

# Supplementary material for Orthogonal Parametric Non-negative Matrix Tri-Factorization with $\alpha$ -Divergence for Co-clustering

Saeid Hoseinipour, Mina Aminghafari , Adel Mohammadpour

*Department of Mathematics and Computer Science, Amirkabir University of Technology  
(Tehran Polytechnic), Iran.*

---

---

## 1. Introduction

In this supplementary material, we present a couple of complementary elements which were omitted in the main paper due to limitation the length of paper and redundancy of them. First, we present synthetic datasets for Bernoulli, Poisson, and Truncated Gaussian, and then implement OPNMTF on these. Second, we show more visualization of some datasets.

## 2. Synthetic data

We propose to use a Latent Block Model (LBM) ([Govaert and Nadif, 2013](#)) for generating a non-negative data matrix. This generative model considers  $\mathcal{F}$  and  $\mathcal{G}$  as the sets of possible labels  $\mathbf{F}$  (rows) and  $\mathbf{G}$  (columns), also  $\varphi(X_{ij}; \delta_{kh})$  denoted distribution of the values  $x_{ij}$  for co-cluster (blocks)  $kh$ . The total probability density function is as follows

$$f(\mathbf{X}; \boldsymbol{\theta}) = \sum_{(\mathbf{F}, \mathbf{G}) \in \mathcal{F} \times \mathcal{G}} \prod_{i,k} \pi_k^{F_{ik}} \prod_{j,h} \rho_h^{G_{jh}} \prod_{i,j,k,h} \varphi(X_{ij}; \delta_{kh}), \quad (1)$$

where  $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\rho}, \boldsymbol{\delta})$ ,  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_g)$  and  $\boldsymbol{\rho} = (\rho_1, \dots, \rho_s)$ , also  $\boldsymbol{\delta} = (\delta_{11}, \dots, \delta_{gs})$  are all parameters co-clusters that are defined according to the probability density function.

- **Bernoulli LBM:** The values  $x_{ij}$  are distributed according to a Bernoulli distribution  $\mathcal{B}(\gamma_{kh})$  with  $\gamma_{kh} \in [0, 1]$ , therefore

$$\varphi(x_{ij}; \gamma_{kh}) = \gamma_{kh}^{x_{ij}} (1 - \gamma_{kh})^{1-x_{ij}}. \quad (2)$$

- **Poisson LBM:** The values  $x_{ij}$  are distributed according to a Poisson distribution  $\mathcal{P}(\omega_i \nu_j \alpha_{kh})$  where  $\omega_i$  and  $\nu_j$  are the effects of row  $i$  and column  $j$ , respectively. Also,  $\alpha_{kh}$  is the effect of co-cluster  $kh$  for  $x_{ij}$  that belong to the  $k$ -th row cluster and  $h$ -th column cluster and

$$\varphi(x_{ij}; \alpha_{kh}) = \frac{e^{-\omega_i \nu_j \alpha_{kh}} (\omega_i \nu_j \alpha_{kh})^{x_{ij}}}{x_{ij}!}. \quad (3)$$

- **Truncated Gaussian LBM:** The values  $x_{ij}$  belong to co-cluster  $kh$  are distributed according to a Truncated Gaussian distribution  $\mathcal{TG}(\mu_{kh}, \sigma_{kh})$  with  $a \leq x_{ij} \leq b$ , mean  $\mu_{kh}$ , and variance  $\sigma_{kh}$ , therefore

$$\varphi(x_{ij}; \mu_{kh}, \sigma_{kh}, a, b) = \frac{1}{\sigma_{kh}} \frac{\phi\left(\frac{x_{ij} - \mu_{kh}}{\sigma_{kh}}\right)}{\Phi\left(\frac{b - \mu_{kh}}{\sigma_{kh}}\right) - \Phi\left(\frac{a - \mu_{kh}}{\sigma_{kh}}\right)}, \quad (4)$$

where  $\phi$  and  $\Phi$  are standard Gaussian distribution and cumulative distribution functions, respectively.

Our motivation is to implement the introduced algorithms on the data matrix with a known distribution of  $X_{ij}$ . Due to having row and column true labels, we use a two-way accuracy evaluation criterion. Algorithm 1

---

**Algorithm 1** Simulating data of the Latent Block Model

---

**Input:**  $n, m, g, s, \boldsymbol{\pi}, \boldsymbol{\rho}$

- 1 Simulate  $\mathbf{F}$  according to a Multinomial distribution with parameters  $\boldsymbol{\pi} = (1, \pi_1, \dots, \pi_g)$ .
- 2 Simulate  $\mathbf{G}$  according to a Multinomial distribution with parameters  $\boldsymbol{\rho} = (1, \rho_1, \dots, \rho_s)$ .
- 3 Simulate each co-cluster  $\mathbf{X}_{kh}$  according to Bernoulli (2), Poisson (3), and Truncated Gaussian (4) distributions.
- 4 Permutation all  $\mathbf{X}_{kh}$ .

**Output:** Data matrix  $\mathbf{X}$  size of  $n \times m$

---

present process simulation data of LBM, also [NMTFcoclust](#) as repository GitHub is available.

### 2.0.1. Setting

Table 1 shows the parameters that we need to generate a non-negative data matrix. Figure 1 illustrates three types of simulation data matrices. Figures 2 and 3 shows row and column clustering comparison with difference tuple parameter  $(\lambda, \mu) \in \{(10^{-9}, 10^{-9}), (10^{-9}, 10^{+9}), (10^{+9}, 10^{-9}), (10^{+9}, 10^{+9})\}$  and  $\alpha \in \{0, 0.5, 1\}$ . We conducted experiments on synthetic and real-world datasets to validate the effectiveness of our proposed algorithm.

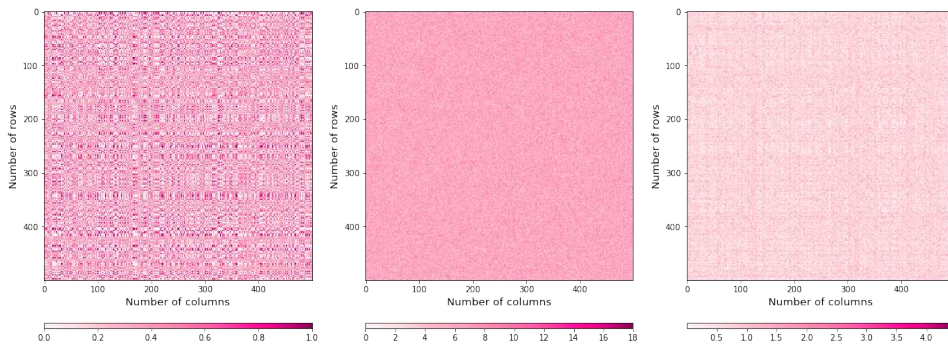


Figure 1: (Left) Data matrix **Bernoulli LBM**. (Middle) Data matrix **Poisson LBM**. (Right) Data matrix **Truncated Gaussian LBM**.

Table 1: Parameters required to generate synthetic data from LBM.

Inputs and Parameters	Setting
$(n, m)$	(500, 500)
$(g, s)$	(3, 3)
LBM	$\pi = (1/3, 1/3, 1/3)$ $\rho = (1/3, 1/3, 1/3)$
Bernoulli LBM	$\gamma = \begin{pmatrix} 0.9 & 0.01 & 0.01 \\ 0.01 & 0.9 & 0.01 \\ 0.01 & 0.01 & 0.9 \end{pmatrix}$
Poisson LBM	$\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ $\omega = (2, 2, 6)$ $\nu = (7, 2, 1)$
Truncated Gaussian LBM	$\mu = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \sigma^2 = \begin{pmatrix} 10 & 1 & 1 \\ 1 & 100 & 1 \\ 1 & 1 & 1000 \end{pmatrix}$

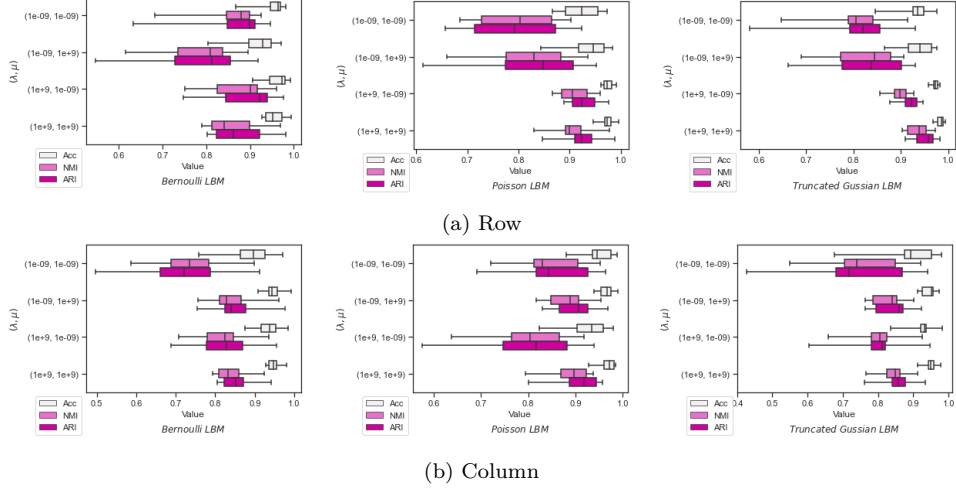


Figure 2: Results on three synthetic data (**Bernoulli, Poisson, Truncated Gaussian**) according to OPNMTF with  $\alpha = 0.5$  and  $(\lambda, \mu) \in \{(10^{-9}, 10^{-9}), (10^{-9}, 10^{+9}), (10^{+9}, 10^{-9}), (10^{+9}, 10^{+9})\}$ .

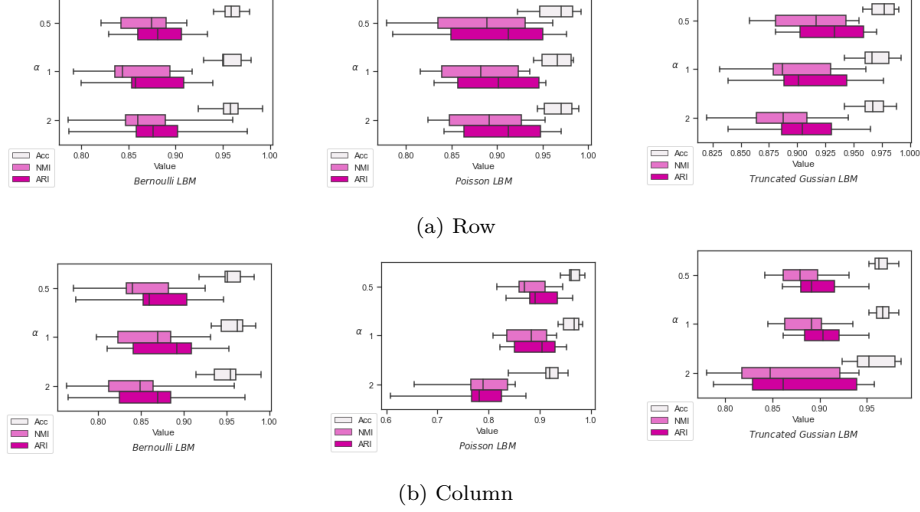


Figure 3: Box plots of measures (Acc, NMI, ARI) for three synthetic data (**B**ernoulli, **P**oisson, **T**runcated **G**aussian) based on 100 iterations of OPNMTF with  $\lambda = 0.03$ ,  $\mu = 1$ , and  $\alpha \in \{0.5, 1\}$ .

### 3. Visualization

We show a triple reorganization plot based on OPNMTF on **WebACE** dataset in Figure 4, also Figures 5 and 6 show different numbers of column clusters for **CSTR** and **Classic3**, respectively. We focus on **Classic3** with PCA in 7, as well as the co-clustering reorganization plot near by word clouds  $3 \times 3$  in Figure 8.

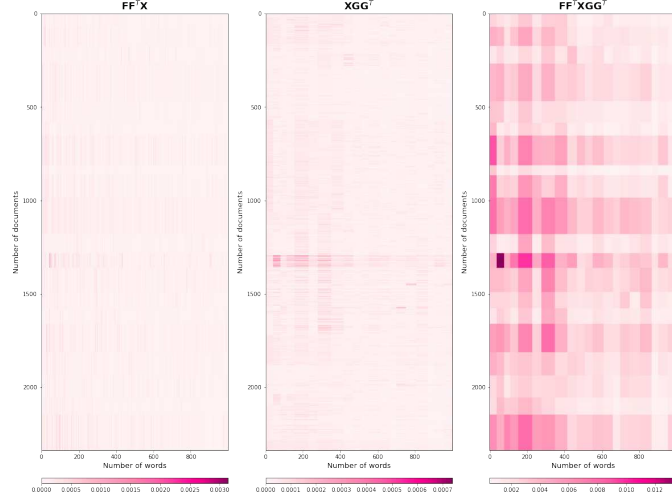


Figure 4: Result by OPNMTF with  $\alpha = 1$ ,  $\lambda = 0.03$ ,  $\mu = 1$ ,  $g = 20$ , and  $s = 10$  on **WebACE**, (Left) Clustering row data matrix ( $\mathbf{FF}^T \mathbf{X}$ ). (Middle) Clustering column data matrix ( $\mathbf{XGG}^T$ ). (Right) Co-clustering reorganization data matrix ( $\mathbf{FF}^T \mathbf{XGG}^T$ ).

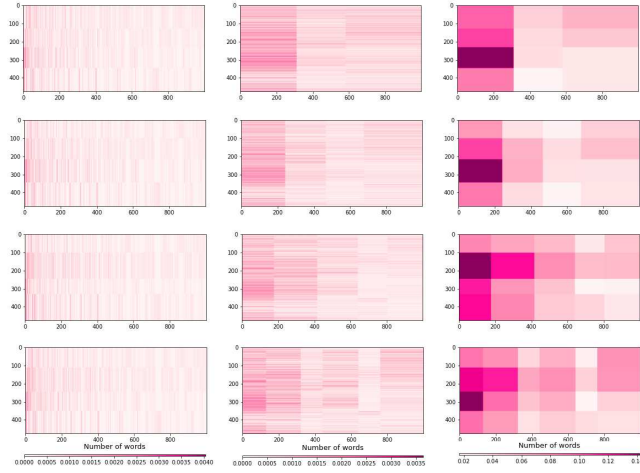


Figure 5: Result by OPNMTF with  $\alpha = 0.4$ ,  $\lambda = 0.4$ ,  $\mu = 0.6$ , and  $g = 4$  on **CSTR**, (Left) Clustering rows data matrix ( $\mathbf{FF}^T \mathbf{X}$ ). (Middle) Clustering columns data matrix ( $\mathbf{XGG}^T$ ) with  $s = 3, 4, 5, 6$ . (Right) Co-clustering reorganization data matrix, ( $\mathbf{FF}^T \mathbf{XGG}^T$ ).

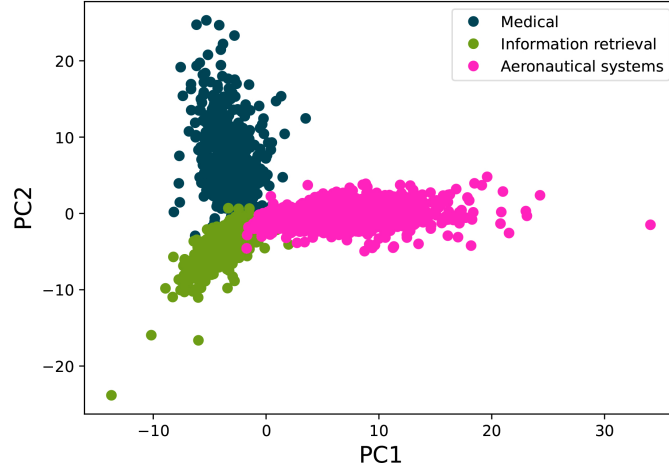


Figure 7: Principal Components Analysis (PCA) on **Classic3** for two components.

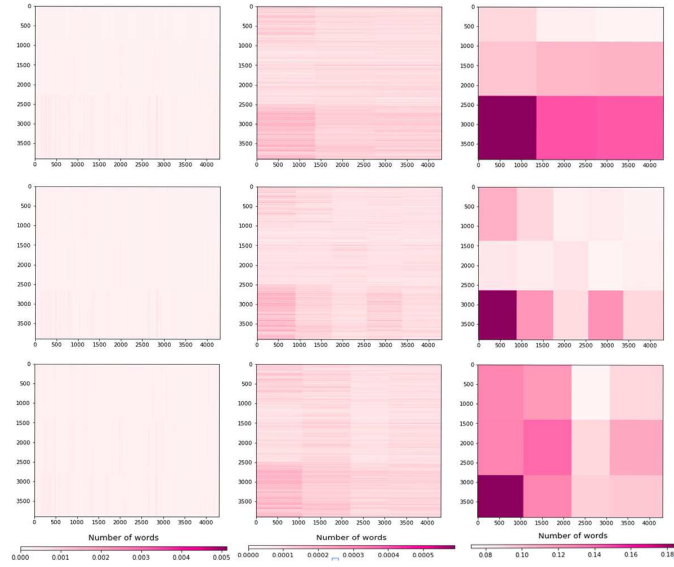


Figure 6: Result by OPNMTF with  $\alpha = 0.4$ ,  $\lambda = 0.03$ ,  $\mu = 1$ , and  $g = 3$  on **Classic3**, (Left) Clustering row data matrix ( $\mathbf{F}\mathbf{F}^T\mathbf{X}$ ). (Middle) Clustering column data matrix ( $\mathbf{X}\mathbf{G}\mathbf{G}^T$ ) with  $s = 3, 4, 5$ . (Right) Co-clustering reorganization data matrix ( $\mathbf{F}\mathbf{F}^T\mathbf{X}\mathbf{G}\mathbf{G}^T$ ).





