Nonnegative Matrix Factorization: A Review

Abdul bin Ismail

School of Engineering and Computing, Manipal International University, Malaysia

E-mail: engabdulbinismail@gmail.com

Abstract

Recent developments in Non-negative Matrix Factorization (NMF) have focused on addressing several challenges and advancing its applicability. New algorithmic variations, such as robust NMF, deep NMF, and graph-regularized NMF, have emerged to improve NMF's performance in various domains. These developments aim to enhance the interpretability, scalability, and robustness of NMF-based solutions. NMF is now widely used in audio source separation, text mining, recommendation systems, and image processing. However, NMF still faces challenges, including sensitivity to initialization, the determination of the appropriate rank, and computational complexity. Overlapping sources in audio and data sparsity in some applications remain challenging issues. Additionally, ensuring the consistency and stability of NMF results in noisy environments is a subject of ongoing research. The quest for more efficient and scalable NMF algorithms continues, especially for handling large datasets. While NMF has made significant strides in recent years, addressing these challenges is crucial for unlocking its full potential in diverse data analysis and source separation tasks.

Keywords: Non-Negative Matrix Factorization, Review, Survey, Challenge

1. Introduction

Non-negative Matrix Factorization (NMF) [1] is a dimensionality reduction and feature extraction technique widely used in machine learning and data analysis. It is particularly useful when dealing with data that has a non-negative structure, such as image data, text data, and many other types of data where values represent quantities that cannot be negative, like pixel intensities or term frequencies. The primary goal of NMF is to factorize a given non-negative matrix into two lower-dimensional matrices such that when multiplied, they approximate the original matrix.

These two lower-dimensional matrices are often referred to as the "basis" and "coefficient" matrices. NMF works as follows [2]. Begins with a non-negative data matrix **X** of dimensions $m \times n$, where m is the number of samples or data points, and n is the number of features. NMF seeks to factorize **X** into two non-negative matrices, **G** ($m \times d$) and **H** ($d \times n$), where d is a user-defined parameter that represents the desired reduced dimensionality. The elements of **G** and **H** are constrained to be non-negative. The product of **G** and **H**, **GH**, is an approximation of the original matrix **X**. The goal is to find **G** and **H** such that **GH** is as close to **X** as possible while satisfying the non-negativity constraints. Let $\|\cdot\|_2$denote L_2 norms.

$$\min \|\mathbf{X} - \mathbf{GH}\|_2^2, \quad \text{s.t.}\, \mathbf{G}_{m,d} > 0, \mathbf{H}_{d,n} > 0.$$

NMF typically minimizes a cost function, such as the Frobenius norm, Kullback-Leibler divergence [2], or other measures, to quantify the dissimilarity between **X** and **GH**. NMF has found applications in various fields, including image processing, text mining, topic modelling, and bioinformatics. In image processing, for example, NMF can be used for feature extraction, where the basis matrix **G** represents image patterns, and the coefficient matrix **H** indicates the presence and extent of those patterns in different images. NMF is known for its ability to reveal meaningful and interpretable patterns in data, making it valuable for extracting features and reducing the dimensionality of data while preserving the non-negativity and sparsity of the original data. This interpretability has led to its adoption in a wide range of data analysis tasks [3]. While NMF has made significant strides in recent years, addressing these challenges is crucial for unlocking its full potential in diverse data analysis tasks [4, 5].

Non-negative constraints ensure that the factor matrices and the resulting approximation of the original matrix contain only non-negative values. In many applications, negative values do not have a physical or meaningful interpretation. For example, in recommendation systems, it doesn't make sense to have negative user preferences or item characteristics. Non-negative matrix factorization aims to decompose data into a set of additive components. Non-negativity ensures that each component's contribution is additive and non-cancelling, which can lead to more interpretable results. Non-negative constraints encourage part-based representations. This means that each component in the factor matrices represents a part or a feature, and the combination of these parts results in a positive value. This aligns with the idea of additive or compositional models, which are common in various domains. Non-negativity enforces sparsity in factor matrices, meaning that most entries in the factor matrices are zero, and only a few components contribute to the approximation. This sparsity often results in a more selective representation, where only the most relevant components are considered.

Non-negative Matrix Factorization (NMF) ensures stability and consistency in noisy areas through several key strategies. First, NMF enforces non-negativity constraints, which help separate signal from noise, ensuring that factorization components are additive and interpretable. Secondly, NMF encourages sparsity in factor matrices, reducing the influence of noisy data by emphasizing only the most relevant components. Third, regularization terms are introduced to control the impact of noise, preventing overfitting to noisy data points. Fourth, proper data pre-processing techniques are employed to detect and handle noisy data points, reducing their impact on the results [4, 5]. Finally, NMF often involves multiple runs with different initializations, and the most stable and consistent factorization is selected, reducing sensitivity to noise-related local minima [1]. These strategies collectively make NMF a robust tool for data analysis, even in the presence of noisy data.

The rest of this study is organized as follows. Section 2 shows related works. Section 3 is related applications with challenges mentioned in Section 4. Section 5 is the conclusion.

2. Related Work

Non-negative Matrix Factorization (NMF) techniques can be categorized based on various criteria, including the objective function, regularization methods, algorithmic approaches, and applications. Here is a categorization of NMF techniques. Table 1 show the developments, challenges, and effectiveness of NMF. This study used systematic literature review (SLR). This study defines a research question, establishes inclusion and exclusion criteria, and systematically searches databases, academic journals, and other relevant sources for papers, articles, and reports that meet these criteria. This method ensures a comprehensive and structured overview of existing research. This study also used academic databases like Springer, IEEE, Google Scholar, and Elsevier to search for peer-reviewed journal articles, conference papers, theses, and reports.

Table 1. Recent Developments, Challenges, and Effectiveness of NMF

Aspect	Recent Developments	Effectiveness	Challenges
Initialization	Advanced techniques such as sparse initialization and randomization improve convergence.	Proper initialization remains critical for stability.	Initialization sensitivity can lead to local optima.
Regularization	Regularized NMF variants mitigate overfitting and enhance stability.	Improved control over model complexity and robustness.	Selecting appropriate regularization terms and parameters.
Sparsity Control	Methods to enforce sparsity, such as sparse NMF, provide better feature selection.	Sparsity encourages more interpretable factorizations.	Balancing sparsity with information loss is a delicate task.
Robustness	Robust NMF variants that consider outliers enhance stability.	Improved performance in the presence of noisy data.	Identifying and handling outliers is a complex task.
Variants	Diverse NMF variants like probabilistic NMF, hierarchical NMF, and deep NMF offer flexibility.	Tailored solutions for various data types and applications.	The plethora of variants requires selecting the most suitable one.
Applications	Expanding use in fields like healthcare, bioinformatics, and natural language processing.	Effectively extracts patterns and structures in diverse datasets.	Application-specific challenges require adaptation.

2.1 Objective Functions

The typical is Standard Non-negative Matrix Factorization (NMF) techniques form a class of algorithms designed for the decomposition of a given non-negative data matrix into two lower-dimensional matrices. The primary objective of Standard NMF is to minimize a specific objective function, which can be either the Frobenius norm or the Kullback-Leibler (KL) divergence, measuring the dissimilarity between the original data matrix and its approximation derived from the product of the factor matrices. For Frobenius norm-based NMF, in this approach, the objective is to minimize the Frobenius norm, which computes the sum of squared differences between corresponding elements of the original data matrix and its approximation. This form of NMF is well-suited for applications where the focus is on preserving the overall structure and magnitude of the data.

For divergence-based NMF, e.g., Kullback-Leibler divergence, it is a measure of the relative entropy between probability distributions. NMF algorithms that use KL divergence seek to minimize the information loss when approximating the original data matrix with the product of the factor matrices. KL divergence-based NMF is often preferred in applications where data sparsity and relative information preservation are crucial, such as in text mining or topic modeling. For divergence-based NMF, it encompasses a broader category of algorithms that extend beyond the standard Frobenius norm or KL divergence. In this category, the focus is on optimizing different divergence measures, such as the Itakura-Saito divergence [2] or the generalized Kullback-Leibler divergence. For example, Itakura-Saito divergence-based NMF is another divergence measure often used for NMF. It is especially well-suited for applications where a more robust approach to non-negativity is required. NMF algorithms employing Itakura-Saito divergence aim to minimize this specific divergence metric, which is known for its sensitivity to extreme values. Consequently, Itakura-Saito divergence-based NMF is used in scenarios where robustness to outliers is essential. For example, Generalized Kullback-Leibler divergence-based NMF [2] provides a more flexible approach compared to the standard KL divergence. NMF algorithms within this category seek to minimize the generalized KL divergence, which can better capture the information content between the original data and its approximation. For example, generalized KL divergence-based NMF is advantageous when dealing with complex, multi-modal data and when a more adaptive divergence measure is required. These two categories, Standard NMF and Divergence-Based NMF, represent different approaches to Non-negative Matrix Factorization, each with its particular strengths and suitability for various applications. The choice between them often depends on the nature of the data and the specific goals of the factorization task.

Half-Quadratic Non-negative Matrix Factorization (HQ-NMF) and robust NMF [6] are variations of Non-negative Matrix Factorization that incorporates a half-quadratic penalty term

[7-19], $L_{2,0}$ norms [10], and $L_{2,1}$ norms [11, 12] into the objective function during the factorization process. HQ-NMF is used for factorizing non-negative data matrices while promoting specific properties in the factor matrices. In HQ-NMF, the objective function typically includes the following components: A data-fitting term and a half-quadratic regularization term [13-15]. The regularization term is the key distinguishing feature of HQ-NMF [16]. The objective function is designed to find factor matrices that minimize the data reconstruction error while adhering to certain constraints or properties. The half-quadratic term is introduced to impose particular structure or characteristics on the factor matrices [17]. It can take different forms, depending on the specific requirements of the factorization task. The half-quadratic term is often designed to promote properties such as sparsity, smoothness, or structured patterns in the factor matrices. It serves as a regularization mechanism to control the factorization process [6, 18].

2.2 Regularization Methods

Regularization methods [19] in Non-negative Matrix Factorization are techniques used to introduce additional constraints or penalties into the NMF optimization process. These constraints serve various purposes, including promoting sparsity, avoiding overfitting, enhancing interpretability, and achieving specific factorization goals. For sparsity constraints: L_1-norm regularization [20, 21] encourages sparsity by adding the L_1-norm of the factor matrices to the objective function. It effectively enforces a sparse representation, which means that many elements of the factor matrices become zero, leading to a parts-based representation of the data. Concerning $L_{2,1}$-norm regularization, it promotes group sparsity. It encourages entire rows (or columns) of the factor matrices to be zero, making it suitable for situations where variables or features can be grouped together. As for smoothness constraints, L_2-norm regularization adds the squared L_2-norm of the factor matrices to the objective function. It encourages a smoother and more continuous representation of the data. This is useful when preserving the overall structure and relationships within the data is important. For example, nuclear norm regularization is often used in matrix factorization settings. It encourages low rank factor matrices, which is valuable in data compression and collaborative filtering tasks. About orthogonal constraints, e.g., orthogonal NMF [22], one or both factor matrices are constrained to be orthogonal, which helps in capturing mutually exclusive components. Orthogonal NMF can be useful for separating sources in signal processing. Regarding non-negativity constraints, while NMF inherently enforces non-negativity constraints on the factor matrices, some regularization methods explicitly emphasize non-negativity to ensure that the factors remain positive throughout the optimization process. Considering temporal and spatial constraints, for data with temporal or spatial dependencies, regularization methods can be designed to incorporate these constraints. This is often seen in applications like video analysis and spatial data factorization. Regularization can be used to balance the contributions of the factor matrices. This can be valuable when we want one factor matrix to play a more dominant role in the factorization. In some cases, a combination of regularization methods [23] is used to achieve specific objectives. For example, combining L_1-norm and L_2-norm regularization can promote both sparsity and smoothness in the factorization. Regularization methods in NMF are selected based on the specific problem and objectives. The choice of regularization can significantly impact the quality of the factorization and the interpretability of the results.

Graph regularization [24] in non-negative matrix factorization is a technique that combines matrix factorization with information from a graph structure. This approach is particularly useful when dealing with data that has a natural graph representation, such as social networks, citation networks, or co-occurrence networks. Graph regularization methods aim to incorporate the relationships or connections between data points as constraints or regularization terms during the NMF optimization process. Graph-based regularization in NMF involves integrating a graph representation of the data into the factorization process [24]. The graph is typically represented as an adjacency matrix, where entries reflect the strength of connections between data points (nodes). In the context of NMF, the graph can be directed or undirected, weighted or unweighted, and it can have various structures, such as *k*-nearest neighbor graphs or fully connected graphs. Graph regularization techniques include Graph Laplacian regularization and graph-based sparsity constraints. The graph Laplacian is a matrix derived from the adjacency matrix, representing the graph's structure. It can be used as a regularization term to encourage smoothness or consistency of factor matrices with respect to the graph. By minimizing the graph Laplacian regularization term, NMF aims to capture underlying patterns in the data while respecting the graph's topology or even multiple graphs. In some cases, graph-based sparsity constraints can be used to enforce sparsity constraints on the factor matrices. For example, nodes in the graph may represent features, and the edges may represent relationships. By promoting sparsity in factor matrices, NMF can discover distinct patterns or clusters of data points based on the graph structure.

2.3 Algorithmic Update Approaches

Algorithmic update approaches (shown in Table 2) in Non-negative Matrix Factorization (NMF) pertain to the methods and strategies used to iteratively update the factor matrices to minimize the objective function during the factorization process [26-31]. NMF aims to factorize a given non-negative data matrix into two lower-dimensional matrices with non-negative elements. These updates are performed until convergence is achieved, and the factor matrices provide a satisfactory approximation of the original data. For multiplicative updates, multiplicative update rules are the most common update methods used in NMF. These updates iteratively adjust the elements of the factor matrices based on the gradients of the objective function. The elements are updated multiplicatively to maintain non-negativity. Multiplicative updates [32] are computationally efficient and have proven to be highly effective in practice. Alternating Least Squares (ALS) [32] is an iterative optimization technique that alternates between updating one of the factor matrices while keeping the other fixed. It proceeds by minimizing the objective function with respect to one factor matrix at a time. ALS can be used with different optimization criteria, including Frobenius norm, Kullback-Leibler divergence, and other divergence measures. Block Coordinate Descent methods [33] update entire columns (or rows) of factor matrices simultaneously, which can lead to faster convergence and may be more computationally efficient than multiplicative updates [34, 35]. These methods involve splitting the factor matrices into blocks and updating one block at a time. Projected Gradient Descent [36] combines gradient descent with a projection step that enforces non-negativity. This approach updates the elements of the factor matrices in the direction of the gradient and then projects them back into the non-negative space. Projected Gradient Descent can be used with various regularization terms and loss functions. There are still many other algorithms [37].

Table 2. presents the Summary of multiplicative updates, alternating least squares, and hierarchical NMF along with their strengths and weaknesses.

Table 2. Summary of Other NMF Algorithms

Aspect	Multiplicative Updates	Alternating Least Squares	Hierarchical NMF
Strength	Simple and computationally efficient. Suitable for large-scale datasets. Often converges quickly.	Can be more stable and less sensitive to initialization. Can provide more accurate factorization under certain conditions.	Captures hierarchical structures in data. Useful for tasks where data has multi-level dependencies. Allows for hierarchical representation.
Weakness	Prone to getting stuck in local minima. Sensitive to initializations. Can be less accurate in some cases.	May require more iterations to converge. Not as intuitive as MU for some applications. May not scale well to very large datasets.	Computationally expensive for deep hierarchies. Complexity increases with each additional layer. Interpretability may decrease with depth.
Notable Variants	Sparse NMF, robust NMF, and non-negative tensor factorization [28-31].	Weighted NMF, constrained NMF, and semi-supervised NMF [32].	Convolutional NMF and non-negative hierarchical latent variable models [34, 35].

2.4 Deep Learning

Deep Non-negative Matrix Factorization (Deep NMF) [38-40] is an extension of the traditional Non-negative Matrix Factorization (NMF) method that introduces multiple layers or stages of factorization, allowing for more complex and hierarchical representations of the data. Deep NMF leverages the principles of deep learning to discover intricate hierarchical patterns in non-negative data [40]. Deep NMF introduces multiple layers, typically three or more, in the factorization process. Each layer is responsible for uncovering specific patterns or representations in the data. Factorization is performed in a layer-wise manner, starting with the input data and proceeding to the subsequent layers. Each layer factorizes the output of the previous layer into lower-level representations. Each layer may incorporate non-linear activation functions, such as ReLU (Rectified Linear Unit) or sigmoid, which add non-linearity to the factorization process. These nonlinearities enable Deep NMF to capture complex, non-linear patterns in the data. The hierarchical structure of Deep NMF allows for the discovery of intricate and abstract representations of the data.

3. Related Applications

NMF has a wide range of applications across various domains due to its ability to uncover latent structures and patterns in data. Some of the key applications of NMF include:

3.1 Image Processing

For image compression [41, 42], NMF is used to reduce the dimensionality of images while preserving important visual features. Image compression using Non-negative Matrix Factorization (NMF) is a technique that aims to reduce the amount of data required to represent an image while preserving its essential visual content [5]. NMF, with its non-negativity constraint, is well-suited for image compression because it naturally separates an image into its constituent parts, making it possible to represent the image using a smaller number of components. An image is represented as a data matrix, where each row corresponds to a pixel, and each column corresponds to a colour channel (e.g., red, green, blue). The data matrix is factorized into two non-negative matrices, often referred to as the basis and coefficient matrices. The basis matrix contains the basis images or components that represent different features or patterns present in the image. The coefficient matrix contains the coefficients that specify how much of each basis image is required to reconstruct the original image. To achieve compression, the number of basis images is reduced compared to the original image dimensions. This results in a lower-dimensional representation of the image. The basis matrix and the coefficient matrix, along with the reduced number of basis images, are stored. These matrices and the compressed representation occupy significantly less space than the original image [43]. To view or process the compressed image, the basis matrix, coefficient matrix, and the reduced number of basis images are used to reconstruct the original image. The quality of the reconstruction depends on the number of basis images retained.

For face/object recognition, NMF is a technique that leverages NMF to identify and classify faces or objects within images or video streams [42, 46-48]. NMF is used to extract meaningful features from visual data, making it a valuable tool for recognition tasks. In face recognition, NMF is used to extract discriminative features from facial images. These features may include facial landmarks, textures, and other relevant characteristics. The face images are typically represented as a data matrix, where each row corresponds to a different facial image, and each column corresponds to pixel values. During the training phase, NMF learns the basis images and coefficients from a dataset of labelled facial images. This phase creates a dictionary of facial features that can be used for recognition. In the recognition phase, NMF is applied to an input facial image to extract its coefficients based on the learned basis images. The extracted coefficients are then used for comparison with coefficients from the training dataset to identify the person [38, 40].

3.2 Recommendation Systems

NMF is applied to analysis of user-item interaction data to make personalized recommendations, such as in movie or product recommendations [42]. Recommendation systems using Non-negative Matrix Factorization (NMF) are designed to provide personalized recommendations to users by analysing their past behaviour and preferences. NMF is applied to factorize user-item interaction data into interpretable components, enabling recommendation systems to suggest relevant items or content to users. User-item interaction data is represented in a matrix format, where rows represent users, columns represent items (e.g., movies, products, articles), and the values in the matrix represent user-item interactions, such as ratings, clicks, or purchase history. NMF is applied to factorize the user-item interaction matrix into two non-negative matrices: the user matrix and the item matrix. The user matrix represents user preferences or characteristics, while the item matrix represents item features or attributes. The factorization process reduces the dimensionality of the data, capturing latent factors that explain user-item interactions. To generate recommendations, NMF leverages the user and item matrices to predict missing or unobserved interactions. The dot product of the user matrix and item matrix provides a recommendation score for each user-item pair. Top items with the highest recommendation scores are suggested to users. Recommendation system techniques can be applied to data imputation [43]. In essence, NMF-based recommendation system techniques, or equivalently collaborative filtering, provide a framework for factorizing matrices to understand underlying patterns. By extending these techniques to data imputation, we can leverage the same principles to fill in missing values, particularly in situations where understanding the relationships and latent factors between variables is essential for accurate imputation. This adaptability makes NMF a versatile tool for handling missing data in various types of datasets even in industry [44].

3.3 Text Mining and Natural Language Processing

Text mining and Natural Language Processing (NLP) using Non-negative Matrix Factorization (NMF) [45] is a technique that involves the analysis of textual data to discover patterns, extract topics, and reveal hidden structures within text documents. NMF is particularly useful for dimensionality reduction and topic modelling in text data. In text data representation, text documents are represented as a document-term matrix, where rows correspond to documents, and columns represent terms or words. The values in the matrix can be binary (presence or absence of words), term frequency (number of times a term appears in a document), or term frequency-inverse document frequency (TF-IDF) scores. NMF is used to factorize the document-term matrix into two non-negative matrices: the document-topic matrix and the topic-term matrix. NMF generates topics that are non-negative linear combinations of terms. These topics are interpretable and can be used to label and categorize documents based on their content. By examining the terms associated with each topic, it is possible to identify the main themes or topics present in a collection of documents [45].

3.4 Speech and Audio Analysis

Audio source separation [42,] using Non-negative Matrix Factorization (NMF) is a signal processing technique that aims to separate mixed audio signals into their constituent source components. NMF is particularly useful for extracting underlying sources from complex audio recordings, such as music, speech, or environmental sounds. In spectrogram representation, audio signals are transformed into spectrograms, which represent the audio content in the time-frequency domain. Spectrograms are matrices where rows correspond to frequency bins, columns correspond to time frames, and the values represent the magnitude or power of each frequency component at each time frame. For audio source separation with NMF, NMF is applied to the spectrogram of the mixed audio signal, decomposing it into two non-negative matrices: the basis matrix and the coefficient matrix [47]. The basis matrix represents spectral templates or source components, and the coefficient matrix encodes the activations of these sources at each time frame. By applying NMF, it's possible to extract individual sources by multiplying the basis matrix by the coefficient matrix for each source. Each source component is reconstructed from the spectrogram by taking the product of the basis corresponding to that source and the corresponding coefficients in the coefficient matrix. Once individual sources are separated, they can be enhanced or isolated from the mixed audio signal. This enables the extraction of specific sound sources, such as vocals, instruments, or background noise, from a complex audio recording [46].

4. Challenges

Non-negative Matrix Factorization (NMF) is a powerful technique with a wide range of applications, but it also comes with several challenges and limitations [47]. Researchers and practitioners should be aware of these limitations when applying NMF to various tasks and datasets and should consider alternatives or complementary techniques when NMF may not be the best fit for a particular problem.

1. Initialization Sensitivity: The quality of the NMF factorization can be sensitive to the choice of initial values. Different initializations can lead to different local minima, affecting the accuracy and consistency of the results.

2. Determination of Rank: Selecting the appropriate rank (the number of components or factors) for factorization is often a challenge. An incorrect choice can lead to overfitting or under-fitting, impacting the quality of the decomposition.

3. Overfitting: NMF can over-fit the data if the rank is chosen to be too high relative to the amount of available data. This can result in the extraction of spurious patterns or over-segmentation of sources.

4. Overlapping Sources: NMF may not be effective in scenarios where sources overlap significantly in time or frequency. Separating such sources can be a challenging problem.

5. Sparse Data: NMF may not perform well with very sparse data, where most values are zeros. Sparse data can lead to unstable factorization results.

6. Convergence and Local Minima: NMF algorithms may converge to local minima rather than the global minimum of the objective function. Techniques to address this issue, such as using multiple initializations, are often required.

7. Consistency and Stability: Ensuring the consistency and stability of NMF results, especially in the presence of noise or variations, can be a complex problem. Addressing these challenges often requires careful parameter tuning, algorithm selection, and domain-specific knowledge.

5. Conclusion

Recent developments in Non-negative Matrix Factorization (NMF) have expanded its applications and capabilities across various domains. New algorithmic variations and extensions have emerged, addressing specific challenges and enhancing NMF's versatility. For example, robust NMF techniques have been introduced to improve NMF's performance in the presence of noise and outliers, making it more resilient to real-world data imperfections.

In the field of audio source separation, NMF-based solutions have advanced, enabling the extraction of individual sound sources from complex mixtures, benefiting music production, speech enhancement, and environmental sound analysis. In text mining and natural language processing, NMF plays a pivotal role in topic modelling, document clustering, and content recommendation. Its non-negativity constraint and interpretability make it suitable for understanding and categorizing large text corpora. Challenges persist in the NMF landscape, including sensitivity to initialization, rank determination, and computational complexity. Overlapping sources in audio recordings pose difficulties for separation, and data sparsity in certain applications demands more robust strategies. Consistency and stability in noisy environments continue to be areas of research focus. Ensuring the scalability of NMF algorithms, especially for handling large datasets, remains essential.

In summary, NMF's recent developments have expanded its use and effectiveness in multiple fields, while the challenges it faces necessitate ongoing research and innovation. By addressing these challenges, NMF can continue to contribute significantly to data analysis, signal processing, and information extraction in a broad spectrum of applications.

References

[1] Z.-Y. Zhang, "Nonnegative matrix factorization: Models, algorithms and applications," in Data Mining: Foundations and Intelligent Paradigms, vol. 2, D. E. Holmes and L. C. Jain, Eds. Berlin, Heidelberg, German: Springer Berlin Heidelberg, 2012, pp. 99–134.

[2] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 6, pp. 1336–1353, Jun. 2013.

[3] B.-W. Chen, W. Ji, S. Rho, and Y. Gu, "Supervised collaborative filtering based on ridge alternating least squares and iterative projection pursuit," IEEE Access, vol. 5, pp. 6600–6607, Mar. 2017.

[4] W. Yan, B. Zhang, and Z. Yang, "A novel regularized nonnegative matrix factorization for spectral-spatial dimension reduction of hyperspectral imagery," IEEE Access, vol. 6, pp. 77953–77964, Dec. 2018.

[5] W. Yan, B. Zhang, Z. Yang, and S. Xie, "Similarity learning-induced symmetric nonnegative matrix factorization for image clustering," IEEE Access, vol. 7, pp. 166380–166389, Nov. 2019.

[6] J. Huang, F. Nie, H. Huang, and C. Ding, "Robust Manifold Nonnegative Matrix Factorization," ACM Transactions on Knowledge Discovery from Data, vol. 8, no. 3, pp. 1-21, Jun. 2014.

[7] C. Bo, "Symmetric nonnegative matrix factorization based on box-constrained half-quadratic optimization," IEEE Access, vol. 8, pp. 170976–170990, Sep. 2020.

[8] L. Du, X. Li, and Y.-D. Shen, "Robust nonnegative matrix factorization via half-quadratic minimization," in Proc. 12th IEEE International Conference on Data Mining, Brussels, Belgium, 2012, Dec. 10–13, pp. 201–210.

[9] Y. He, F. Wang, Y. Li, J. Qin, and B. Chen, "Robust matrix completion via maximum correntropy criterion and half-quadratic optimization," IEEE Transactions on Signal Processing, vol. 68, pp. 181–195, 2020.

[10] W. Min, T. Xu, X. Wan, and T.-H. Chang, "Structured sparse non-negative matrix factorization with ℓ2,0-norm," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 8, pp. 8584–8595, Aug. 2023.

[11] D. Kong, C. Ding, and H. Huang, “Robust nonnegative matrix factorization using L21-norm,” in Proc. 20th ACM International Conference on Information and Knowledge Management, Glasgow, Scotland, United Kingdom, 2011, Oct. 24–28, pp. 673–682.

[12] B. Wu, E. Wang, Z. Zhu, W. Chen, and P. Xiao, “Manifold NMF with L21 norm for clustering,” Neurocomputing, vol. 273, pp. 78–88, Jan. 2018.

[13] R. He, W.-S. Zheng, T. Tan, and Z. Sun, “Half-quadratic-based iterative minimization for robust sparse representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 2, pp. 261–275, May 2014.

[14] T. Jin, R. Ji, Y. Gao, X. Sun, X. Zhao, and D. Tao, “Correntropy-induced robust low-rank hypergraph,” IEEE Transactions on Image Processing, vol. 28, no. 6, pp. 2755–2769, Jun. 2019.

[15] H. J. Yang and C. B. Wei, “Robust data-aware sensing based on half quadratic hashing for the Visual Internet of Things,” IEEE Internet of Things Journal, vol. 10, no. 5, pp. 3890–3900, Mar. 2023.

[16] A. Liutkus, D. Fitzgerald, and R. Badeau, “Cauchy nonnegative matrix factorization,” in Proc. 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, United States, 2015, Oct. 18–21.

[17] A. Zach and G. Bourmaud, “Multiplicative vs. additive half-quadratic minimization for robust cost optimization,” in Proc. 28th British Machine Vision Conference, Newcastle, United Kingdom, 2018, Sep. 03–06.

[18] Y. Zhang, Z. Sun, R. He, and T. Tan, “Robust low-rank representation via correntropy,” in Proc. 2013 2nd IAPR Asian Conference on Pattern Recognition, Naha, Japan, 2013, Nov. 05–08, pp. 461–465.

[19] X. Fu, K. Huang, N. D. Sidiropoulos, and W.-K. Ma, “Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications,” IEEE Signal Processing Magazine, vol. 36, no. 2, pp. 59–80, Mar. 2019.

[20] P. O. Hoyer, “Non-negative matrix factorization with sparseness constraints,” Journal of Machine Learning Research, vol. 5, pp. 1457–1469, Dec. 2004.

[21] X. Dai, X. Su, W. Zhang, F. Xue, and H. Li, “Robust Manhattan non-negative matrix factorization for image recovery and representation,” Information Sciences, vol. 527, pp. 70–87, Jul. 2020.

[22] S. Choi, “Algorithms for orthogonal nonnegative matrix factorization,” in Proc. 2008 IEEE International Joint Conference on Neural Networks, Hong Kong, China, 2008, Jun. 01–08, pp. 1828–1832.

[23] Y. Li and A. Ngom, “Versatile sparse matrix factorization: Theory and applications,” Neurocomputing, vol. 145, no. 5, pp. 23–29, Dec. 2014.

[24] S. Yang, C. Hou, C. Zhang, Y. Wu, and S. Weng, “Robust non-negative matrix factorization via joint sparse and graph regularization,” in Proc. 2013 International Joint Conference on Neural Networks, Dallas, Texas, United States, 2013, Aug. 04–09, pp. 1–5

[25] Z. Yang and E. Oja, “Linear and nonlinear projective nonnegative matrix factorization,” IEEE Transactions on Neural Networks, vol. 21, no. 5, pp. 734–749, May 2010.

[26] Z. Yuan and E. Oja, “Projective nonnegative matrix factorization for image compression and feature extraction,” in Proc. Scandinavian Conference on Image Analysis, Joensuu, Finland, 2005, Jun. 19–22, pp. 333–342.

[27] N. Gillis and F. Glineur, “Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization,” Neural Computation, vol. 24, no. 4, pp. 1085–1105, Apr. 2012.

[28] A. Vandaele, N. Gillis, Q. Lei, K. Zhong, and I. Dhillon, “Coordinate descent methods for symmetric nonnegative matrix factorization,” IEEE Transactions on Signal Processing, vol. 64, no. 21, pp. 5571–5584, May 2016.

[29] H. Kim and H. Park, “Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method,” SIAM Journal on Matrix Analysis and Applications, vol. 30, no. 2, pp. 713–730, 2008.

[30] J. Kim and H. Park, “Toward faster nonnegative matrix factorization: A new algorithm and comparisons,” in Proc. 8th IEEE International Conference on Data Mining, Pisa, Italy, 2008, Dec. 15–19, pp. 353–362.

[31] C.-J. Lin, “Projected gradient methods for nonnegative matrix factorization,” Neural Computation, vol. 19, no. 10, pp. 2756–2779, Oct. 2007.

[32] N. Guan, D. Tao, Z. Luo, and B. Yuan, “Online nonnegative matrix factorization with robust stochastic approximation,” IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 7, pp. 1087–1099, Jul. 2012.

[33] P. de Handschutter and N. Gillis, “Deep orthogonal matrix factorization as a hierarchical clustering technique,” in Proc. 29th European Signal Processing Conference, Dublin, Ireland, 2021, Aug. 23–27, pp. 1466–1470.

[34] P. de Handschutter and N. Gillis, “A consistent and flexible framework for deep matrix factorizations,” Pattern Recognition, vol. 134, Feb. 2023.

[35] P. de Handschutter, N. Gillis, and X. Siebert, “A survey on deep matrix factorizations,” Computer Science Review, vol. 42, Nov. 2021.

[36] H. Wang, W. Yang, and N. Guan, “Cauchy sparse NMF with manifold regularization: A robust method for hyperspectral unmixing,” Knowledge-Based Systems, vol. 184, Nov. 2019.

[37] P. de Handschutter, N. Gillis, A. Vandaele, and X. Siebert, “Near-convex archetypal analysis,” IEEE Signal Processing Letters, vol. 27, pp. 81–85, Dec. 2020.

[38] B.-W. Chen and K.-L. Hou, “Half quadratic dual learning for fuzzy multiconcepts of partially-observed images,” IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 4, pp. 994–1007, Feb. 2022.

[39] Y. Zhou, D. Wilkinson, R. Schreiber, and R. Pan, “Large-scale parallel collaborative filtering for the Netflix prize,” in Proc. 4th International Conference on Algorithmic Applications in Management, Shanghai, China, 2008, Jun. 23–25, pp. 337–348.

[40] J. Wang, “Incomplete data analysis,” in Pattern Recognition, C. Travieso-Gonzalez, Ed. London, United Kingdom: IntechOpen, 2020.

[41] W. Ye, “Low-error data recovery based on collaborative filtering with nonlinear inequality constraints for manufacturing processes,” IEEE Transactions on Automation Science and Engineering, 2020.

[42] N. Seichepine, S. Essid, C. Févotte, and O. Cappé, “Soft nonnegative matrix co-factorization,” IEEE Transactions on Signal Processing, vol. 62, no. 22, pp. 5940–5949, Nov. 2014.

[43] N. Gillis and R. Luce, “Robust near-separable nonnegative matrix factorization using linear optimization,” Journal of Machine Learning Research, vol. 15, no. 1, pp. 1249–1280, Jan. 2014.

[44] K. Yoshii, K. Itoyama, and M. Goto, “Student's T nonnegative matrix factorization and positive semidefinite tensor factorization for single-channel audio source separation,” in

Proc. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016, Mar. 20–25, pp. 51–55.

[45] N. Nadisic, N. Gillis, and C. Kervazo, "Smoothed separable nonnegative matrix factorization," Linear Algebra and its Applications, vol. 676, pp. 174–204, Nov. 2023.

[46] N. Guan, L. Lan, D. Tao, Z. Luo, and X. Yang, "Transductive nonnegative matrix factorization for semi-supervised high-performance speech separation," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, 2014, May 04–09 pp. 2534–2538.

[47] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Non-negative patch alignment framework," IEEE Transactions on Neural Networks, vol. 22, no. 8, pp. 1218–1230, Aug. 2011.

Author's biography

Abdul bin Ismail is with School of Engineering and Computing, Manipal International University, Malaysia.