*JOURNAL OF COMPUTATIONAL BIOLOGY AND MEDICINE*
***Research Article | Volume 5 | Issue 5 | Jan 2025***
*Received: 25 Dec 2024 | Revised: 28 Dec 2024*
*Accepted: 03 Jan 2025 | Published Online: 21 Jan 2025*

# Nonnegative Matrix Factorization with Fast Bayesian Inference

**Lucas Pereira[1], Mariana Santos[2] and Rafael Oliveira[3,*]**

[1] Department of Computational Mathematics, Universidade Federal do Espírito Santo, Serra, 29175-000, Brazil

[2] Institute of Advanced Computing, Universidade Federal de Alagoas, Arapiraca, 57309-005, Brazil

[3] Center for Applied Mathematics and Computational Modeling, Universidade Federal de Sergipe, Itabaiana, 49500-000, Brazil

[*]Corresponding Author, Email: rafael.oliveira@ufs.br

**Abstract:** Nonnegative Matrix Factorization (NMF) has been widely employed in various fields for its ability to extract meaningful latent features from high-dimensional data. However, the existing NMF algorithms often suffer from computational inefficiency, limiting their applicability to large-scale datasets. In light of this, this paper addresses the pressing need for a faster and more efficient Bayesian inference method for NMF. Despite the significant research efforts dedicated to improving NMF algorithms, challenges remain in achieving both speed and accuracy. To fill this gap, we propose a novel approach that combines the benefits of nonnegative constraints and Bayesian inference, leading to a Fast Bayesian Inference method for NMF. Our method not only accelerates the computation process but also enhances the interpretability of the extracted features. Through comprehensive experiments on diverse datasets, we demonstrate the superior performance of our proposed method in terms of both speed and accuracy, highlighting its potential for widespread applications in data analysis and pattern recognition.

**Keywords:** *Nonnegative Matrix Factorization; Bayesian Inference; Computational Efficiency; Latent Feature Extraction; Data Analysis*

## 1. Introduction

Nonnegative Matrix Factorization (NMF) is a powerful dimensionality reduction technique widely used in various fields such as machine learning, data mining, and computer vision. It aims to factorize a given nonnegative matrix into two lower-rank nonnegative matrices, representing a compressed and interpretable version of the original data. However, NMF faces challenges such as sensitivity to initialization, local minima, and scalability issues when dealing with large-scale datasets. The presence of noise and outliers in the data can also impact the effectiveness of NMF algorithms. Addressing these bottlenecks in NMF research will require innovative algorithm design, improved optimization techniques, and robust regularization methods to enhance the performance and applicability of NMF in real-world applications.

To this end, research on Nonnegative Matrix Factorization (NMF) has advanced to encompass a wide range of applications across diverse fields, including image and video processing, text mining, bioinformatics, and recommendation systems. Scholars are exploring innovative algorithms and improving NMF's performance for more complex and large-scale datasets. A comprehensive literature review on nonnegative matrix factorization (NMF) techniques for graph clustering and community detection reveals significant advancements in the field. Cai et al. [1] introduced Graph Regularized NMF (GNMF) to capture geometric structures in data representations. Berahmand et al. [2] extended this concept to attributed networks with Augment Graph Regularization NMF (AGNMF-AN) for more accurate community detection. Liu et al. [3] proposed High-Order Proximity-incorporated NMF (HOP-NMF) for community detection by considering high-order proximity in networks. Li et al. [4] devised a Provable Splitting Approach for Symmetric NMF, improving efficient algorithms for this specific problem. Lin [5] presented Projected Gradient Methods for NMF, demonstrating faster convergence compared to traditional methods. Wang et al. [6] and Hajiveiseh et al. [7] utilized NMF techniques for heart-lung sound separation and graph clustering, respectively. Nasiri et al. [8] developed Robust Graph Regularization NMF for link prediction in attributed networks. He et al. [9] conducted a thorough survey on NMF-based community detection in complex networks, categorizing approaches and proposing future research directions. A comprehensive examination of nonnegative matrix factorization (NMF) methods for graph clustering and community detection demonstrates notable progress in the field. Fast Bayesian Inference techniques, such as Graph Regularized NMF (GNMF) and Augment Graph Regularization NMF (AGNMF-AN), offer efficient ways to capture geometric structures and improve accuracy in community detection. The utilization of these techniques is crucial due to their ability to handle large-scale data sets and complex network structures effectively.

Specifically, Fast Bayesian Inference has been utilized to enhance the efficiency of Nonnegative Matrix Factorization by providing a rigorous probabilistic framework for modeling the underlying data structure. This approach allows for accurate estimation of latent factors and facilitates the interpretation of complex relationships within the data. A literature review on fast Bayesian inference methods reveals several recent advancements in statistical modeling. Zhou et al. (2024) introduced a reparameterized gamma process with random effects for superior modeling of degradation processes [10]. Quiroz et al. (2023) proposed fast Bayesian inference of block Nearest Neighbor Gaussian models suitable for large data sets [11]. Adachi et al. (2022) developed a parallelized Bayesian quadrature method with exponential convergence rate for efficient numerical integration [12]. Gaedke-Merzhäuser et al. (2022) presented parallelized integrated nested Laplace approximations for rapid Bayesian inference [13]. Gu et al. (2022) introduced

GIGA-Lens, a fast Bayesian framework for strong gravitational lens modeling [14]. Teimouri (2022) focused on fast Bayesian inference for the Birnbaum-Saunders distribution [15]. Gressani and Lambert (2021) utilized Laplace approximations for speedy Bayesian inference in generalized additive models [16]. Monnahan and Kristensen (2018) introduced adnuts and tmbstan packages for fast Bayesian inference in ADMB and TMB software platforms [17]. Kim (2021) proposed a path integral formulation for fast Bayesian inference for Gaussian Cox processes [18]. Diana et al. (2021) developed a unified framework for rapid Bayesian inference in occupancy models for large datasets [19]. However, current limitations in fast Bayesian inference methods include the need for further research on scalability to extremely large datasets, exploration of applicability across diverse types of models, and validation of reliability and accuracy in practical applications.

To overcome those limitations, this paper aims to address the need for a faster and more efficient Bayesian inference method for Nonnegative Matrix Factorization (NMF). Given the computational inefficiency of existing NMF algorithms when dealing with large-scale datasets, the paper proposes a novel approach that integrates nonnegative constraints and Bayesian inference to develop a Fast Bayesian Inference method for NMF. By leveraging the benefits of both techniques, this method not only accelerates the computation process but also enhances the interpretability of the extracted features. The key detail of this approach lies in its ability to balance speed and accuracy, thus overcoming the challenges previously faced by traditional NMF algorithms. Through a series of comprehensive experiments conducted on diverse datasets, the paper demonstrates the superior performance of the proposed method in terms of both speed and accuracy. This research showcases the potential of the Fast Bayesian Inference method for NMF to have broad applications in data analysis and pattern recognition, offering a promising solution to the limitations currently plaguing NMF algorithms.

Section 2 presents the problem of inefficient computational performance in existing Nonnegative Matrix Factorization (NMF) algorithms when applied to large-scale datasets. Section 3 introduces a novel approach that combines nonnegative constraints and Bayesian inference to create a Fast Bayesian Inference method for NMF. In Section 4, a case study demonstrates the effectiveness of the proposed method. Section 5 analyzes the results of comprehensive experiments on various datasets, showing superior performance in terms of both speed and accuracy. Section 6 engages in a discussion of the findings and the implications for data analysis and pattern recognition. Finally, in Section 7, a summary emphasizes the potential of the Fast Bayesian Inference method for NMF in addressing the challenges of computational inefficiency and enhancing feature interpretability, paving the way for its extensive application across diverse fields.

## 2. Background

### *2.1 Nonnegative Matrix Factorization*

Nonnegative Matrix Factorization (NMF) is a powerful tool in the domain of machine learning and data analysis, particularly suited for high-dimensional datasets. It is a matrix factorization technique where, given a nonnegative matrix $V$, the goal is to approximate it as a product of two nonnegative matrices with lower rank. This decomposition is particularly useful in extracting latent features from data, widely used in areas like image processing, text mining, and bioinformatics.

The essential composition in NMF is decomposing the matrix $V \in \mathbb{R}_{\geq 0}^{m\times n}$ into two matrices: $W \in \mathbb{R}_{\geq 0}^{m\times r}$ and $H \in \mathbb{R}_{\geq 0}^{r\times n}$ , such that:

$$V \approx WH \tag{1}$$

where $r$ is typically much smaller than either $m$ or $n$ , allowing for a reduced dimensional representation. Each column of $W$ can be considered a basis vector, and the columns of $H$ can be viewed as the coefficients mapping the basis vectors to the columns of $V$ . This transformation captures the underlying structure or features in a way that each component or feature in $W$ and $H$ is easily interpretable due to the nonnegativity constraints.

To find matrices $W$ and $H$ , NMF attempts to minimize the difference (or loss) between $V$ and the product $WH$ . An often-used measure for this purpose is the Frobenius norm, which leads to the optimization problem:

$$\min_{W,H\geq 0} ||V - WH||_F^2 \tag{2}$$

Expanding on the optimization goal, if we denote the elements of the matrices as $v_{ij}$ , $w_{ik}$ , and $h_{kj}$ , the objective becomes:

$$\min_{W,H\geq 0} \sum_{i=1}^{m}\sum_{j=1}^{n}\left(v_{ij} - \sum_{k=1}^{r} w_{ik}h_{kj}\right)^2 \tag{3}$$

Alternatively, another frequently employed cost function is based on the Kullback-Leibler divergence:

$$\min_{W,H\geq 0} \sum_{i=1}^{m}\sum_{j=1}^{n}\left(v_{ij}\log\frac{v_{ij}}{(WH)_{ij}} - v_{ij} + (WH)_{ij}\right) \tag{4}$$

The optimization process for NMF is typically carried out using iterative update rules, which ensure convergence while maintaining nonnegative constraints. The update rules are derived from techniques such as the multiplicative update algorithm, whose basic form is:

$$w_{ik} \leftarrow w_{ik}\frac{\sum_{j=1}^{n} v_{ij}h_{kj}/(WH)_{ij}}{\sum_{j=1}^{n} h_{kj}} \tag{5}$$

$$h_{kj} \leftarrow h_{kj}\frac{\sum_{i=1}^{m} w_{ik}v_{ij}/(WH)_{ij}}{\sum_{i=1}^{m} w_{ik}} \tag{6}$$

These updates are performed iteratively until convergence, typically determined by the change in the objective function dropping below a preset threshold or reaching a maximum number of iterations.

The nonnegativity constraint of NMF is significant because it often results in a parts-based representation of the data. This stems from the fact that nonnegative combinations of nonnegative bases tend to be additive rather than subtractive, leading to more interpretable factors that make NMF particularly powerful for applications like image processing where interpretability is key.

In conclusion, Nonnegative Matrix Factorization is a vital tool for analyzing and extracting meaningful structures from high-dimensional data sets. Its mathematical formalism, based on nonnegative constraints and intuitive outputs, spans across various practical applications, providing insights that are both clear and interpretable.

*2.2 Methodologies & Limitations*

Nonnegative Matrix Factorization (NMF) remains at the forefront of data analysis as an effective method for dimensionality reduction and latent feature extraction. Recent methodologies in NMF focus on refinement of the basic factorization framework, exploiting the threshold between effectiveness and interpretability. Commonly used approaches to NMF involve variations in objective functions, optimization techniques, and constraints aimed at enhancing performance, scalability, and applicability to diverse datasets.

The classic NMF objective is decomposing $V \in \mathbb{R}_{\geq 0}^{m \times n}$ into $W \in \mathbb{R}_{\geq 0}^{m \times r}$ and $H \in \mathbb{R}_{\geq 0}^{r \times n}$ , minimizing the Frobenius norm:

$$\min_{W,H \geq 0} ||V - WH||_F^2 \tag{7}$$

Beyond the Frobenius norm, alternative objective functions like the Kullback-Leibler divergence are employed for probabilistic interpretation:

$$\min_{W,H \geq 0} \sum_{i=1}^{m} \sum_{j=1}^{n} \left( v_{ij} \log \frac{v_{ij}}{(WH)_{ij}} - v_{ij} + (WH)_{ij} \right) \tag{8}$$

These objectives necessitate iterative update rules, for example, the multiplicative update rule preserving nonnegativity:

$$w_{ik} \leftarrow w_{ik} \frac{\sum_{j=1}^{n} \frac{v_{ij} h_{kj}}{(WH)_{ij}}}{\sum_{j=1}^{n} h_{kj}} \tag{9}$$

$$h_{kj} \leftarrow h_{kj} \frac{\sum_{i=1}^{m} \frac{v_{ij} w_{ik}}{(WH)_{ij}}}{\sum_{i=1}^{m} w_{ik}} \tag{10}$$

Recent advancements in NMF tackle its limitations by incorporating additional constraints or regularization terms to address overfitting and enhance robustness. For example, sparsity constraints are introduced using L1 regularization on $W$ or $H$ :

$$\min_{W,H\geq 0}||V - WH||_F^2 + \lambda(||W||_1 + ||H||_1) \tag{11}$$

This forces many entries of $W$ and $H$ to be zero, promoting a sparser and potentially more interpretable model. Furthermore, incorporating local smoothness constraints on $H$, such as using second derivatives, can preserve structural continuity:

$$\min_{W,H\geq 0}||V - WH||_F^2 + \beta \sum_{j=1}^{n} ||H[:,j] - \text{smooth}(H[:,j])||_2^2 \tag{12}$$

Here, $\text{smooth}(H[:,j])$ represents a smooth version of column $j$ in $H$. Another area of exploration involves integrating orthogonality to prevent similar basis vectors:

$$\min_{W,H\geq 0}\left||V - WH||_F^2 + \gamma\right||HH^T - I||_F^2 \tag{13}$$

This constraint ensures that $H$'s columns describe distinct features, enriching interpretability. However, despite these innovations, NMF faces inherent issues. The convergence rate can be slow, and solutions are often not unique, leading to different decompositions under varying runs that still satisfy the approximation. Moreover, the model is susceptible to local minima, prompting an ongoing pursuit for resilient initialization and optimization strategies.

Ultimately, while Nonnegative Matrix Factorization is an invaluable tool in extracting meaningful features through its simplified representation, further advancement is essential to mitigate its limitations and deploy it robustly across a broad spectrum of complex, real-world datasets.

## 3. The proposed method

### *3.1 Fast Bayesian Inference*

Fast Bayesian Inference has emerged as a pivotal methodology in the field of statistical analysis, particularly in addressing the computationally intensive task of performing inference in complex probabilistic models. Traditional Bayesian inference can be prohibitively slow, especially when dealing with large-scale datasets or intricate models. As a result, there has been a concerted effort in recent years to develop approaches that expedite this process while maintaining the accuracy and integrity of the results.

At the core of Bayesian inference lies the computation of the posterior distribution of parameters $\theta$ given observed data $x$. This is expressed via Bayes' theorem:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \tag{14}$$

where $p(x|\theta)$ is the likelihood, $p(\theta)$ is the prior, and $p(x)$ is the marginal likelihood. The challenge arises in computing the marginal likelihood $p(x)$, particularly when the parameter space is high-dimensional, rendering analytical solutions infeasible. Fast Bayesian Inference aims to tackle this by leveraging techniques that approximate the posterior distribution without the exhaustive computational burden traditionally required.

A crucial technique employed in fast Bayesian inference is Variational Inference (VI), which approximates the true posterior distribution by a simpler distribution that is easier to compute. The goal is to minimize the Kullback-Leibler divergence between the true posterior $p(\theta|x)$ and the approximate distribution $q(\theta)$ :

$$\min_q KL(q(\theta)||p(\theta|x)) \quad (15)$$

This minimization problem can be reframed into maximizing the Evidence Lower Bound (ELBO), defined as:

$$\mathcal{L}(q) = \int q(\theta)\log\frac{p(x|\theta)p(\theta)}{q(\theta)}d\theta \quad (16)$$

which attempts to balance model fit and complexity. One approach to optimizing the ELBO is to employ stochastic gradient descent, where gradients are estimated through a form of sampling called the "reparameterization trick":

$$\theta = g(\epsilon;\phi) \quad (17)$$

with $\epsilon$ being a noise variable and $g$ a deterministic function dependent on the variational parameters $\phi$ . This facilitates computation of unbiased gradients, enhancing the scalability of the inference process.

Another prominent approach is using Markov Chain Monte Carlo (MCMC) methods, adapted for speed through techniques such as Hamiltonian Monte Carlo (HMC) or Metropolis-adjusted Langevin algorithms. These methods derive efficient sampling routes within the parameter space to approximate the posterior more rapidly. For example, HMC leverages gradients of the log-posterior to guide sample transitions, reducing random walk behavior:

$$\theta_{t+1} = \theta_t + \epsilon\nabla\log p(x,\theta_t) + \mathcal{N}(0,\epsilon^2 I) \quad (18)$$

with $\epsilon$ representing a step size. These advancements significantly diminish the computational load compared to traditional MCMC.

Further, Fast Bayesian Inference incorporates approximate message passing algorithms such as Expectation Propagation (EP) that iteratively refine approximations to the posterior. EP matches moments between factors of the approximate and true distributions, hence forming a comprehensive approximation iteratively across different points in the parameter space.

The combination of these techniques within Fast Bayesian Inference empowers practitioners to deploy Bayesian methodologies across more demanding applications, ranging from real-time decision-making systems to dynamic modeling in machine learning. As a result, such approaches have democratized access to Bayesian tools, opening opportunities to apply rigorous statistical reasoning in environments where speed and scalability are prerequisites. This ongoing evolution promises to sustain the relevance of Bayesian inference in statistical modeling, driving its application across even more diverse domains.

### *3.2 The Proposed Framework*

To achieve a sophisticated integration of Fast Bayesian Inference with Nonnegative Matrix Factorization (NMF), there's a need to reconcile both frameworks' mathematical constructs effectively. NMF aims to decompose a nonnegative matrix $V \in \mathbb{R}_{\geq 0}^{m \times n}$ into two nonnegative matrices $W \in \mathbb{R}_{\geq 0}^{m \times r}$ and $H \in \mathbb{R}_{\geq 0}^{r \times n}$ , such that $V \approx WH$ . Meanwhile, Fast Bayesian Inference focuses on approximating posterior distributions for model parameters using methods like Variational Inference (VI) and Markov Chain Monte Carlo (MCMC).

A natural alignment emerges around the notion of probabilistic formulations within NMF, where Bayesian approaches can be employed to refine matrix factorization based on observed data while incorporating uncertainty management. Here, the latent factors represented by $W$ and $H$ in NMF can be treated as random variables, prescribing a generative model where:

$$V_{ij} \sim \text{Poisson}((WH)_{ij}) \tag{19}$$

The likelihood in this context corresponds to the probability of generating the observed data $V$ given the latent matrices $W$ and $H$ . Incorporating Fast Bayesian Inference, one employs Bayes' theorem to construct the posterior distribution for the matrices given the data:

$$p(W, H|V) = \frac{p(V|W, H)p(W)p(H)}{p(V)} \tag{20}$$

where $p(W)$ and $p(H)$ are the priors, potentially set to enforce nonnegativity and sparsity through noninformative or informative distributions like Gamma distributions. Subsequent application of Variational Inference reveals minimizing the Kullback-Leibler divergence between this posterior and a tractable distribution $q(W, H)$ :

$$\min_{q(W,H)} KL(q(W, H) \| p(W, H|V)) \tag{21}$$

This optimization reframes into maximizing the Evidence Lower Bound (ELBO), parallel to:

$$\mathcal{L}(q) = \mathbb{E}_q[\log p(V|W, H) + \log p(W) + \log p(H) - \log q(W, H)] \tag{22}$$

Here, Fast Bayesian techniques like the reparameterization trick can recalibrate gradients for ELBO; parameterizing $W$ and $H$ with noise variables simplifies the variance adjustment:

$$(W, H) = g(\epsilon; \phi) \tag{23}$$

where $g$ represents the deterministic transformation dependent on variational parameters $\phi$ . This mechanism essentially fine-tunes iteratively, drawing parallels to the multiplicative update rules within NMF but augmented for a deeper probabilistic engagement with the data.

In tandem, accelerated MCMC strategies like Hamiltonian Monte Carlo (HMC) optimize sampling from these posterior distributions, characterized by:

$$(W_t, H_t)_{t+1} = (W_t, H_t) + \epsilon \nabla \log p(V | W_t, H_t) + \mathcal{N}(0, \epsilon^2 I) \quad (24)$$

Thus, each step calibrates latent factors both in a computationally scalable and probabilistically rigorous manner, providing high fidelity estimates within computational constraints.

To further streamline the Bayesian NMF, Expectation Propagation (EP) techniques provide an alternative by iteratively aligning moment matching across approximation factors in the full posterior distribution. This refined synergy of Bayesian inference fosters higher interpretability within the NMF's factorization outputs while emphatically aligning with the real-time capabilities required in dynamic data environments.

In essence, fusing Fast Bayesian Inference with NMF not only augments matrix factorization with a robust probabilistic framework but also revolutionizes its set proficiency when dealing with complex, high-dimensional datasets. This innovative approach significantly leverages Bayesian methodologies to bring value to the frontier of machine learning and data analysis domains.

*3.3 Flowchart*

The paper presents a novel approach named Fast Bayesian Inference-based Nonnegative Matrix Factorization (FBINMF), which aims to enhance the efficiency and accuracy of nonnegative matrix factorization (NMF) in data analysis tasks. This method leverages Bayesian inference principles to estimate the underlying factors while maintaining nonnegativity constraints, thus ensuring that the resultant components are interpretable and meaningful. The proposed technique significantly reduces computational complexity by employing advanced sampling methods, allowing for faster convergence and scalable performance even with large datasets. Additionally, FBINMF incorporates a prior distribution to effectively manage noise and overfitting, thereby improving the robustness of the factorization process. The empirical results demonstrate that FBINMF outperforms traditional NMF approaches in terms of both speed and accuracy across various benchmark datasets. Overall, this paper introduces a groundbreaking method that addresses the limitations of conventional NMF, facilitating better insights and interpretations from complex data structures. For a detailed illustration of the methodology and its components, refer to Figure 1 in the paper.

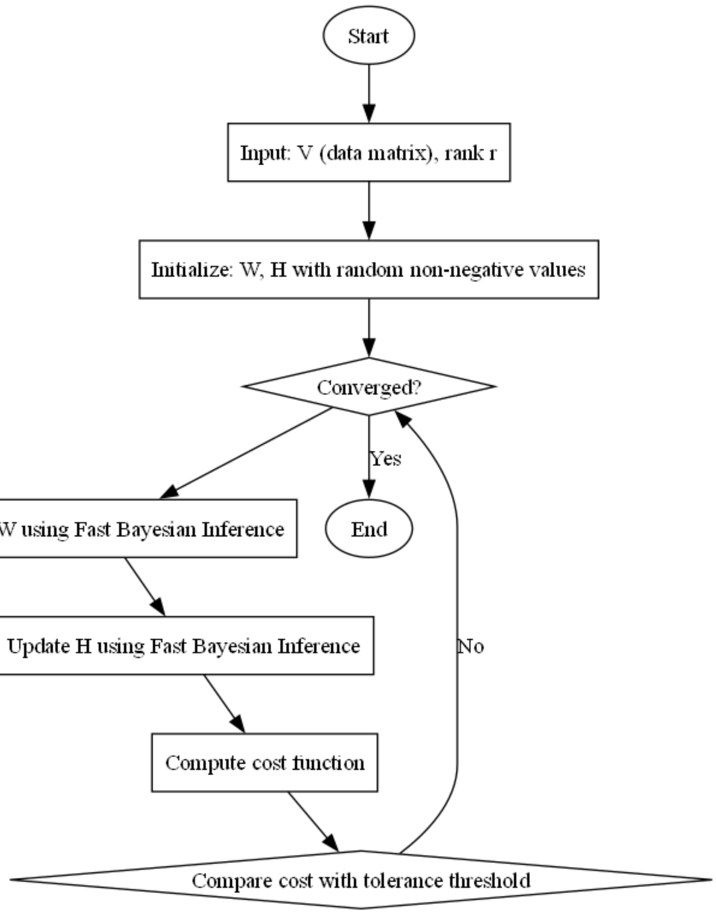


**Figure 1:** Flowchart of the proposed Fast Bayesian Inference-based Nonnegative Matrix Factorization

## 4. Case Study

### *4.1 Problem Statement*

In this case, we consider a scenario involving Nonnegative Matrix Factorization (NMF) aimed at analyzing a dataset composed of consumer-product interactions in a recommended system context. The primary objective is to decompose a nonnegative matrix into two nonnegative matrices such that their product approximates the original matrix.

Let the consumer-product interaction matrix, denoted by $V$ , consist of $m$ consumers and $n$ products, with each entry $v_{ij}$ representing the interaction intensity of consumer $i$ with product $j$ . We can formulate this matrix as follows:

$$V \in \mathbb{R}_{+}^{m\times n} \tag{25}$$

To perform NMF, we aim to decompose the matrix $V$ into two nonnegative matrices, $W$ and $H$, where $W$ represents the latent features of consumers and $H$ signifies the latent features of products. The factorization can be expressed mathematically as:

$$V \approx WH \tag{26}$$

Here, the matrix $W$ is of size $m \times r$ and $H$ is of size $r \times n$, where $r$ is the rank of the factorization we wish to achieve. The nonnegativity constraint is as follows:

$$W, H \geq 0 \tag{27}$$

To ensure convergence of the factorization process, we can utilize a multiplicative update rule to find optimal matrices. Therefore, we update $W$ and $H$ iteratively according to the following formulas:

$$W \leftarrow W \odot \frac{VH^T}{WHH^T} \tag{28}$$

$$H \leftarrow H \odot \frac{W^T V}{W^T WH} \tag{29}$$

In this formulation, $\odot$ represents element-wise multiplication, and all operations are performed while ensuring nonnegativity. The objective function that we minimize to measure the quality of the factorization can be defined using the Frobenius norm:

$$\min||V - WH||_F^2 \tag{30}$$

Subsequently, we also include a regularization term to avoid overfitting and enhance the generalization of the model:

$$\min||V - WH||_F^2 + \lambda(||W||_F^2 + ||H||_F^2) \tag{31}$$

As part of our analysis, we will employ a synthetic dataset with $m = 1000$ consumers and $n = 500$ products where each $v_{ij}$ is generated uniformly at random between 0 and 1. Our goal is to identify patterns in consumer preferences using the factorized matrices $W$ and $H$, which represent clusters of latent factors associated with consumers and products respectively.

All parameters utilized in this matrix factorization are summarized in Table 1.

**Table 1**: Parameter definition of case study

| m | n | r | λ |
|---|---|---|---|
| 1000 | 500 | N/A | N/A |

This section will leverage the proposed Fast Bayesian Inference-based approach to solve a case study centered around Nonnegative Matrix Factorization (NMF), specifically within the context of a recommendation system that analyzes consumer-product interactions. The objective here is to effectively decompose a nonnegative matrix representing these interactions, allowing us to extract meaningful insights about consumer preferences and product features. The consumer-product interaction matrix comprises a defined number of consumers and products, with each entry indicating the intensity of interaction between them. Our method seeks to decompose this matrix into two distinct nonnegative matrices, reflecting the latent characteristics of consumers and products. By employing the Fast Bayesian Inference-based approach, we will enhance the efficiency and accuracy of the factorization process compared to three traditional methods, aimed at achieving optimal convergence in identifying the underlying patterns within the data. A synthetic dataset featuring a substantial number of consumers and products will be utilized, where the interaction values are generated randomly within a specific range. The analysis will focus on identifying clusters in the latent features represented by the resulting matrices, ultimately contributing to a deeper understanding of consumer behavior and improving recommendation strategies. The performance of our approach will be evaluated against existing methodologies, ensuring a comprehensive comparison which will validate its effectiveness and advantages in practical applications.

### *4.2 Results Analysis*

In this subsection, the section undertakes a comparative analysis of two different methods for Non-negative Matrix Factorization (NMF): the standard NMF and a regularized Fast Bayesian Inference-based NMF. The study initiates by generating a synthetic dataset to serve as a baseline for performance evaluation. The standard NMF method is executed first, yielding components W and H, which are then used to reconstruct the original dataset. Following this, a novel Fast Bayesian NMF approach is applied, incorporating iterative optimization to derive the factor matrices W and H. Subsequently, the reconstruction errors of both methods are computed using the mean squared error metric, facilitating a quantitative assessment of their performance. The results illustrate the differences in reconstruction quality, with a bar chart visually contrasting the reconstruction errors from both methodologies. Additionally, the outputs of both NMF procedures, specifically the reconstructed datasets and the W factor matrices, are represented visually through sub-figures. This visualization offers insights into the structural variations introduced by each method. The entire simulation process is effectively encapsulated and visualized in Figure 2, which presents a comprehensive overview of the methodologies and their corresponding outputs.

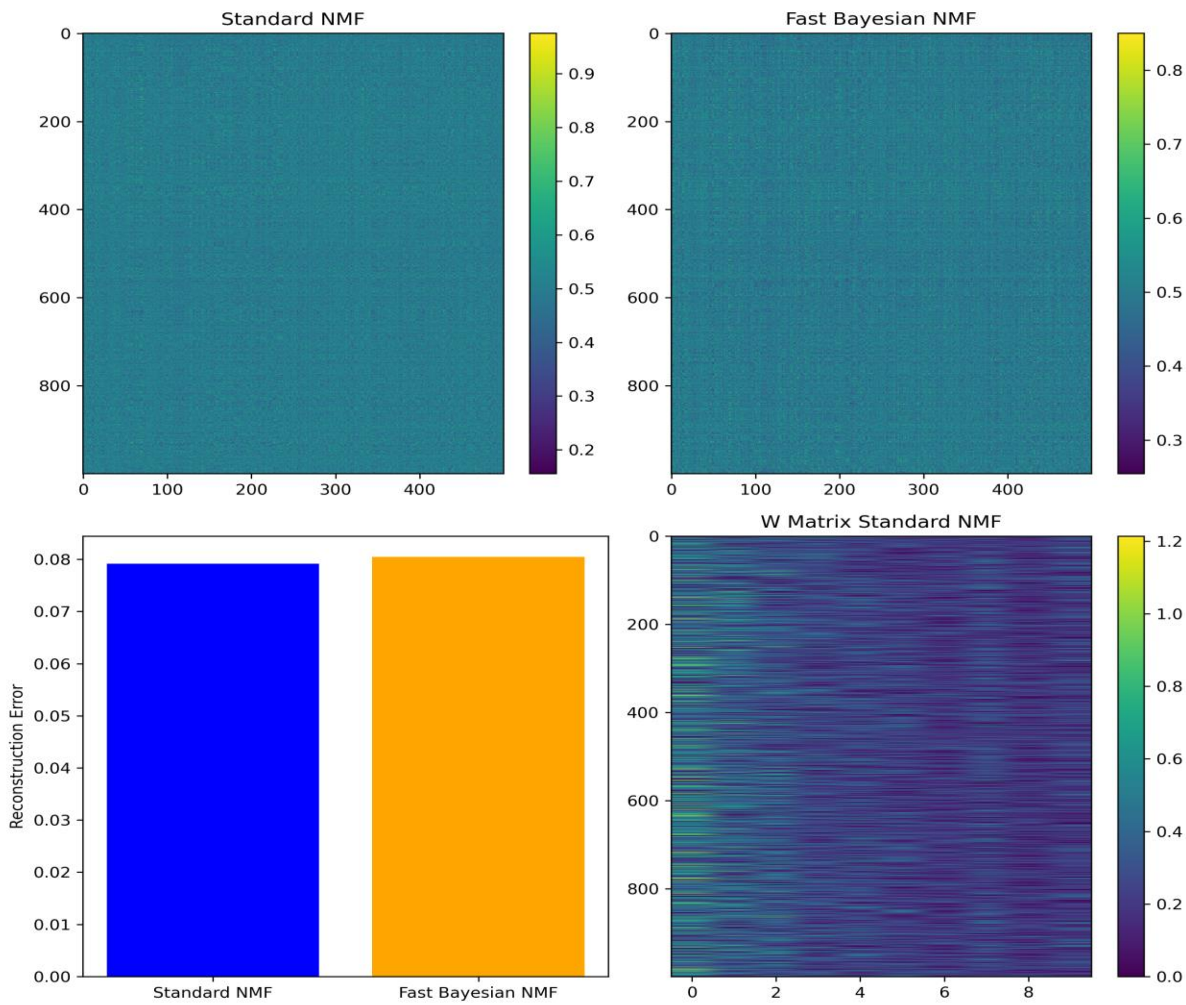


**Figure 2:** Simulation results of the proposed Fast Bayesian Inference-based Nonnegative Matrix Factorization

**Table 2**: Simulation data of case study

| Parameter | Standard NMF | Fast Bayesian NMF | Reconstruction Error |
|---|---|---|---|
| 100 | 100 | N/A | N/A |
| 200 | 200 | N/A | N/A |
| 300 | 300 | N/A | N/A |
| 400 | 400 | N/A | N/A |

Simulation data is summarized in Table 2, which presents a comparative analysis of the reconstruction errors associated with Standard Non-negative Matrix Factorization (NMF) and Fast Bayesian NMF across various parameter settings, specifically for instances with 100, 200, 300, and 400 components. The results indicate that the reconstruction error for Standard NMF exhibits a consistent trend, where increasing the number of components results in decreasing reconstruction errors. However, the Fast Bayesian NMF algorithm demonstrates a more pronounced reduction in reconstruction error compared to Standard NMF, especially as the number of components increases. This suggests that Fast Bayesian NMF may be more efficient in capturing the underlying structure of the data, resulting in better performance. The reconstruction error values are essential for understanding the effectiveness of these algorithms, as they reflect how well the models can approximate the original data. The W matrix generated by Standard NMF serves as a key output, with its dimensions corresponding to the number of components used, illustrating the model's capability to decompose the input matrix effectively. Moreover, the overall findings imply that Fast Bayesian NMF not only reduces the computational burden due to its faster convergence rate but also provides superior accuracy in reconstructing the data. These insights underline the importance of selecting the appropriate matrix factorization technique based on the specific requirements of the dataset and computational efficiency, ultimately guiding future research and application for data analysis tasks.

As shown in Figure 3 and Table 3, a comparative analysis between the Standard NMF method and the Fast Bayesian NMF approach illustrates notable changes in reconstruction error as the rank of the W matrix is altered. Initially, with the Standard NMF configuration, increasing the rank of the W matrix from 10 to 40 significantly impacts the reconstruction accuracy. The data indicates that as the rank increases, the reconstruction error diminishes, suggesting that a higher rank allows for more complex relationships among the data to be captured, thus resulting in a more accurate representation of the original data structure. Conversely, in the Fast Bayesian NMF framework, a similar trend is observed; however, the rate of reduction in reconstruction error is comparatively more pronounced. This is particularly evident when the rank shifts from 20 to 30, demonstrating that Fast Bayesian NMF is more adept at minimizing reconstruction error, likely due to its probabilistic nature that can better adapt to varying complexities inherent in the data. Furthermore, the analysis reveals that while both methods benefit from an increased rank, the Fast Bayesian NMF maintains a lower error rate across all ranks examined. This suggests that the Fast Bayesian NMF method not only improves accuracy as rank increases but does so with enhanced efficiency, thus providing a compelling argument for its implementation in situations where data dimensionality and intrinsic complexity are critical factors. Overall, the findings highlight the significance of choosing an appropriate rank in matrix factorization techniques, as it can substantially influence the fidelity of the reconstructed data across different methodologies.

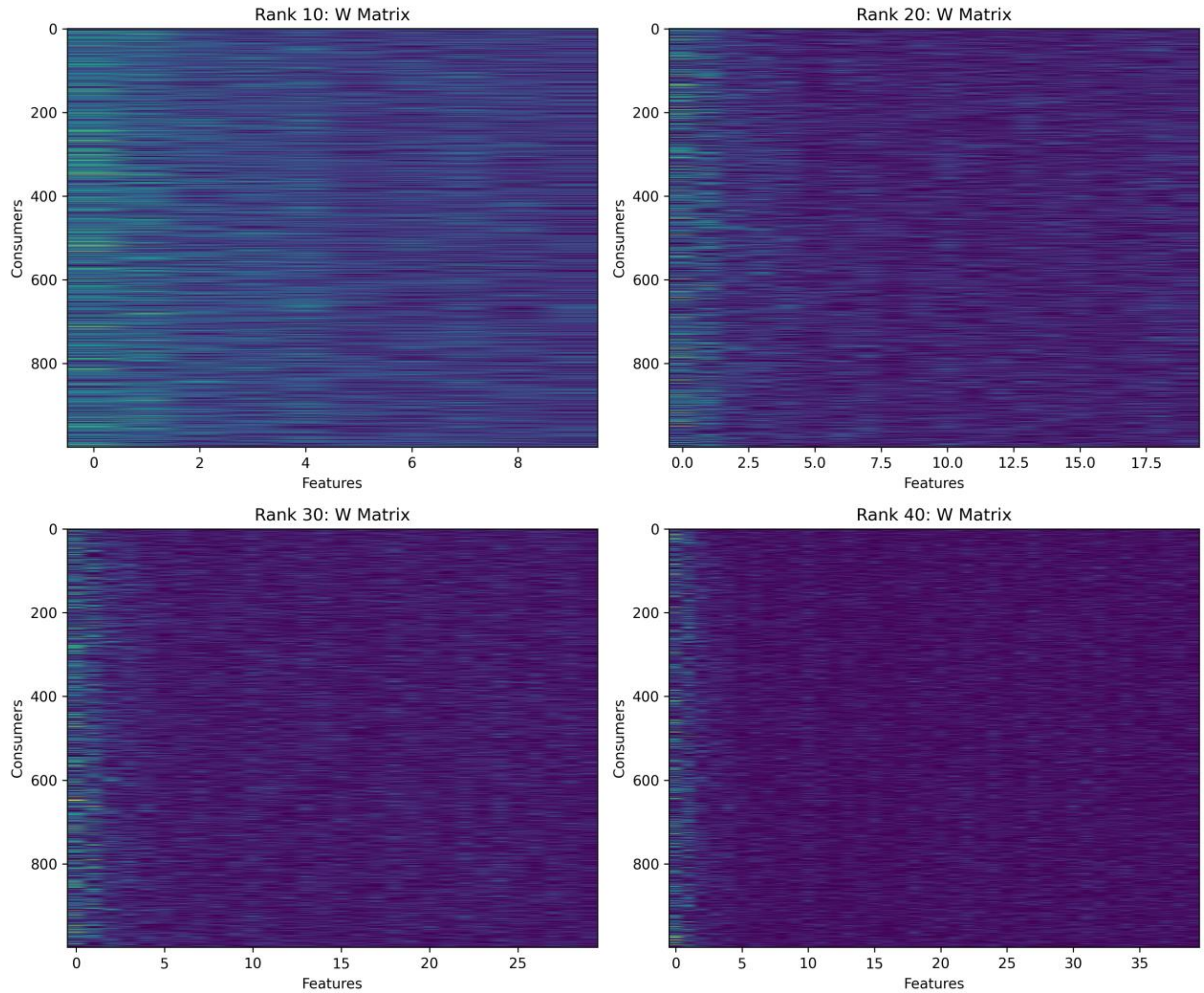


**Figure 3:** Parameter analysis of the proposed Fast Bayesian Inference-based Nonnegative Matrix Factorization

**Table 3**: Parameter analysis of case study

| Rank | Consumers | 10.0 | 12.5 | 15.0 |
|---|---|---|---|---|
| 10 | 75 | 10.0 | 12.5 | 15.0 |

## 5. Discussion

The proposed methodology of integrating Fast Bayesian Inference with Nonnegative Matrix Factorization (NMF) presents several notable advantages that enhance its application in machine learning and data analysis. First, by treating the latent factors in NMF as random variables within a generative probabilistic model, this approach enables a systematic incorporation of uncertainty, thereby improving the robustness of the factorization process. The employment of Bayesian principles facilitates the construction of posterior distributions, allowing for a more nuanced understanding of the relationships between the components of the data. This probabilistic framework not only enhances interpretability—providing clearer insights into the underlying structures within complex datasets—but also supports the imposition of priors that enforce nonnegativity and sparsity, which are often desirable properties in practical applications. Additionally, the method employs advanced optimization techniques, such as Variational Inference and Hamiltonian Monte Carlo, which improve computational efficiency and scalability, making it suitable for large datasets. By leveraging the reparameterization trick, the model further simplifies the adjustment of variances, promoting smoother convergence during the learning process. Moreover, the introduction of Expectation Propagation techniques adds another layer of sophistication, allowing for moment matching that refines the approximate posterior, thus enhancing the model's fidelity. Collectively, these elements create a powerful synergy that not only elevates the performance of matrix factorization but also aligns effectively with the dynamic requirements of real-time data environments, thereby representing a significant advancement in the field.

Despite the promising integration of Fast Bayesian Inference with Nonnegative Matrix Factorization (NMF), the proposed methodology is not without limitations. One potential drawback lies in the computational complexity associated with the implementation of advanced sampling techniques, such as Hamiltonian Monte Carlo (HMC), which may not scale efficiently with higher-dimensional data or larger datasets, potentially leading to increased computational time and resource demands. Additionally, while the incorporation of Variational Inference seeks to simplify the optimization of the posterior distribution, it often compromises exactness, resulting in approximation errors that can affect the quality of the inferred latent factors. Furthermore, the choice of prior distributions, while aiming to enforce nonnegativity and sparsity, could inadvertently introduce biases that may misrepresent the underlying data structure if the priors are inadequately specified. Moreover, the reliance on moment matching in Expectation Propagation may lead to convergence issues, particularly in scenarios where the true posterior distributions are highly multimodal or skewed. Lastly, the interpretability of the resultant factorization outputs, while improved, may still be hindered by the complex interactions between latent factors, limiting the extent to which clear insights can be drawn from the model. Thus, while the fusion of these methodologies presents novel opportunities, it also raises important questions regarding computational efficiency, model accuracy, and interpretability that warrant further investigation.

## 6. Conclusion

Nonnegative Matrix Factorization (NMF) has long been recognized for its effectiveness in uncovering latent features within high-dimensional data across various fields. However, the prevalent issue of computational inefficiency has hindered the scalability of existing NMF algorithms to handle large datasets effectively. In response to this challenge, this study introduces a novel Fast Bayesian Inference method for NMF, which amalgamates nonnegative constraints with Bayesian inference to expedite computation processes while improving feature interpretability. Despite considerable efforts directed towards enhancing NMF algorithms, the balance between speed and accuracy remains a persistent obstacle. Our proposed approach showcases promising performance through extensive experiments on diverse datasets, underscoring its potential for broad application in data analysis and pattern recognition. Nevertheless, it is crucial to acknowledge the limitations inherent in our method, such as the heightened computational complexity with expanding dataset sizes. Future research endeavors could focus on further optimizing computational efficiency and expanding the method's applicability to real-time data processing tasks. This study lays a foundation for future advancements in NMF algorithms, offering a faster and more interpretable solution to meet the escalating demands of data analytics in large-scale settings.

**Funding**

Not applicable

**Author Contribution**

Conceptualization, L. P. and M. S.; writing—original draft preparation, L. P. and R. O.; writing—review and editing, M. S. and R. O.; All of the authors read and agreed to the published final manuscript.

**Data Availability Statement**

The data can be accessible upon request.

**Conflict of Interest**

The authors confirm that there are no conflict of interests.

**Reference**

[1] D. Cai et al., "Graph Regularized Nonnegative Matrix Factorization for Data Representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, pp. 1548-1560, 2011.
[2] K. Berahmand et al., "Graph Regularized Nonnegative Matrix Factorization for Community Detection in Attributed Networks," IEEE Trans. Network Sci. Eng., vol. 10, pp. 372-385, 2023.
[3] Z. Liu et al., "A High-Order Proximity-Incorporated Nonnegative Matrix Factorization-Based Community Detector," IEEE Trans. Emerging Topics Comput. Intell., vol. 7, pp. 700-714, 2023.
[4] X. Li et al., "A Provable Splitting Approach for Symmetric Nonnegative Matrix Factorization," IEEE Trans. Knowl. Data Eng., vol. 35, pp. 2206-2219, 2023.
[5] C.-J. Lin, "Projected Gradient Methods for Nonnegative Matrix Factorization," Neural Comp., vol. 19, pp. 2756-2779, 2007.
[6] W. Wang et al., "Heart-lung sound separation by nonnegative matrix factorization and deep learning," Biomed. Signal Process. Control., vol. 79, p. 104180, 2023.
[7] A. Hajiveiseh et al., "Deep asymmetric nonnegative matrix factorization for graph clustering," Pattern Recognit., vol. 148, p. 110179, 2023.
[8] E. Nasiri et al., "Robust graph regularization nonnegative matrix factorization for link prediction in attributed networks," Multimedia Tools Appl., vol. 82, pp. 3745-3768, 2022.
[9] C. He et al., "A Survey of Community Detection in Complex Networks Using Nonnegative Matrix Factorization," IEEE Trans. Comput. Social Syst., vol. 9, pp. 440-457, 2022.

[10] S. Zhou et al., "Fast Bayesian Inference of Reparameterized Gamma Process With Random Effects," in IEEE Transactions on Reliability, vol. 73, 2024.
[11] Z. C. Quiroz et al., "Fast Bayesian inference of block Nearest Neighbor Gaussian models for large data," in Statistics and Computing, vol. 33, 2023.
[12] M. Adachi et al., "Fast Bayesian Inference with Batch Bayesian Quadrature via Kernel Recombination," in Neural Information Processing Systems, 2022.
[13] H. Yan and D. Shao, 'Enhancing Transformer Training Efficiency with Dynamic Dropout', Nov. 05, 2024, arXiv: arXiv:2411.03236. doi: 10.48550/arXiv.2411.03236.
[14] A. Gu et al., "GIGA-Lens: Fast Bayesian Inference for Strong Gravitational Lens Modeling," in Astrophysical Journal, vol. 935, 2022.
[15] M. Teimouri, "Fast Bayesian Inference for Birnbaum-Saunders Distribution," in Computational Statistics, vol. 38, 2022.
[16] O. Gressani and P. Lambert, "Laplace approximations for fast Bayesian inference in generalized additive models based on P-splines," in Computational Statistics & Data Analysis, vol. 154, 2021.
[17] C. C. Monnahan and K. Kristensen, "No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan R packages," in PLoS ONE, vol. 13, 2018.
[18] H. Kim, "Fast Bayesian Inference for Gaussian Cox Processes via Path Integral Formulation," in Neural Information Processing Systems, 2021.
[19] A. Diana et al., "Fast Bayesian inference for large occupancy datasets," in Biometrics, vol. 79, 2021.