since the nonnegative rank and Boolean rank can be different, using NMF-specific model selection algorithms could potentially produce incorrect estimation for the latent dimension in the Boolean case.

Ideally, one would like a model selection that allows for the identification of unique signatures when present. The following example demonstrates a challenge of this criteria. Namely, that small perturbations of a binary matrix could result in a loss of uniqueness in the Boolean decomposition:

Example 9. Let X be the Boolean matrix given by

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Then $\text{rk}_B(X) = 3$. Moreover, X contains an 3×3 permutation matrix. Thus, $\text{frk}_B(X) = 3$ by Proposition 7. By Theorem 5, X has unique Boolean factorization. Now, consider the matrix Y obtained by a single bit flip:

$$Y = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & \mathbf{1} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

From this minimal perturbation, we no longer have unique decomposition for Y . Indeed, the matrices W_1, W_2 given by

$$W_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad W_2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

can both be used to generate the data Y .

What Example 9 shows is that a small perturbation of the original matrix X could result in a loss of uniqueness. Thus unique factorizations, when they exist, are not stable solutions. In practical applications obtaining stable approximations is often more desirable than unstable exact solutions due to the presence of noise in the data and subsequent interpretations of the decomposition. We therefore desire a model selection algorithm which 1) discovers a stable approximation, such that 2) uniqueness is recovered when it exists.

Here we introduce a heuristic model selection algorithm for estimating the true number, K , of original Boolean sources, based on a stability criteria. We refer to this method as *BMFk*. *BMFk* is an analog for the Boolean semiring of a recent NMF automatic model determination method called *NMFk* [2]. It has been shown that when applied to a large number of synthetic datasets with predetermined latent dimensions, *NMFk* demonstrated a superior performance in comparison to other heuristics [29]. We aim to seek similar robust results for the Boolean semiring. In this section, we describe the *NMFk* and *BMFk* algorithms.

5.1 Robust Model Selection Algorithm

Algorithm 1: Stable Latent Dimension Selection Procedure

Input:
 $X \in \mathcal{S}^{M \times N}$: data
 $K_{\text{range}} \subset \mathbb{N}$: a list of explored latent dimensions k
 $S \in \mathbb{N}$: number of random resamplings and decompositions
Output: Silhouette scores, s_k , relative errors, e_k
 /* Computation part */

```

1 for  $k$  in  $K_{\text{range}}$  do
2   for  $s$  from 1 to  $S$  do
3      $X^{(s)} \sim \mathcal{P}(X)$ ; // Draw Sample from Random Ensemble
4      $W^{(s)}, H^{(s)} \leftarrow \text{decompose}(X^{(s)}, k)$ ; // Decompose Sample of Ensemble
5      $C \leftarrow \text{CustomCluster}(W^{(1)}, \dots, W^{(S)})$ ; // Cluster Solutions
6      $s_k \leftarrow \text{SilhouetteScore}(C)$ ; // Evaluate Cluster Stability
7      $e_k \leftarrow \text{ReconstructError}(X, C)$ ; // Evaluate Cluster Fit

```

NMF k and our proposed BMF k are heuristic methods that select latent dimensions of the initial matrix that give stable, or comparatively unique, nonnegative and Boolean approximations respectively. The general procedure for both of these algorithms checks each candidate latent dimension by: decomposing an ensemble of random matrices with mean the initial matrix, clustering the solutions, and evaluating the clusters stability and fit. Algorithm 7 outlines the unifying procedure that NMF k and BMF k follow to select latent dimensions that provide stable approximations for their respective decompositions.

In Algorithm 7, the drawing of matrices from a prescribed random distribution on line 3 constructs an ensemble of random matrices, that are slight perturbations of the initial matrix, to mitigate overfitting, to be decomposed with the appropriate algebra on line 4. The details for these two steps vary between NMF k and BMF k and are reported in the following sections. Following their respective random samplings and decompositions, custom clustering is applied in line 5 to the left factors of the set of decompositions, $\{W^{(1)}, \dots, W^{(S)}\}$. This custom clustering is an iterative centroid based algorithm similar to k -means for NMF k and k -medians for BMF k . The clustering assigns each column of each $W^{(s)}$ to a centroid, and updates the centroids based on the clusters [29]. Our custom clustering differs from k -means and k -medians with the additional constraint that each centroid can only be assigned one column of each $W^{(s)}$. This additional constraint enforces a 1-1 and onto mapping between the columns of each solution $W^{(s)}$ and the current centroids. This constraint simplifies the cluster assignment problem into solving S linear sum assignment problems on distance matrices between each $W^{(s)}$ and the current centroids. After each column assignment is done, the centroids are updated, and the procedure is iterated until converged.

The evaluation of the resulting clusters is done with two metrics, the silhouette score [32] on line 6 that measures the clusters stability, and a relative error metric on line 7 which measures the quality of fit. Silhouette scores range between -1, the poorest quality clustering, to 1, the highest quality clustering. We aggregate the silhouettes into a single score by taking the minimum of the average of the silhouettes for each cluster. A minimum silhouette score close to 1 indicates that the solutions of each $X^{(s)}$ in the ensemble form tight clusters. The relative error measure is evaluated from the medoids of the clusters and an H computed through nonnegative or Boolean regression. Since this process is done for each latent dimension of interest, L-statistics [39] can then be used to determine the optimal stable dimension, the elbows of the reconstruction errors and the silhouette scores. Further details of resampling, decomposing, and clustering for the individual implementations of NMF k and BMF k are described in the following sections.

5.2 NMF k

NMF k discovers the latent dimension that gives a robust nonnegative approximation for a given nonnegative matrix X by following the procedure given in Algorithm 7. The NMF k algorithm begins by drawing samples from a random distribution to form a random ensemble. While there is an abundance of suitable distributions to draw from for nonnegative data, our reported results use the elementwise resampling defined by $\mathcal{P}(x) = x \cdot \mathcal{U}(1 - \epsilon, 1 + \epsilon)$ where $\mathcal{U}(a, b)$ is a uniform distribution on the interval $[a, b)$. Alternative noise models can be selected appropriately for various matrices and applications.

For each element of the random ensemble, the sample is decomposed according to a prescribed noise model and decomposition algorithm on line 4. There are numerous NMF objective functions corresponding to different noise models, e.g. Kullback-Liebler divergence, Frobenius norm, Itakura-Saito divergence, as well as numerous algorithms for each objective function e.g. Multiplicative Update, Alternating Direction Method of Multipliers, Block Principal Pivoting [8]. NMF k decomposes each matrix in the ensemble, $X^{(s)}$, using a multiplicative update algorithm [21] to solve the Frobenius norm NMF problem:

$$\begin{aligned} \arg \min_{W, H} \quad & \frac{1}{2} \|X^{(s)} - WH\|_F^2 \\ \text{subject to} \quad & W \in \mathbb{R}_+^{N \times k}, \\ & H \in \mathbb{R}_+^{k \times M}. \end{aligned} \tag{1}$$

Multiplicative updates is an iterative algorithm that alternates optimizing over W and H

with the update rules

$$\begin{aligned} W &\leftarrow W \frac{XH^\top}{WHH^\top} \\ H &\leftarrow H \frac{W^\top X}{W^\top WH} \end{aligned} \quad (2)$$

always using the updated variables in the subsequent computation. This procedure is iterated until a fixed number of steps, or some convergence criteria is met. The multiplicative update algorithm preserves nonnegativity given nonnegative initializations, which were taken to be uniform random matrices.

In the clustering step of NMF k we employed a cosine similarity in the linear sum assignment problem to determine cluster membership. After the cluster assignments, the cluster centroids were updated as the medians of the assigned members to remove outlying solutions. Similarly, a cosine distance was employed in the silhouette statistic computation. In practice, the noise resampling technique, the NMF objective function and algorithm, and the distance metrics can all be selected for the application and the NMF k procedure successfully aids in the selection of a latent dimension that provides stable solutions.

5.3 BMF k

The BMF k algorithm follows the series of steps in Algorithm 7, adjusted for Boolean factorizations. Once again, an ensemble of BMF solutions $(W_s, H_s)_{s=1}^S$ is obtained from an ensemble of perturbed matrices $\{X^{(1)}, \dots, X^{(S)}\}$ for each permissible latent dimension k . For BMF k , the ensemble of random matrices is generated by randomly flipping elements of the original matrix X according a Bernoulli distribution. This corresponds to resampling the matrix X using the element-wise distribution $\mathcal{P}(x) = x +_f \mathcal{B}(p)$, where the $+_f$ operator is the bit flipping operator, and $\mathcal{B}(p)$ is the Bernoulli distribution with parameter p . The sampled matrices are then decomposed into a Boolean factorization using one of many BMF algorithms. In this work our reported results use the BANMF algorithm [38] with random initial conditions. For recent developments in the area of BMF, please see the review [25]. The BANMF algorithm solves the Boolean matrix factorization problem for each $X^{(s)}$ through the auxiliary optimization problem:

$$\begin{aligned} \arg \min_{Y, W, H} \quad & \|Y - WH\|_F \\ \text{subject to} \quad & 1 \leq Y_{ij} \leq k, \text{ if } X_{ij}^{(s)} = 1 \\ & Y_{ij} = 0, \text{ if } X_{ij}^{(s)} = 0 \\ & W \in \mathbb{R}_+^{N \times k}, \\ & H \in \mathbb{R}_+^{k \times M}. \end{aligned} \quad (3)$$

The factors W and H are updated via an alternating optimization procedure similar to the multiplicative updates, while the *auxiliary matrix* Y is updated by being projected to the constraint set according to the update rules,

$$\begin{aligned} W &\leftarrow W \frac{YH^\top}{WHH^\top} \\ H &\leftarrow H \frac{W^\top Y}{W^\top WH} \\ Y &\leftarrow X^{(s)} \star \text{maximum}(\text{minimum}(WH, k), 1), \end{aligned} \quad (4)$$

where \star , maximum, and minimum are the elementwise multiplication, maximum operator, and minimum operator respectively. The solution of this optimization problem will then be thresholded according to the following minimization problem,

$$\min_{w \in \mathbb{R}, h \in \mathbb{R}} \|X^{(s)} - (W \geq w) \otimes_B (H \geq h)\|_F \quad (5)$$

to arrive at a BMF solution. The grid-search thresholding algorithm is shown in [38].

In the clustering step, since the factors are Boolean, the custom clustering algorithm uses hamming distance to determine cluster assignment for each of the K features of each $W^{(s)}$. The

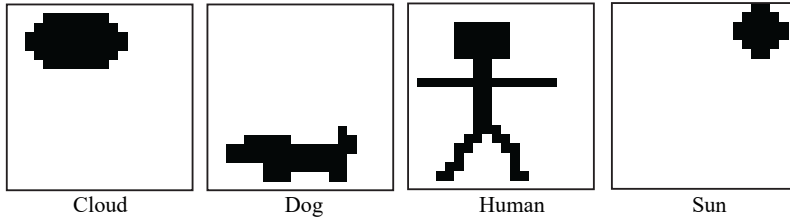


Figure 2: Four binary images are used as Boolean latent features to generate the synthetic image data.

centroids are updated by taking the elementwise medians of the vectors assigned to each cluster as this minimizes the hamming distance objective function, $\min_{c_k} \sum_{w \in \mathcal{W}_k} \text{hamming}(c_k, w)$, to find a centroid c_k for a cluster $\mathcal{W}_k = \{W_k^{(1)}, W_k^{(2)}, \dots, W_k^{(S)}\}$. Silhouette scores with hamming distance are used to determine how well each latent dimension K was clustered.

5.4 Experimental evaluation

We demonstrate the effectiveness of $\text{BMF}k$ at discovering the stable latent dimension in several synthetically generated datasets. Our first example compares the extracted signals from $\text{BMF}k$ to $\text{NMF}k$ on a matrix with different nonnegative and Boolean ranks. For our second example, we generate a large collection of Boolean matrices with unique BMF, and show that not only does $\text{BMF}k$ identify the correct latent dimension, but also the unique factors. We then inject noise into a random collection of Boolean matrices with unique BMF and show that $\text{BMF}k$ again discovers the correct latent dimension, and the discovered features are highly correlated with the “true” hidden features.

To evaluate an extracted feature, a_1 , relative to a corresponding generative feature, a_2 , we rely on the cosine similarity metric,

$$\frac{a_1^\top \cdot a_2}{\|a_1\| \|a_2\|}.$$

The cosine similarity applied to all pairs of recovered and generative features constructs a cosine similarity matrix.

To evaluate a decomposition, we aggregate the cosine similarities of both the left and right factors into a single *score metric* to measure how well the ground-truth decomposition is recovered [3]. For a pair of rank-one matrices $X_1 = a_1 b_1^\top$ and $X_2 = a_2 b_2^\top$, the score metric is defined as:

$$\text{score}(X_1, X_2) = \frac{a_1^\top a_2}{\|a_1\| \|a_2\|} \frac{b_1^\top b_2}{\|b_1\| \|b_2\|}.$$

For higher rank matrices and decompositions, the average of the scores of all rank one factors is taken, after the decompositions are permuted to maximize the scores. For reference, we also report results of other Boolean model selection strategies: (1) Minimum Description Length (MDL) [32] and (2) Covarge Quality (CQ) [37]. While $\text{BMF}k$ focuses on the solution stability, MDL minimizes the description of the data using the BMF model and CQ measures the change of the angle in the coverage (or reconstruction) curve.

5.4.1 Dataset with $\text{rk}_{\mathcal{B}}(X) \neq \text{rk}_{\mathbb{R}_+}(X)$

This example demonstrates that the underlying semiring can affect the latent dimension in practice. A synthetic image dataset is generated whose latent features are four binary 20×20 images, as shown in Figure 2. All possible (nonempty) Boolean combination images are generated, producing a dataset X of size 400×15 with Boolean rank $\text{rk}_{\mathcal{B}}(X) = 4$. The overlap between the images ‘human’ and ‘dog’, and between ‘cloud’ and ‘human’ forces $\text{rk}_{\mathbb{R}_+}(X) = 6$. Each image has an entry with a value of 1 where each other has a value of 0. It follows that X contains a 4×4 permutation matrix, so that by Theorem 5, X has a unique BMF.

Figure 3 (top) depicts the resulting silhouette and relative error scores of NMF_k and BMF_k for each candidate latent dimension. NMF_k correctly identifies the nonnegative rank at $K = 6$ and BMF_k correctly identifies the Boolean rank of the data of $K = 4$. Figure 3 (bottom) additionally depicts the score metric between the recovered factors of each method, and the Boolean images used to generate the matrix. Clearly, BMF_k perfectly recovered the factors with the scores of 1, while several of NMF_k 's scores are lower.

Figure 4 shows the extracted images from both methods. Note that the stable features identified from NMF_k include the overlapping regions. This is a more robust solution than say, completely separated features where small perturbations could disrupt such signals for reconstruction. The BMF_k technique has correctly identified the four unique signals.

Table 2 compares the selected dimension of several Boolean latent dimension selection models on the image data. Clearly, only BMF_k and $BANMF$ -MDL correctly identified the latent dimension, while $BANMF$ -CQ selected a lower latent dimension corresponding to a higher relative error plateau seen at $k = 2$ and $k = 3$.

Datasets	Boolean rank	BMF_k	$BANMF$ -MDL	$BANMF$ -CQ
image	4	4	4	2

Table 2: Comparisons of latent Boolean dimension selection algorithms applied to the image dataset.

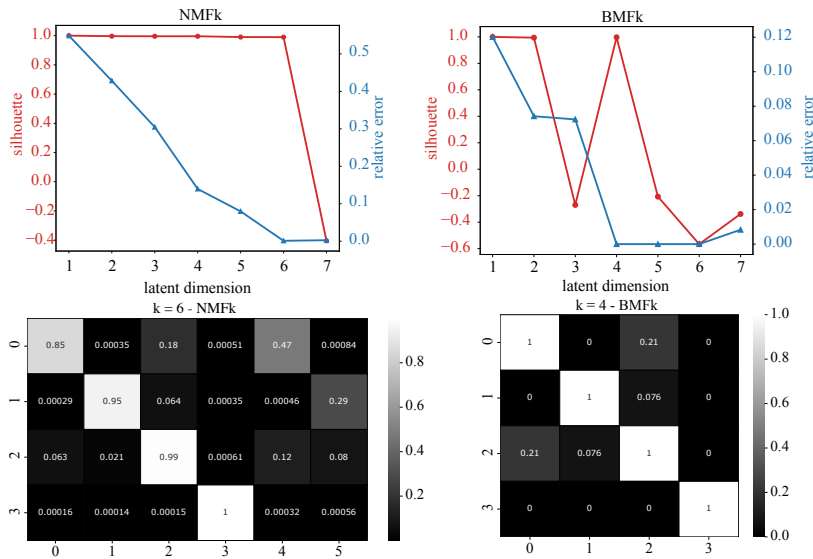


Figure 3: *Top* - NMF_k and BMF_k silhouette scores and relative errors for candidate latent dimensions. *Bottom* - Cosine similarity matrices comparing the extracted features to the generative Boolean features.

5.4.2 BMF_k stable rank finds unique features

The next experiments demonstrate that for matrices generated with unique Boolean features, BMF_k identifies the latent dimension corresponding to these unique decompositions, and extracts these unique patterns. We first show that for a class of matrices X with unique BMF, BMF_k exactly finds the unique features. We then add noise to these matrices and show that the BMF_k algorithm successfully recovers the correct latent dimension and features that are highly correlated with the original pre-corrupted unique features.

The dataset used for this experiment is generated based on the uniqueness result from Theorem 5. A single data matrix X is generated as following: First, a $W' \in \mathcal{B}^{15,5}$ and $H' \in \mathcal{B}^{5,15}$

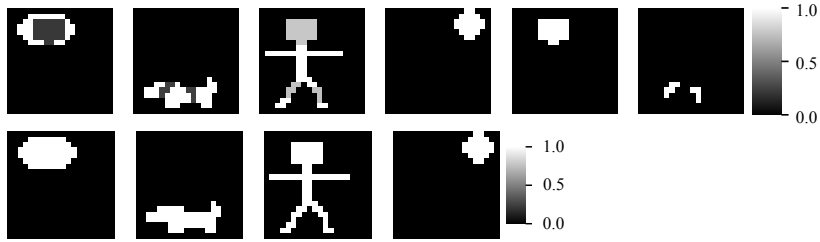


Figure 4: *Top* - NMF k extracted factors. *Bottom* - BMF k extracted factors are perfectly correlated with the true unique factors.

with a density of 0.3 are randomly generated. Then $X = \begin{pmatrix} I \\ W' \end{pmatrix} (I \ H')$. We apply BMF k on 1600 randomly generated dataset and collect the silhouette scores, relative Boolean error, and the score between the extracted features and the true factors. Figure 5 shows the average with one standard deviation band of silhouette scores and relative Boolean error of 1600 random matrices. Here we see that BMF k successfully identified the correct Boolean latent dimension at $k = 5$ for all matrices. Moreover, the factor scores are ones for all matrices indicating that BMF k can perfectly recover the original factors.

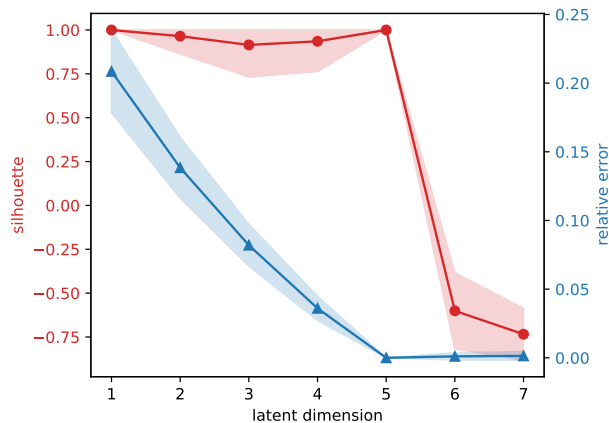


Figure 5: The mean values with one standard deviation band of silhouette scores and Boolean relative error of 1600 random matrices that have a unique factorization at $\text{rk}_B = 5$. BMF k correctly identify the Boolean rank, and extracted the unique features for all matrices.

Next, we generate a dataset using $W' \in \mathcal{B}^{15,4}$, $H' \in \mathcal{B}^{4,15}$, density of 0.3 and $X = \begin{pmatrix} I \\ W' \end{pmatrix} (I \ H')$. We then take this collection of matrices with unique BMF and corrupt the data by applying modest Bernoulli noise. Here, 5% of the entries of each matrix X were flipped. Once again, we have plotted the average and standard deviation of the silhouette scores with relative error in Figure 6 (a). In Figure 6 (b), we have plotted the distribution of the metric scores from all datasets. In this case, where the input data is noisy, BMF k can extract the ground truth latent features quite well.

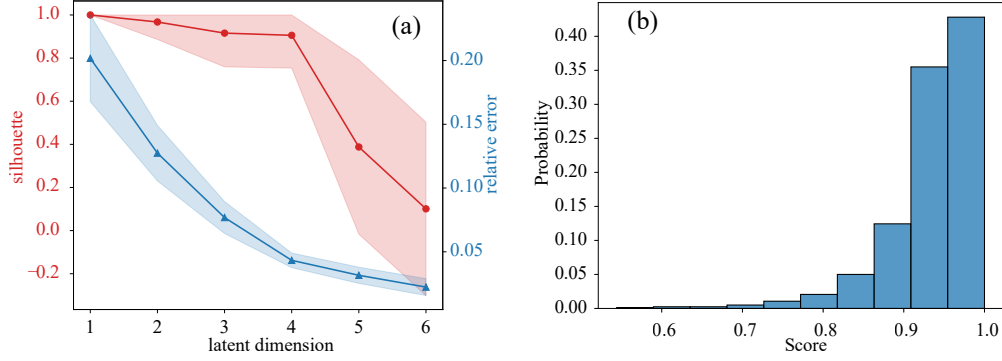


Figure 6: The mean values with one standard deviation band of silhouette scores and Boolean relative error of 1600 random datasets **with 5% flipping noise** that have a unique factorization at $\text{rk}_{\mathcal{B}} = 4$.

Acknowledgements

This work was supported by the LDRD program of Los Alamos National Laboratory under project number 20190020DR and the Center for Nonlinear Studies. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001).

6 Appendix

6.1 Proofs for rank results

Here we provide detailed proofs of Proposition 1 and Theorem 2:

Proposition 1. Let $X \in \mathbb{B}^{N,M}$. If $X = WH$ is a nonnegative factorization of X , then $\hat{W} \wedge \hat{H}$ is a Boolean factorization of X .

Proof. Note that from the nonnegative factorization,

$$\sum_{k=1}^K W_{i,k} H_{k,j} = X_{i,j} \in \{0, 1\}.$$

Because each entry in the summand is nonnegative, the only way the above sum results in zero is if each element is itself zero. Similarly, the only way the above sum results in one is if at least one of the products $W_{i,k} H_{k,j}$ is non-zero.

Now consider the Boolean product $Y = \hat{W} \wedge \hat{H}$. We will show that $Y = X$. By definition,

$$Y_{i,j} = \bigvee_{k=1}^K \hat{W}_{i,k} \wedge \hat{H}_{k,j}.$$

If $X_{i,j} = 0$, then by above each $W_{i,k} H_{k,j} = 0$, and hence each $\hat{W}_{i,k} \wedge \hat{H}_{k,j} = 0$. In this case, $Y_{i,j} = 0$. Similarly, if $X_{i,j} = 1$, then because there exists some product $W_{i,k} H_{k,j} \neq 0$, we have that $\hat{W}_{i,k} \wedge \hat{H}_{k,j} = 1$. In this case, $Y_{i,j} = 1$. Therefore, we have $Y = X$. \square

Next, we show that if $X \in \mathbb{B}^{N,M}$, then $\text{rk}_{\mathbb{Z}_2}(X) \leq \text{rk}_{\mathbb{R}}(X)$. Recall that if X is an $R \times R$ matrix over a field \mathbb{F} , one can define its determinant. Concretely, if $X = (X_{i,j})_{i,j=1}^R$, then

$$\det(X) := \sum_{\sigma \in \mathcal{S}_R} \text{sgn}(\sigma) \prod_{i=1}^R X_{\sigma(i),i},$$

where σ is a permutation of $\{1, \dots, R\}$, S_R is the set of all permutations, and $\text{sgn}(\sigma) \in \{-1, 1\}$ is the sign of the permutation σ .

If X is an $R \times R$ matrix over a field \mathbb{F} , then $\det(X) \neq 0$ if and only if X is invertible [18]. For a $N \times M$ matrix X over a field \mathbb{F} , the row rank, column rank, and factor rank are all equal [33]. Therefore X has rank R if and only if X contains a $R \times R$ submatrix X' with full rank, and every larger square submatrix is not full rank. Thus $X \in \mathbb{F}^{N,M}$ has rank R if and only if X contains a $R \times R$ submatrix X' with $\det(X') \neq 0$ and every larger square submatrix has determinant zero.

Theorem 2. Let $X \in \mathbb{B}^{N,M}$. Then $\text{rk}_{\mathbb{Z}_2}(X) \leq \text{rk}_{\mathbb{R}}(X)$.

Proof. Let $\text{rk}_{\mathbb{R}}(X) = R$. By the comments preceding the theorem, if $K > R$, then every $K \times K$ submatrix X' of X will have real determinant 0. That is,

$$\det_{\mathbb{R}}(X') = \sum_{\sigma \in S_K} \text{sgn}(\sigma) \prod_{i=1}^K X_{\sigma(i),i} = 0$$

For each σ , $\prod_{i=1}^K X_{\sigma(i),i}$ is either 0 or 1. Hence, the above summation is a some quantity of 1's perfectly balanced by an equal quantity of -1 's. Therefore, the summation consists of an even number of non-zero elements.

Now, consider the \mathbb{Z}_2 determinant of X' . Since $-1 = 1$ in \mathbb{Z}_2 , $\text{sgn}(\sigma) = 1$ for each σ . Therefore, we have that

$$\det_{\mathbb{Z}_2}(X') := \bigoplus_{\sigma \in S_K} \bigotimes_{i=1}^K X_{\sigma(i),i},$$

where now the sum and product are happening over \mathbb{Z}_2 . Note the product $\bigotimes_{i=1}^K X_{\sigma(i),i}$ happening within \mathbb{Z}_2 is the same as within \mathbb{R} , and so remains unchanged. From above, we know there is an even number of non-zero elements in this sum. Hence $\det_{\mathbb{Z}_2}(X') = 0$. By the comments preceding the theorem, $\text{rk}_{\mathbb{Z}_2}(X) \leq R = \text{rk}_{\mathbb{R}}(X)$. □

6.2 Proof that minimal generating set for Boolean cone unique

We will show that a minimal generating set for a cone \mathcal{C} over the Boolean semiring \mathcal{B}^N is necessarily unique. To prove this, we require the following order notation:

Definition 9. Let

$$[0, y) := \{x \in \mathcal{C} : x < y\}$$

be the set of vectors strictly dominated by y . We say that y is *minimal* in \mathcal{C} if

$$\bigvee_{x \in [0, y)} x < y.$$

Finally, we define

$$\min(\mathcal{C}) := \{y \in \mathcal{C} : y \text{ is minimal in } \mathcal{C}, y \neq 0\}.$$

For example, consider the cone

$$\mathcal{C} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

Then $\min(\mathcal{C}) = \{(1, 0, 0), (0, 1, 0), (1, 0, 1)\}$. We will show that $\min(\mathcal{C})$ is the unique minimal generating set, and therefore the set of extreme rays are unique.

Theorem 6. Let $\mathcal{C} \subset \mathcal{B}^N$ be a cone. Then the set of extreme rays are unique, and are equal to $\min(\mathcal{C})$.

Proof. We will show that $\min(\mathcal{C})$ is a subset of any generating set. We will then show that $\min(\mathcal{C})$ also generates \mathcal{C} . It will then follow that the unique minimal generating set is $\min(\mathcal{C})$.

First, we note that if $G \subset \mathcal{C}$ is a generating set, then $\min(\mathcal{C}) \subset G$. By definition, if $y \in \min(\mathcal{C})$ then $y \in \mathcal{C}$ and nothing else from \mathcal{C} can be added to get y . Hence, it must be in any generating set G .

Next, we show that $\min(\mathcal{C})$ spans \mathcal{C} . Let $\mathcal{C}' = \text{span}_{\vee}(\min(\mathcal{C}))$. Suppose to the contrary that there exists some $y \in \mathcal{C}$ but $y \notin \mathcal{C}'$. Then certainly, $y \notin \min(\mathcal{C})$ so that

$$\bigvee_{x \in [0, y)} x = y.$$

But consider now $[0, y)$. Since $y \notin \mathcal{C}'$, there must exist a $y_1 \in [0, y)$ such that $y_1 \notin \mathcal{C}'$ and

$$\bigvee_{x \in [0, y) \setminus \{y_1\}} x < y.$$

One can then apply this process again on the new vector y_1 . Since any vector in \mathcal{B}^N has at most N non-zero entries, this process must terminate after a finite number steps k . This means that $y_k \notin \mathcal{C}'$, but yet any sub-vector that is required to build y_k must belong to \mathcal{C}' . This is clearly a contradiction, implying that y must have belonged to \mathcal{C}' . \square

References

- [1] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- [2] Ludmil B Alexandrov, Serena Nik-Zainal, David C Wedge, Samuel AJR Aparicio, Sam Behjati, Andrew V Biankin, Graham R Bignell, Niccolo Bolli, Ake Borg, Anne-Lise Børresen-Dale, et al. Signatures of mutational processes in human cancer. *Nature*, 500(7463):415–421, 2013.
- [3] Casey Battaglino, Grey Ballard, and Tamara G Kolda. A practical randomized cp tensor decomposition. *SIAM Journal on Matrix Analysis and Applications*, 39(2):876–901, 2018.
- [4] LeRoy B Beasley. Isolation number versus boolean rank. *Linear algebra and its applications*, 436(9):3469–3474, 2012.
- [5] Christopher M Bishop. Bayesian pca. *Advances in neural information processing systems*, pages 382–388, 1999.
- [6] Thomas Andrew Brown, Mario Leon Juncosa, and VL Klee. Invertibly positive linear operators on spaces of continuous functions. Technical report, RAND CORP SANTA MONICA CALIF, 1968.
- [7] Jean-Philippe Brunet, Pablo Tamayo, Todd R Golub, and Jill P Mesirov. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the national academy of sciences*, 101(12):4164–4169, 2004.
- [8] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.
- [9] Joel E Cohen and Uriel G Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra and its Applications*, 190:149–168, 1993.
- [10] Faun CC Doherty, J Richard Lundgren, and Daluss J Siewert. Biclique covers and partitions of bipartite graphs and digraphs and related matrix ranks of $\{0, 1\}$ -matrices. *Congressus Numerantium*, pages 73–96, 1999.
- [11] Cédric Févotte and A Taylan Cemgil. Nonnegative matrix factorizations as probabilistic inference in composite models. In *2009 17th European Signal Processing Conference*, pages 1913–1917. IEEE, 2009.
- [12] Stéphane Gaubert. *Théorie des systèmes linéaires dans les dioïdes*. PhD thesis, Paris, ENMP, 1992.

- [13] Floris Geerts, Bart Goethals, and Taneli Mielikäinen. Tiling databases. In *International conference on discovery science*, pages 278–289. Springer, 2004.
- [14] Nicolas Gillis. Sparse and unique nonnegative matrix factorization through data preprocessing. *The Journal of Machine Learning Research*, 13(1):3349–3386, 2012.
- [15] Michel Gondran and Michel Minoux. *Graphs, dioids and semirings: new models and algorithms*, volume 41. Springer Science & Business Media, 2008.
- [16] Harold W Gutch, Peter Gruber, Arie Yeredor, and Fabian J Theis. Ica over finite fields—separability and algorithms. *Signal Processing*, 92(8):1796–1808, 2012.
- [17] Per Christian Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM review*, 34(4):561–580, 1992.
- [18] Serge Lang. Algebra. 3rd. *Edition Addison–Wesley*, 1993.
- [19] Hans Laurberg, Mads Græsbøll Christensen, Mark D Plumbley, Lars Kai Hansen, and Søren Holdt Jensen. Theorems on positive data: On the uniqueness of nmf. *Computational intelligence and neuroscience*, 2008, 2008.
- [20] Robert S Ledley. The inverse of a boolean matrix. Technical report, NATIONAL BIOMEDICAL RESEARCH FOUNDATION WASHINGTON DC, 1965.
- [21] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [22] Tao Li. A general model for clustering binary data. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 188–197, 2005.
- [23] David JC MacKay and Radford M Neal. Automatic relevance determination for neural networks. In *Technical Report in preparation*. Cambridge University, 1994.
- [24] Tatiana Makhlova and Martin Trneca. From-below boolean matrix factorization algorithm based on mdl. *Advances in Data Analysis and Classification*, 15(1):37–56, 2021.
- [25] Pauli Miettinen and Stefan Neumann. Recent developments in boolean matrix factorization. *arXiv preprint arXiv:2012.03127*, 2020.
- [26] Pauli Miettinen and Jilles Vreeken. Mdl4bmf: Minimum description length for boolean matrix factorization. *ACM transactions on knowledge discovery from data (TKDD)*, 8(4):1–31, 2014.
- [27] Sebastian Miron, Mamadou Diop, Anthony Larue, Eddy Robin, and David Brie. Boolean decomposition of binary matrices using a post-nonlinear mixture approach. *Signal Processing*, 178:107809, 2021.
- [28] Morten Mørup and Lars Kai Hansen. Tuning pruning in sparse non-negative matrix factorization. In *2009 17th European Signal Processing Conference*, pages 1923–1927. IEEE, 2009.
- [29] Benjamin T Nebgen, Raviteja Vangara, Miguel A Hombrados-Herrera, Svetlana Kuksova, and Boian S Alexandrov. A neural network for determination of latent dimensionality in non-negative matrix factorization. *Machine Learning: Science and Technology*, 2(2):025012, 2021.
- [30] James Orlin et al. Contentment in graph theory: covering graphs with cliques. In *Indagationes Mathematicae (Proceedings)*, volume 80, pages 406–424. North-Holland, 1977.
- [31] Christophe Reutenauer and Howard Straubing. Inversion of matrices over a commutative semiring. *Journal of Algebra*, 88(2):350–360, 1984.
- [32] Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.
- [33] Steven Roman, S Axler, and FW Gehring. *Advanced linear algebra*, volume 3. Springer, 2005.
- [34] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

- [35] Gilbert W Stewart. On the early history of the singular value decomposition. *SIAM review*, 35(4):551–566, 1993.
- [36] Bit-Shun Tam. A geometric treatment of generalized inverses and semigroups of nonnegative matrices. *Linear Algebra and its Applications*, 41:225–272, 1981.
- [37] Martin Trnecka and Marketa Trneckova. Model order selection for approximate boolean matrix factorization problem. *Knowledge-Based Systems*, page 107184, 2021.
- [38] Duc P. Truong, Erik Skau, Derek Desantis, and Boian Alexandrov. Boolean matrix factorization via nonnegative auxiliary optimization. *IEEE Access*, 9:117169–117177, 2021.
- [39] Raviteja Vangara, Manish Bhattarai, Erik Skau, Gopinath Chennupati, Hristo Djidjev, Thomas Tierney, James P Smith, Valentin G Stanev, and Boian S Alexandrov. Finding the number of latent topics with semantic non-negative matrix factorization. *IEEE Access*, 2021.
- [40] Arie Yeredor. Independent component analysis over galois fields of prime order. *IEEE Transactions on Information Theory*, 57(8):5342–5359, 2011.
- [41] Zhongyuan Zhang, Tao Li, Chris Ding, and Xiangsun Zhang. Binary matrix factorization with applications. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pages 391–400. IEEE, 2007.