

A Post-Nonlinear Mixture Model Approach to Binary Matrix Factorization (BMF)

Mamadou Diop

CEA Tech Grand-Est Metz Technopole and
Université de Lorraine, CRAN, UMR 7039

Email: mamadou.diop@univ-lorraine.fr
mamadou.diop@cea.fr

Anthony Larue

CEA, LIST
Gif-sur-Yvette Cedex, 91191, France

Email: anthony.larue@cea.fr

Sebastian Miron and David Brie

Université de Lorraine, CRAN, UMR 7039
Vandœuvre, F-54506, France

Email: sebastian.miron@univ-lorraine.fr
david.brie@univ-lorraine.fr

Abstract—In this paper, we address the Binary Matrix Factorization (BMF) problem which is the restriction of the nonnegative matrix factorization (NMF) to the binary matrix case. A necessary and sufficient condition for the identifiability for the BMF model is given. We propose to approach the BMF problem by the NMF problem using a nonlinear function which guarantees the binarity of the reconstructed data. Two new algorithms are introduced and compared in simulations with the state of art BMF algorithms.

I. INTRODUCTION

The data processed in most applications are real or complex-valued but there is also a lot of data that are naturally binary, *i.e.*, taking two discrete values, most oftenly 0 and 1. The factorization of binary matrices has a large number of applications such as: association rule mining for agaricus-lepiota mushroom data sets [1], high-dimensional discrete-attribute data mining [2], biclustering structure identification for gene expression data sets [3], market basket data clustering [4], digits reconstruction for USPS data sets [5], pattern discovery for gene expression pattern images [6], or recommendation systems [7].

In this article, we focus on Binary Matrix Factorization (BMF) methods which can be seen as restrictions of Non-negative Matrix Factorization (NMF) [8], [9] to the binary case. The principle of BMF is to factorize a binary matrix \mathbf{X} into two binary matrices \mathbf{W} and \mathbf{H} such that: $\mathbf{X} \approx \mathbf{W} \odot \mathbf{H}^T$; (\odot) is the binary matrix product which will be defined in section 2.

Several BMF algorithms have been proposed in the literature. In [10], the BFM problem is seen as a Discrete Basis Problem (DBP). A greedy-like algorithm (ASSO) for solving DBP problem based on the associations between the columns of \mathbf{X} is proposed. In [11], Tu *et al.* proposed a binary matrix factorization algorithm under the Bayesian Ying-Yang (BYY) learning, to predict protein complexes from protein-protein interactions networks. The proposed BYY-BMF algorithm automatically determines the number of clusters, while this number is pre-given for most existing BMF algorithms. In [12], Jiang and Heath proposed a variant of BMF where the real matrix product is restricted to the class of binary matrices, by exploring the relationship between BMF and special classes of clustering problems. Belohlavek and Vychodil studied in [13] the problem of factor analysis of three-way binary data.

Thus the problem consists of finding a decomposition of binary data into three binary matrices: an object-factor matrix, an attribute factor matrix and a condition factor matrix. The factors are provided by triadic concepts developed in formal concept analysis. In [14], Yeredor studied the Independent Component Analysis (ICA) for the case of binary sources, where addition had the meaning of the boolean exclusive OR (XOR) operation. In [15], Zhang *et al.* extended the standard NMF to BMF. They proposed two algorithms which solve the NMF problem with binary constraints on \mathbf{W} and \mathbf{H} . In this paper, we introduce a post-nonlinear model that approximates better the binary mixture model $\mathbf{W} \odot \mathbf{H}^T$ than the model introduced in [15], and propose two algorithms to solve the unmixing problem. We also address the identifiability of BMF model and provide a necessary and sufficient condition of identifiability.

The notations hereafter are used:

- \mathbf{X} : matrix
- \mathbf{x}_j : j^{th} vector column of matrix \mathbf{X}
- \mathbf{X}_{ij} : entry (i, j) of matrix \mathbf{X}
- k, K : scalar values
- $\mathbf{1}_{N \times M}$: all-one matrix of size $N \times M$
- \mathbf{X}_k : the k^{th} source (rank-1) matrix in the decomposition of \mathbf{X}

The rest of this paper is organized as follows. In Section 2, we define the BMF problem and study algorithms proposed in [15]. In Section 3, we provide a theoretical identifiability condition for the BMF model. In Section 4, the proposed post-nonlinear mixture approach is presented along with the two novel estimation algorithms. In section 5, the performances of the proposed algorithms are evaluated in numerical simulations.

II. BINARY MATRIX FACTORIZATION (BMF)

A. The BMF problem

1) *Direct problem formulation*: The direct BMF problem is to compute a binary matrix \mathbf{X} ($\mathbf{X}_{ij} \in \{0, 1\}$) of size $N \times M$, given by two binary matrices \mathbf{W} and \mathbf{H} of respective sizes $N \times K$ and $M \times K$ as:

$$\mathbf{X} = \mathbf{W} \odot \mathbf{H}^T. \quad (1)$$

(\odot) is called binary matrix product and is defined as [16], [17]: $\mathbf{X}_{ij} = \bigvee_{k=1}^K (\mathbf{W}_{ik} \wedge \mathbf{H}_{jk})$, where (\vee) and (\wedge) are OR and AND logical operators, respectively.

Thus, hereafter, we study the possible inverse problem formulations for the direct problem in (1).

2) *Inverse problem formulation*: The inverse problem corresponding to the direct problem in (1) is to find, from the binary matrix \mathbf{X} of size $N \times M$, two binary matrices \mathbf{W} and \mathbf{H} of respective sizes $N \times K$ and $M \times K$ such that:

$$\{\mathbf{W}, \mathbf{H}\} = \underset{\mathbf{W}, \mathbf{H} \in \{0,1\}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{W} \odot \mathbf{H}^T\|_2^2. \quad (2)$$

The l_2 -norm in (2) can be replaced equivalently (in the binary case) by the l_1 or the l_0 -norm. The problem in (2) is NP-complete [10] and therefore, in order to develop efficient algorithms for solving this inverse problem, reformulations of (2) have been proposed in the literature.

B. A BMF linear mixture model

In this reformulation of the BMF, the binary matrix product (\odot) has been replaced by the real matrix product [15]. In other words, the direct problem (1) is replaced by the linear problem:

$$\mathbf{X} = \mathbf{W}\mathbf{H}^T. \quad (3)$$

Thus, the corresponding inverse problem comes down to the NMF problem with an additional constraint on the binarity of \mathbf{W} and \mathbf{H} . Thus, the inverse problem for (3) can be expressed as:

$$\{\mathbf{W}, \mathbf{H}\} = \underset{\mathbf{W}, \mathbf{H} \in \{0,1\}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{W}\mathbf{H}^T\|_2^2. \quad (4)$$

The two algorithms developed in [15] solve this inverse problem. One of them is based on a penalty function (PF) and uses a gradient descent with multiplicative update rule, similar to NMF, to minimize the cost function:

$$J(\mathbf{W}, \mathbf{H}) = \sum_{i,j} (\mathbf{X}_{ij} - (\mathbf{W}\mathbf{H}^T)_{ij})^2 + \frac{1}{2}\lambda \sum_{i,k} (\mathbf{W}_{ik}^2 - \mathbf{W}_{ik})^2 + \frac{1}{2}\lambda \sum_{j,k} (\mathbf{H}_{jk}^2 - \mathbf{H}_{jk})^2. \quad (5)$$

A second algorithm proposed in [18], is based on a thresholding (TH) procedure of the values of matrices \mathbf{W} and \mathbf{H} resulting from standard NMF. The idea is to find two thresholds w and h for the two matrices \mathbf{W} and \mathbf{H} , respectively, that minimize the cost function:

$$F(w, h) = \sum_{i,j} \left(\mathbf{X}_{ij} - (\theta(\mathbf{W} - \mathbf{1}_{N \times K} \cdot w) \theta(\mathbf{H} - \mathbf{1}_{M \times K} \cdot h))^T \right)_{ij}^2, \quad (6)$$

where θ is the Heaviside step function applied element-wise. These two algorithms, based on the linear model mixture, will be used in this paper as benchmark for our approach. Before introducing the proposed approach, we study in the next section, the identifiability of the BMF model (1).

III. IDENTIFIABILITY OF THE BMF MIXTURE MODEL

In this paper, the identifiability of the BMF model is understood as the uniqueness of the BMF decomposition $\mathbf{X} = \mathbf{W} \odot \mathbf{H}^T$. By uniqueness we mean that, if another couple of matrices ($\bar{\mathbf{W}}, \bar{\mathbf{H}}$) verify (1), i.e.,:

$$\mathbf{X} = \mathbf{W} \odot \mathbf{H}^T = \bar{\mathbf{W}} \odot \bar{\mathbf{H}}^T, \quad (7)$$

then (\mathbf{W}, \mathbf{H}) and ($\bar{\mathbf{W}}, \bar{\mathbf{H}}$) are the same up to a joint permutation of the columns of \mathbf{W} and \mathbf{H} . Note that, because of the binary nature of \mathbf{W} and \mathbf{H} , there is no scale indeterminacy in this case.

Definition III.1. (Partial identifiability)

We say that the model (1) is *partially identifiable* if only one or several columns of \mathbf{W} and \mathbf{H} can be uniquely estimated from \mathbf{X} .

$\operatorname{supp}(\mathbf{x}) = \{i, \mathbf{x}_i \neq 0\}$ denotes the support of vector \mathbf{x} and $\operatorname{supp}(\mathbf{X}) = \{(i, j), \mathbf{X}_{ij} \neq 0\}$ denotes the support of matrix \mathbf{X} .

Theorem III.1 (Partial identifiability). *The ℓ^{th} column of \mathbf{W} i.e., \mathbf{w}_ℓ can be uniquely estimated from \mathbf{X} iff:*

$$\forall n = 1, \dots, N, \text{ with } \mathbf{w}_\ell(n) \neq 0, \operatorname{supp}(\mathbf{e}_n \odot \mathbf{h}_\ell^T) \not\subseteq \bigcup_{k \neq \ell}^K \operatorname{supp}(\mathbf{X}_k) \quad (8)$$

with $\mathbf{e}_n = [0, \dots, 1, \dots, 0]^T$ the n^{th} vector of the canonical basis of \mathbb{R}^N

Proof. Let K be the number of columns of \mathbf{W} and \mathbf{H} . We can write:

$$\mathbf{X} = \bigvee_{k=1}^K (\mathbf{w}_k \odot \mathbf{h}_k^T) = \bigvee_{k=1}^K \mathbf{X}_k = \bigvee_{k \neq \ell}^K \mathbf{X}_k \vee (\mathbf{w}_\ell \odot \mathbf{h}_\ell^T) \quad (9)$$

Suppose that $\exists n \in \{1, \dots, N\}$ with $\mathbf{w}_\ell(n) \neq 0$ such as:

$$\operatorname{supp}(\mathbf{e}_n \odot \mathbf{h}_\ell^T) \subseteq \bigcup_{k \neq \ell}^K \operatorname{supp}(\mathbf{X}_k).$$

Then, from (9) we have:

$$\operatorname{supp}(\mathbf{X}) = \bigcup_{k=1}^K \operatorname{supp}(\mathbf{X}_k) = \bigcup_{k \neq \ell}^K \operatorname{supp}(\mathbf{X}_k) \cup \operatorname{supp}(\mathbf{w}_\ell \odot \mathbf{h}_\ell^T)$$

As $\mathbf{w}_\ell(n) \neq 0$, the following relation holds: $\mathbf{w}_\ell = \mathbf{w}_\ell \oplus \mathbf{e}_n \vee \mathbf{e}_n$, where (\oplus) and (\vee) denote the XOR and OR logical operator element-wise, respectively. Thus we can write:

$$\begin{aligned} \operatorname{supp}(\mathbf{X}) &= \bigcup_{k \neq \ell}^K \operatorname{supp}(\mathbf{X}_k) \cup \operatorname{supp}((\mathbf{w}_\ell \oplus \mathbf{e}_n \vee \mathbf{e}_n) \odot \mathbf{h}_\ell^T) \\ &= \bigcup_{k \neq \ell}^K \operatorname{supp}(\mathbf{X}_k) \cup \operatorname{supp}((\mathbf{w}_\ell \oplus \mathbf{e}_n) \odot \mathbf{h}_\ell^T) \cup \operatorname{supp}(\mathbf{e}_n \odot \mathbf{h}_\ell^T) \end{aligned}$$

As $\operatorname{supp}(\mathbf{e}_n \odot \mathbf{h}_\ell^T) \subseteq \bigcup_{k \neq \ell}^K \operatorname{supp}(\mathbf{X}_k)$, we have:

$$\operatorname{supp}(\mathbf{X}) = \bigcup_{k \neq \ell}^K \operatorname{supp}(\mathbf{X}_k) \cup \operatorname{supp}((\mathbf{w}_\ell \oplus \mathbf{e}_n) \odot \mathbf{h}_\ell^T)$$

\Updownarrow

$$\mathbf{X} = \bigvee_{k \neq \ell}^K \mathbf{X}_k \vee (\mathbf{w}_\ell \oplus \mathbf{e}_n) \odot \mathbf{h}_\ell^T$$

As $\mathbf{w}_\ell \neq (\mathbf{w}_\ell \oplus \mathbf{e}_n)$, this means that there exists another $\bar{\mathbf{w}}_\ell = \mathbf{w}_\ell \oplus \mathbf{e}_n \neq \mathbf{w}_\ell$ that satisfies (7). Therefore, \mathbf{w}_ℓ can not be uniquely estimated, which ends the proof. A similar

condition can be also derived for the columns of \mathbf{H} . \square

Intuitively, Theorem III.1 states that a source $\mathbf{X}_\ell = \mathbf{w}_\ell \odot \mathbf{h}_\ell^T$ can be uniquely estimated from \mathbf{X} if none of its columns and none of its rows has the support included in the union of the supports of all the other sources.

From Theorem III.1, we deduce immediately the following corollary which gives the necessary and sufficient identifiability condition for the BMF model.

Corollary III.1.1 (Identifiability of BMF model (1)). *The BMF model (1) is identifiable iff: $\forall k = 1, \dots, K$, \mathbf{w}_k and \mathbf{h}_k can be uniquely estimated from \mathbf{X} , i.e., Theorem III.1 holds for all the columns of \mathbf{W} and \mathbf{H} .*

By replacing the binary problem (2) with a problem on the real numbers domain, with constraints on \mathbf{W} and \mathbf{H} (4), the BMF problem is reduced to a classical real-valued optimization problem, easier to solve. However, there is no guarantee that, in general, the solution to the inverse problem (4) is identical to the solution of the initial problem (2). That is why, we propose in this next section a post-nonlinear BMF model that better approximates the original model (1) and therefore allows to obtain better estimates of the binary matrices \mathbf{W} and \mathbf{H} .

IV. A POST-NONLINEAR MIXTURE APPROACH TO BMF

We propose in this section, a post-nonlinear mixture model that is equivalent to the binary matrix product (1) when \mathbf{W} and \mathbf{H} are binary. We preserve the idea of replacing the binary matrix product by the real matrix product with binary constraints on \mathbf{W} and \mathbf{H} , but we introduce a nonlinear function (Φ) which guarantees the binarity of the reconstructed data from \mathbf{X} . Thus, the direct BMF problem can be expressed as:

$$\mathbf{X} = \Phi(\mathbf{W}\mathbf{H}^T). \quad (10)$$

In this paper, we choose $\Phi(x) = \frac{1}{1+e^{-\gamma(x-0.5)}}$ as the sigmoid function, with a fixed parameter γ which allows to adjust the sigmoid slope.

The inverse problem for (10) can be expressed as:

$$\{\mathbf{W}, \mathbf{H}\} = \underset{\mathbf{W}, \mathbf{H} \in \{0,1\}}{\operatorname{argmin}} \|\mathbf{X} - \Phi(\mathbf{W}\mathbf{H}^T)\|_2^2. \quad (11)$$

To solve this problem, we propose an algorithm based on a gradient descent with a multiplicative update rule, similar to the PF algorithm proposed in [15]. The proposed algorithm minimizes the cost function hereafter, which is similar to the PF cost function, but takes into account the nonlinearity Φ :

$$G(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \sum_{i,j} (\mathbf{X}_{ij} - \Phi(\mathbf{W}\mathbf{H}^T)_{ij})^2 + \frac{1}{2} \lambda \sum_{i,k} (\mathbf{W}_{ik}^2 - \mathbf{W}_{ik})^2 + \frac{1}{2} \lambda \sum_{j,k} (\mathbf{H}_{jk}^2 - \mathbf{H}_{jk})^2. \quad (12)$$

The binary constraint on \mathbf{H} and \mathbf{W} is imposed by using the penalty terms $\mathbf{H}_{jk}^2 - \mathbf{H}_{jk}$ and $\mathbf{W}_{ik}^2 - \mathbf{W}_{ik}$. To minimize (12),

a gradient descent method with multiplicative update rule, similar to NMF [8], [9] is used. Thus, the proposed algorithm (PNL-PF) can be summarized in algorithm 1.

Algorithm 1 : Post NonLinear Penalty Function algorithm (PNL-PF)

Input: \mathbf{X} , K , Nb_{iter} , ε

Output: \mathbf{W} , \mathbf{H}

STEP 1: Initialization

$\mathbf{W} \leftarrow \text{rand}(N, K)$, $\mathbf{H} \leftarrow \text{rand}(M, K)$

$p = 0$

STEP 2: Update of \mathbf{W} and \mathbf{H}

$$\mathbf{H}_{jk} \leftarrow \mathbf{H}_{jk} \frac{\left(\mathbf{X} \frac{\partial}{\partial \mathbf{H}_{jk}} \Phi(\mathbf{W}\mathbf{H}^T) \right)_{jk} + 3\lambda \mathbf{H}_{jk}^2}{\left(\frac{\partial}{\partial \mathbf{H}_{jk}} \Phi(\mathbf{W}\mathbf{H}^T) \right)_{jk} + 2\lambda \mathbf{H}_{jk}^3 + \lambda \mathbf{H}_{jk}}$$

$$\mathbf{W}_{ik} \leftarrow \mathbf{W}_{ik} \frac{\left(\mathbf{X} \frac{\partial}{\partial \mathbf{W}_{ik}} \Phi(\mathbf{W}\mathbf{H}^T) \right)_{ik} + 3\lambda \mathbf{W}_{ik}^2}{\left(\frac{\partial}{\partial \mathbf{W}_{ik}} \Phi(\mathbf{W}\mathbf{H}^T) \right)_{ik} + 2\lambda \mathbf{W}_{ik}^3 + \lambda \mathbf{W}_{ik}}$$

STEP 3: Normalization

$$\mathbf{W} \leftarrow \mathbf{W} \mathbf{D}_W^{-1/2} \mathbf{D}_H^{1/2} \quad \mathbf{H} \leftarrow \mathbf{H} \mathbf{D}_H^{-1/2} \mathbf{D}_W^{1/2}$$

with $\mathbf{D}_W = \text{diag}(\max(\mathbf{w}_1), \dots, \max(\mathbf{w}_K))$
 $\mathbf{D}_H = \text{diag}(\max(\mathbf{h}_1), \dots, \max(\mathbf{h}_K))$

STEP 4: Stop criterion

$p \leftarrow p + 1$

if $p \geq Nb_{iter}$ **or**

$$\|\mathbf{X} - \Phi(\mathbf{W}\mathbf{H}^T)\|^2 + \sum_{i,k} (\mathbf{W}_{ik}^2 - \mathbf{W}_{ik})^2 + \sum_{j,k} (\mathbf{H}_{jk}^2 - \mathbf{H}_{jk})^2 < \varepsilon$$

then

break

else

return to **STEP 2**

end if

In practice, for a faster convergence, \mathbf{W} and \mathbf{H} are initialized with the result of the original NMF algorithm [8]. In step 3, \mathbf{W} and \mathbf{H} are normalized in order to have the values of \mathbf{W}_{ij} and \mathbf{H}_{ij} within the interval $[0, 1]$. For space reasons, the details of the update rule derivation in step 2 are not given in this paper.

However, the inverse problem (11) is ill-posed in general. In others words, the post-nonlinear mixture model is not always identifiable. To regularize it, additional constraint must be added. Here, we choose to maximize the support of the rank-one terms ($\mathbf{X}_k = \mathbf{w}_k \odot \mathbf{h}_k^T$) by solving the following problem:

$$\{\mathbf{W}, \mathbf{H}\} = \underset{\mathbf{W}, \mathbf{H} \in \{0,1\}}{\operatorname{argmin}} \|\mathbf{X} - \Phi(\mathbf{W}\mathbf{H}^T)\|_2^2$$

$$s.t \min \left(\frac{1}{\sum_{k=1}^K (\mathbf{w}_k \odot \mathbf{h}_k^T)} \right). \quad (13)$$

We define the following cost function for the inverse problem (13):

$$L(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \sum_{i,j} (\mathbf{X}_{ij} - \Phi(\mathbf{W}\mathbf{H}^T)_{ij})^2 + \frac{1}{2} \lambda \sum_{j,k} (\mathbf{H}_{jk}^2 - \mathbf{H}_{jk})^2 + \frac{1}{2} \lambda \sum_{i,k} (\mathbf{W}_{ik}^2 - \mathbf{W}_{ik})^2 + \lambda_1 \frac{1}{\sum_k \left(\sum_{i,j} \mathbf{W}_{ik} \mathbf{H}_{jk} \right)}. \quad (14)$$

The algorithm that solve (14), called Constraint Post Non-Linear Penalty Function algorithm (C-PNL-PF), is similar to the PNL-PF algorithm except for the update rules step. Thus, in C-PNL-PF, the update rule of \mathbf{W} and \mathbf{H} are given by:

$$\mathbf{H}_{jk} \leftarrow \mathbf{H}_{jk} \frac{\left(\mathbf{X} \frac{\partial}{\partial \mathbf{H}_{jk}} \Phi(\mathbf{W}\mathbf{H}^T) \right)_{jk} + 3\lambda \mathbf{H}_{jk}^2 + \lambda_1 \frac{\sum_k \left(\sum_i \mathbf{w}_{ik} \right)}{\left(\sum_k \left(\sum_{i,j} \mathbf{w}_{ik} \mathbf{H}_{jk} \right) \right)^2}}{\left(\frac{\partial}{\partial \mathbf{H}_{jk}} \Phi(\mathbf{W}\mathbf{H}^T) \right)_{jk} + 2\lambda \mathbf{H}_{jk}^3 + \lambda \mathbf{H}_{jk}}$$

$$\mathbf{W}_{ik} \leftarrow \mathbf{W}_{ik} \frac{\left(\mathbf{X} \frac{\partial}{\partial \mathbf{W}_{ik}} \Phi(\mathbf{W}\mathbf{H}^T) \right)_{ik} + 3\lambda \mathbf{W}_{ik}^2 + \lambda_1 \frac{\sum_k \left(\sum_j \mathbf{h}_{kj} \right)}{\left(\sum_k \left(\sum_{i,j} \mathbf{w}_{ik} \mathbf{H}_{jk} \right) \right)^2}}{\left(\frac{\partial}{\partial \mathbf{W}_{ik}} \Phi(\mathbf{W}\mathbf{H}^T) \right)_{ik} + 2\lambda \mathbf{W}_{ik}^3 + \lambda \mathbf{W}_{ik}}$$

In the next section, the different models and algorithms presented in this paper are compared in numerical simulations.

V. NUMERICAL SIMULATIONS

In this section, we compare the two BMF algorithms presented in [15] based on linear mixture model (3) (PF and TH) and the two proposed algorithms based on the post nonlinear model (10) (PNL-PF and C-PNL-PF).

A first experiment compares the different algorithms in the case of $K=2$ uncorrelated sources, *i.e.*, sources having disjoint supports. One can observe that, all algorithms estimate correctly \mathbf{X} , \mathbf{W} and \mathbf{H} (figure 1), as in this case, models (1), (3) and (10) are equivalent. In the second experiment, we have two sources strongly correlated in the identifiable case *i.e.*, each source verify theorem III.1. We can see that, the TH algorithm does not estimate correctly the matrices, the PF algorithm estimates well \mathbf{W} and \mathbf{H} but the reconstructed \mathbf{X} is not accurate. This is because the direct model (3) used in these algorithms is not equivalent to the original BMF model (1) when the sources are correlated. On the other hand, the two proposed algorithms estimate well \mathbf{X} , \mathbf{W} and \mathbf{H} (figure 2). The third experiment compares the PF, PNL-PF and C-PNL-PF algorithms in the case of 3 correlated sources. This mixture is not identifiable because the second source ($\mathbf{X}_2 = \mathbf{w}_2 \odot \mathbf{h}_2^T$) do not verify the condition of theorem III.1. We can observe that, the PF algorithm fails completely. The PNL-PF algorithm estimates well \mathbf{X} and \mathbf{H} but not \mathbf{W} (figure 3). Actually, as the model is not identifiable, the PNL-PF algorithm finds another admissible solution. The C-PNL-PF algorithm allows to find the right solution of \mathbf{H} , \mathbf{W} and \mathbf{X} .

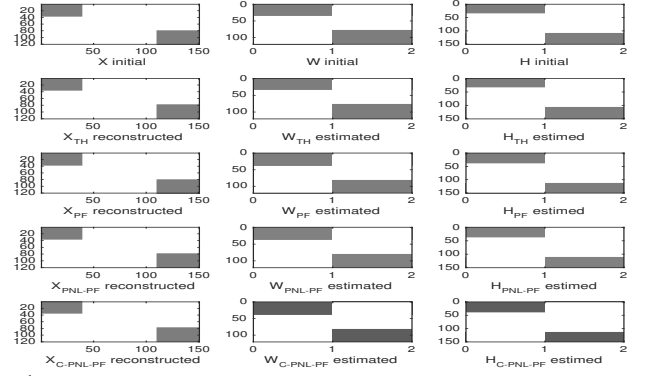


Figure 1: Comparison of TH, PF, PNL-PF and C-PNL-PF algorithms for the case of two uncorrelated sources ($N=120$, $M=150$, $\gamma = 150$, $\lambda = 800$, $\lambda_1 = 10^7$)

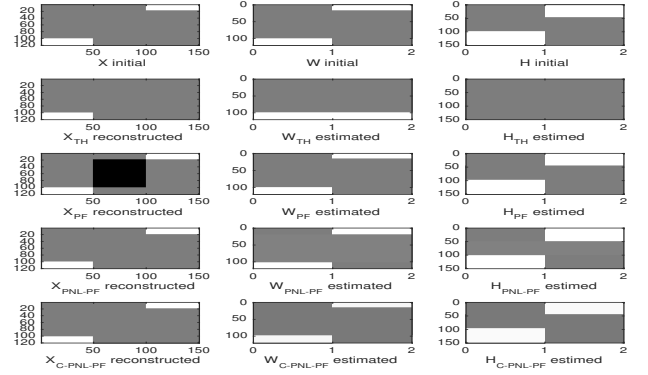


Figure 2: Comparison of TH, PF, PNL-PF and C-PNL-PF algorithms for the case of two correlated sources ($N=120$, $M=150$, $\gamma = 150$, $\lambda = 800$, $\lambda_1 = 10^7$)

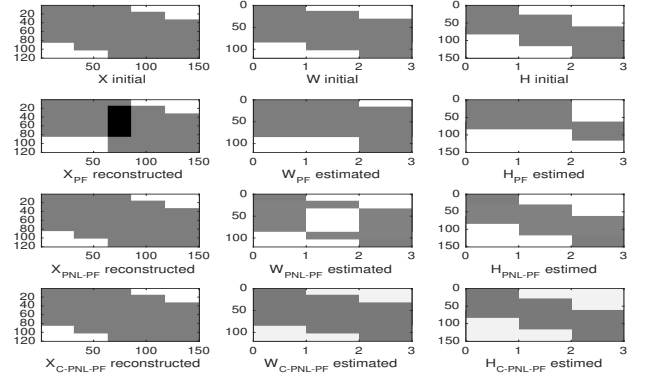


Figure 3: Comparison of PF, PNL-PF and C-PNL-PF algorithms for the case of 3 correlated sources ($N=120$, $M=150$, $\gamma = 150$, $\lambda = 800$, $\lambda_1 = 2 \cdot 10^8$)

The second part of this section aims at evaluating the statistical performance of these algorithms. We consider a rank 5 BMF model where each column of \mathbf{W} and \mathbf{H} follows a Bernoulli distribution of parameter b (%) which represents the probability of non-zero elements. The quality of estimation is assessed by the error rate in the estimation of the columns of \mathbf{W} and \mathbf{H} (mean, [min, max]) obtained by averaging the result over 100 trials of the Bernoulli random variables.

When the Bernoulli parameter is low, the sources correlation (overlapping) is low. When b increases, the sources correlation increases.

Results reported in table I, correspond to $b = 30\%$ which result in binary sources weakly correlated. In that case, all the algorithms yield a low error rate ($< 1\%$) but only PNL-PF and C-PNL-PF yield an error rate equal to 0% which shows that in that case, the identifiability condition was verified. This is no longer true when b increases; in that case only C-PNL-PF gives low error rate (table II).

	X	W	H
TH	0.99 [0.48 1.6]	0.3 [0 1.0]	0.45 [0 1.2]
PF	0.13 [0 0.68]	0.13 [0 0.66]	0 [0 0]
PNL-PF	0 [0 0]	0 [0 0]	0 [0 0]
C-PNL-PF	0 [0 0]	0 [0 0]	0 [0 0]

Table I: The mean, minimum and maximum of the estimation error rate of **W** and **H** columns and **X** sources for $b = 30\%$ ($N=120$, $M=150$, $\gamma = 170$, $\lambda = 800$, $\lambda_1 = 100$)

	X	W	H
TH	20.7 [3.4 37.2]	5.65 [0.4 10.8]	3.64 [0.1 8.3]
PF	19.3 [6.7 36.2]	5.74 [0.9 11.4]	3.2 0[.04 8.3]
PNL-PF	12.6 [1.1 32.2]	3.40 [0.05 9.1]	2.1 [0.0 7.3]
C-PNL-PF	4.3[0.1 13.7]	1.86 [0.9 3.9]	1.5 [0.8 3.5]

Table II: The mean, minimum and maximum of the estimation error rate of **W** and **H** columns and **X** sources for $b = 55\%$ ($N=120$, $M=150$, $\gamma = 170$, $\lambda = 800$, $\lambda_1 = 10^8$)

VI. CONCLUSION

In this paper we studied the identifiability of the binary mixture model and provided a necessary and sufficient identifiability condition. We also introduced a novel post nonlinear mixture model which is equivalent to the binary mixture model when the matrices are strictly binary. Based on this model, two algorithms for binary matrix factorization have been proposed and their performance has been compared in simulations to two state of the art methods. Our approach gives more accurate estimates of the binary matrices, especially when the columns of the matrices are highly correlated.

Future work will study the robustness of the proposed algorithms to binary noise and how to optimally choose the value of the hyper parameters.

REFERENCES

- [1] M. Koyutürk, A. Grama, and N. Ramakrishnan, "Algebraic techniques for analysis of large discrete-valued datasets," in *European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 311–324, Springer, 2002.
- [2] M. Koyutürk, A. Grama, and N. Ramakrishnan, "Compression, clustering, and pattern discovery in very high-dimensional discrete-attribute data sets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 447–461, 2005.
- [3] Z.-Y. Zhang, T. Li, C. Ding, X.-W. Ren, and X.-S. Zhang, "Binary matrix factorization for analyzing gene expression data," *Data Mining and Knowledge Discovery*, vol. 20, no. 1, pp. 28–52, 2010.
- [4] T. Li, "A general model for clustering binary data," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pp. 188–197, ACM, 2005.
- [5] E. Meeds, Z. Ghahramani, R. M. Neal, and S. T. Roweis, "Modeling dyadic data with binary latent factors," in *Advances in neural information processing systems*, pp. 977–984, 2006.
- [6] B.-H. Shen, S. Ji, and J. Ye, "Mining discrete patterns via binary matrix factorization," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 757–766, ACM, 2009.
- [7] E. Nenova, D. I. Ignatov, and A. V. Konstantinov, "An fca-based boolean matrix factorisation for collaborative filtering," *arXiv preprint arXiv:1310.4366*, 2013.
- [8] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [9] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, pp. 556–562, 2001.
- [10] P. Miettinen, T. Mielikainen, A. Gionis, G. Das, and H. Mannila, "The discrete basis problem," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 20, no. 10, pp. 1348–1362, 2008.
- [11] S. Tu, R. Chen, and L. Xu, "A binary matrix factorization algorithm for protein complex prediction," *Proteome Science*, vol. 9, no. 1, p. 1, 2011.
- [12] P. Jiang and M. Heath, "Mining discrete patterns via binary matrix factorization," in *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on*, pp. 1129–1136, IEEE, 2013.
- [13] R. Belohlavek, C. Glodeanu, and V. Vychodil, "Optimal factorization of three-way binary data using triadic concepts," *Order*, vol. 30, no. 2, pp. 437–454, 2013.
- [14] A. Yeredor, "Ica in boolean xor mixtures," in *Independent Component Analysis and Signal Separation*, pp. 827–835, Springer, 2007.
- [15] Z. Zhang, C. Ding, T. Li, and X. Zhang, "Binary matrix factorization with applications," in *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pp. 391–400, IEEE, 2007.
- [16] R. D. Luce, "A note on boolean matrix theory," *Proceedings of the American Mathematical Society*, vol. 3, no. 3, pp. 382–388, 1952.
- [17] R. Belohlavek and V. Vychodil, "Discovery of optimal factors in binary data via a novel method of matrix decomposition," *Journal of Computer and System Sciences*, vol. 76, no. 1, pp. 3–20, 2010.