


OPEN ACCESS
EDITED BY

A. M. Elsayah,
Beijing Normal–Hong Kong Baptist
University, China

REVIEWED BY

Xintao Wu,
University of Arkansas, United States
Manjish Pal,
Indian Institute of Technology Kharagpur,
India

***CORRESPONDENCE**

Erin George
✉ e2george@ucsd.edu

RECEIVED 31 October 2025
 REVISED 14 February 2026
 ACCEPTED 23 February 2026
 PUBLISHED 17 March 2026

CITATION

Kassab L, George E, Needell D, Geng H,
Jafar Nia N and Li A (2026) Fairer
non-negative matrix factorization.
Front. Big Data 9:1737043.
doi: 10.3389/fdata.2026.1737043

COPYRIGHT

© 2026 Kassab, George, Needell, Geng,
Jafar Nia and Li. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Fairer non-negative matrix factorization

Lara Kassab¹, Erin George^{2*}, Deanna Needell³, Haowen Geng⁴,
Nika Jafar Nia⁵ and Aoxi Li⁶

¹Department of Mathematics, California State University, Fullerton, Fullerton, CA, United States,

²Department of Mathematics, University of California, San Diego, La Jolla, CA, United States,

³Department of Mathematics, University of California, Los Angeles, Los Angeles, CA, United States,

⁴McCormick School of Engineering, Northwestern University, Evanston, IL, United States, ⁵Department of Mathematics, Amherst College, Amherst, MA, United States, ⁶Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, CA, United States

There has been a recent critical need to study fairness and bias in machine learning (ML) algorithms. Since there is clearly no one-size-fits-all solution to fairness, ML methods should be developed alongside bias mitigation strategies that are practical and approachable to the practitioner. Motivated by recent work on “fair” PCA, here we consider the more challenging method of non-negative matrix factorization (NMF) as both a showcasing example and a method that is important in its own right for both topic modeling tasks and feature extraction for other ML tasks. We demonstrate that a modification of the objective function, by using a min-max formulation, may *sometimes* be able to offer an improvement in fairness for groups in the population. We derive two methods for the objective minimization, a multiplicative update rule as well as an alternating minimization scheme, and discuss implementation practicalities. We include a suite of synthetic and real experiments that show how the method may improve fairness while also highlighting the important fact that this may sometime increase error for some individuals and fairness is not a rigid definition and method choice should strongly depend on the application at hand.

KEYWORDS

dimensionality reduction, Fairer-NMF, fairness, non-negative matrix factorization, topic modeling

1 Introduction

Machine Learning (ML) and Artificial Intelligence (AI) have seen a huge surge in uses and applications in recent times and are being used in nearly every aspect of society. Despite this, there are serious and critical issues that propagate bias, disseminate unfair outcomes, and affect racial and social justice (CBS News, 2023; Ongweso, 2019; Truong, 2020). This lack of equity typically stems from many sources, from bias in the data to algorithmic bias and even post-processing decisions (Barocas et al., 2023; Mehrabi et al., 2021). In this work, we focus on inequities typically stemming from both the lack of fair representation of the data as well as algorithmic bias in the treatment of that data. ML techniques are used in a wide array of applications ranging from medical diagnostics and predictions to recidivism and sentencing decisions. At the heart of these applications lies the need for an algorithm to uncover hidden themes or features, that either explain some studied medical phenomenon or are used downstream for a learning task like classification or prediction. Because the data itself is often biased and ML methods tend to be designed to perform well *on average*, this can lead to misclassification and poorly explained patterns, particularly for underrepresented populations.

In this work, we focus on addressing this problem for topic modeling, or more generally, (unsupervised) dimensionality reduction tasks. We specifically consider non-negative matrix factorization (NMF), a powerful method used in a wide array of ML applications. It serves as a mathematically tangible example of a method for discovering data trends that highlight the inequities we wish to study (Lee and Seung, 2001, 1999). Indeed, at the core of NMF is a simple objective function that asks that the factorization has a small *average reconstruction error*. Because one only minimizes such an average, it is clear that population subgroups of small size may often be overshadowed (i.e. experience large reconstruction error) even when the total error is quite small relative to the population size. If NMF is being used to study a population's features and the desire is that *all* members or groups of the populations be represented in those features, this is clearly detrimental. Further, if NMF is being used for feature identification along with e.g. classification or some other ML task, this example would imply that those small subgroups experience incredible inaccuracy while the majority of the population benefits from accurate predictions. Especially when such a method is used without a fairness analysis, these may go unnoticed and cause extreme harm, such as is the case in medical, criminal justice, and many other applications.

This work aims to explore a fairer alternative objective function for NMF under a specific framework of fairness and present algorithmic implementations for solving this formulation. Before one can seek to promote fairness, it is essential to first define what fairness means. Defining fairness and fighting against bias and discrimination have existed long before the advent of machine learning (Saxena, 2019). Notably, definitions of fairness vary across tasks and settings and are generally non-universal which contributes to the difficulty of solving such problems (Barocas et al., 2023; Mehrabi et al., 2021). We emphasize here, as we do in the title, that our goal must be humble; we seek a *fairer—not fair*—formulation, and even that will only be fairer for certain contexts and applications. In fact, in some cases, applying a fairer method may also result in increased error (or decreased accuracy) for some individuals, meaning that the use of any method (“fairer” or not) should always be handled with care, and appropriate fairness metrics should be measured regardless. See Section 8 for more discussion on this and other related points. Nonetheless, we view this as an important step forward, which will hopefully lead to ML algorithms with more transparency and flexibility for the end user to identify and mitigate bias.

2 Contributions and organization

2.1 Contribution

We view the contributions of this work in several ways. First, we showcase how NMF, a method used for both transparent direct data analysis and as a precursor or proxy for interpretable feature extraction for other ML methods, may often produce inequitable outcomes. Next, we present a min-max mitigation strategy to improve fairness that is motivated by the so-called fair variant of principal component analysis (PCA) (Samadi et al., 2018), aimed

at mitigating bias stemming from group imbalance in size and complexity. Despite the added complications of non-negativity in the analysis, we derive two algorithms to solve the proposed fairer formulation of NMF: a multiplicative updates scheme as well as an alternating minimization scheme. Lastly and perhaps most importantly, we show through a suite of synthetic and real data experiments that there are settings where this formulation improves fairness, and that also there may be situations where it may not—depending on the desired form of fairness and application. We believe being able to provide this type of transparency is an absolutely critical first step toward a fairer world of ML and AI.

2.2 Organization

In Section 3, we present related works in fair unsupervised learning techniques, particularly dimensionality reduction. We provide an overview of NMF, its applications, and existing algorithms in Section 4. In Section 5, we further discuss the objective function of standard NMF at the group level and define the fairness criterion of our proposed NMF formulation called Fairer-NMF. Then, in Section 6, we present two algorithms for solving Fairer-NMF. Lastly, in Section 7, we present numerical experiments on synthetic and real data to demonstrate the performance of the algorithms.

2.3 Notation

We use boldfaced upper-case Latin letters (e.g., \mathbf{X}) to denote matrices and $\mathbf{X} \in \mathbb{R}_{\geq 0}^{m \times n}$ to denote an $m \times n$ matrix with real non-negative entries. The Frobenius norm of a matrix \mathbf{X} is denoted by $\|\mathbf{X}\|$. The notation \mathbf{A}/\mathbf{B} indicates entrywise division, $\mathbf{A} \odot \mathbf{B}$ entrywise multiplication, and \mathbf{AB} standard matrix multiplication. The notation $\mathbf{e}_\ell \in \mathbb{R}^L$ indicates the standard basis vector in \mathbb{R}^L where ℓ -th entry in the vector \mathbf{e} is 1 and all other entries are zero.

For a dataset partitioned into two (or more) mutually exclusive sample groups, we write the data matrix \mathbf{X} with m number of samples and n number of features in block format as,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix} \in \mathbb{R}_{\geq 0}^{m \times n} \quad \text{where } \mathbf{X}_A \in \mathbb{R}_{\geq 0}^{m_1 \times n}, \mathbf{X}_B \in \mathbb{R}_{\geq 0}^{m_2 \times n}.$$

The matrices \mathbf{X}_A and \mathbf{X}_B are constructed from the rows in \mathbf{X} corresponding to sample group A and B , respectively. In general, we write $|A|$ to denote the size of group A . Here, $|A| = m_1$ and $|B| = m_2$ with $m_1 + m_2 = m$. Additionally, the notation \mathbf{X}_A generally means we restrict to the rows of \mathbf{X} with sample indices given by A .

3 Related works

In this section, we focus on related work that is most pertinent to our study, rather than providing an exhaustive review. We provide more technical presentation of NMF in the section following.

3.1 Standard non-negative matrix factorization

Topic modeling is a machine learning technique used to reveal latent themes or patterns from large datasets. A popular technique for topic modeling that provides a low-rank approximation of a matrix is non-negative matrix factorization (NMF) (Lee and Seung, 1999, 2001). NMF has garnered increasing attention due to its effectiveness in handling large-scale data across various domains. In image processing, NMF is employed for tasks like feature extraction and perceptual hashing (Lee and Seung, 1999; Rajapakse et al., 2004; Tang et al., 2013). In the field of text mining, it has proven useful for document clustering and semantic analysis (Berry and Browne, 2005; Xu et al., 2003). In the medical field, it has been employed on applications ranging from fraud detection (Zhu et al., 2011) and phenotyping (Joshi et al., 2016) to studying trends in health and disease from record or survey data (Hamamoto et al., 2022; Hassaine et al., 2020; Johnson et al., 2024; Vendrow et al., 2020). Indeed, due to the non-negativity constraints, NMF acquires a parts-based, sparse representation of the data (Lee and Seung, 1999). When the features are naturally non-negative, this approach often enhances interpretability compared to traditional methods like Principal Components Analysis (PCA) (Lee and Seung, 1999).

3.2 Fair unsupervised learning

The topic of fairness in clustering has recently gained significant interest in the machine learning community with Chierichetti et al. (2017) leading the first work on fair clustering. Due to the difficulty in defining and enforcing fairness criteria for unsupervised learning tasks, including clustering techniques, many different fairness notions for clustering exist (Backurs et al., 2019; Chen et al., 2019; Ghadiri et al., 2021; Mahabadi and Vakilian, 2020). An overview of fair clustering is given in Chhabra et al. (2021).

Fairness issues in recommender systems have also recently attracted increasing attention, leading to the emergence of works aimed at mitigating bias (e.g., Li et al., 2021; Zhu et al., 2020). In Wang et al. (2023), the authors provide a survey on the fairness of recommender systems. Some works have proposed fairness-aware matrix factorizations for recommender systems (e.g., Togashi and Abe, 2022) including federated approaches (e.g., Liu et al., 2022). We note that the works of Togashi and Abe (2022) and Liu et al. (2022) differ substantially from ours. Both Togashi and Abe (2022) and Liu et al. (2022) are limited to recommendation and ranking contexts, focusing on fairness-aware collaborative filtering for item exposure fairness. They employ standard matrix factorization as a tool for building recommender systems. In contrast, our work directly incorporates group fairness into the NMF formulation while serving as a general dimensionality reduction technique for diverse ML applications, including topic modeling.

While the aforementioned works focus on fairness in recommendation contexts, fairness has also been studied in broader dimensionality reduction settings. In Buet-Golfouse and Utyagulov (2022), the authors investigate fairness of generalized low-rank models (GLRMs), including NMF, in unsupervised learning

settings. We remark that our fairness formulation differs from Buet-Golfouse and Utyagulov (2022), as we detail in Section 5.2.

3.3 Fair principal component analysis

In Samadi et al. (2018), the authors investigate how PCA might inadvertently introduce bias. The numerical experiments show that PCA incurs much higher average reconstruction error for one population than another (e.g., lower- versus higher-educated individuals), even when the populations are of similar sizes. The authors established a formulation called, *Fair PCA*, that addresses this bias under a specific framework of fairness. For a given matrix $Y \in \mathbb{R}^{a \times n}$ denote by $\hat{Y} \in \mathbb{R}^{a \times n}$ the optimal rank- d approximation of Y . Given $Z \in \mathbb{R}^{a \times n}$ with rank at most d , the reconstruction loss is defined as,

$$\text{loss}(Y, Z) := \|Y - Z\|_F^2 - \|Y - \hat{Y}\|_F^2.$$

Consider a data matrix $\begin{bmatrix} A \\ B \end{bmatrix} \in \mathbb{R}^{m \times n}$ where the rows in A and B are samples in the data belonging to a group A and B , respectively. The problem of finding a projection into d -dimensions in *Fair PCA* is defined as solving,

$$\min_{U \in \mathbb{R}^{m \times n}, \text{rank}(U) \leq d} \max \left\{ \frac{1}{|A|} \text{loss}(A, U_A), \frac{1}{|B|} \text{loss}(B, U_B) \right\}. \quad (1)$$

This criterion of Fair PCA, Equation 1, seeks to minimize the maximum of the average reconstruction loss across different groups which fits under what is known as the min-max framework or *social fairness*. In Tantipongpipat et al. (2019), the authors further introduce a multi-criteria dimensionality reduction problem, where multiple objectives are optimized simultaneously. One application of this model is capturing several fairness criteria in dimensionality reduction, such as the Fair PCA problem (Samadi et al., 2018).

Motivated by Samadi et al. (2018), here we ask whether a similar framework can be used for other linear algebraic-based ML approaches. In particular, we explore how the non-negativity constraints, which offer more interpretability than PCA but also introduce analytical challenges, can be adapted to such a framework.

4 Standard NMF formulation

Given a non-negative matrix $X \in \mathbb{R}_{\geq 0}^{m \times n}$ and a target dimension $r \in \mathbb{N}$, NMF decomposes X into a product of two low-dimensional non-negative matrices: $W \in \mathbb{R}_{\geq 0}^{m \times r}$ and $H \in \mathbb{R}_{\geq 0}^{r \times n}$, such that

$$X \approx WH$$

We consider X to be a data matrix where the rows represent data samples and the columns represent data features. Typically, $r > 0$ is chosen such that $r \ll \min\{m, n\}$ to reduce the dimension of the original data matrix or reveal hidden patterns in the data. The matrix W is called the *representation matrix* and H is called the *dictionary matrix*. The rows of H are generally referred to as

topics, which are characterized by features of the dataset. Each row of W provides the approximate representation of the respective row in X in the lower-dimensional space spanned by the rows of H . Thus, the data points are well approximated by an additive linear combination of the topics.

We note that in the NMF literature, r is referred to as the target dimension, the number of desired topics, or the desired non-negative rank. It is a user-specified hyperparameter and can be estimated heuristically. The *non-negative rank* of a matrix X is the smallest integer $r^* > 0$ such that there exists an exact NMF decomposition: $X = WH$ where $W \in \mathbb{R}_{\geq 0}^{m \times r^*}$ and $H \in \mathbb{R}_{\geq 0}^{r^* \times n}$. Computing the exact non-negative rank of a matrix is NP-hard (Vavasis, 2010). Therefore, several formulations for the non-negative approximation, $X \approx WH \in \mathbb{R}_{\geq 0}^{m \times n}$, have been studied (Cichocki et al., 2009; Lee and Seung, 1999, 2001) that seek to minimize the reconstruction error of the decomposition.

Definition 4.1 (Relative reconstruction error). Suppose $X \in \mathbb{R}_{\geq 0}^{m \times n}$ and $r < \min\{m, n\} \in \mathbb{N}$. For a given $W \in \mathbb{R}_{\geq 0}^{m \times r}$ and $H \in \mathbb{R}_{\geq 0}^{r \times n}$ we define the reconstruction error of X as $\|X - WH\|$ and the relative reconstruction error of X as $\|X - WH\|/\|X\|$.

One of the most popular formulations of finding an NMF approximation uses the Frobenius norm as a measure of the reconstruction error,

$$\operatorname{argmin}_{W \in \mathbb{R}_{\geq 0}^{m \times r}, H \in \mathbb{R}_{\geq 0}^{r \times n}} \|X - WH\|^2. \tag{2}$$

Throughout the paper, we refer to this formulation as *rank- r NMF* or *standard NMF with rank r* . For simplicity and as common in the literature of NMF, we will refer to non-negative rank simply as rank.

Many numerical optimization techniques can be applied to find local minima for the NMF problem defined in Equation 2 (e.g., Cichocki et al., 2009; Kim et al., 2008; Kim and Park, 2008; Lin, 2007; Paatero and Tapper, 1994). Note that although Equation 2 is a non-convex optimization problem, it is convex in W when H is held fixed and vice-versa. Thus, an *alternating minimization* (AM) approach (see e.g., Bertsekas, 1997) can be used to find local minima:

$$W^{(k)} \leftarrow \operatorname{argmin}_{W \in \mathbb{R}_{\geq 0}^{m \times r}} \|X - WH^{(k)}\|^2$$

$$H^{(k)} \leftarrow \operatorname{argmin}_{H \in \mathbb{R}_{\geq 0}^{r \times n}} \|X - W^{(k)}H\|^2$$

where k denotes the k -th iteration. Both of these convex problems are non-negative least squares problems, and specialized solvers exist to find solutions.

Another minimization method is the *multiplicative updates* (MU) method proposed in Lee and Seung (2001). The method can be viewed as an entrywise projected gradient descent algorithm. The choice of stepsize for each entry of the updating matrix results in multiplicative (rather than additive) update rules that ensure non-negativity. The algorithm performs alternating steps in

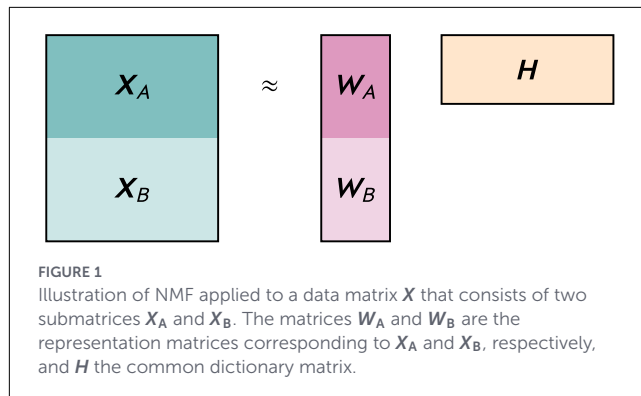


FIGURE 1
Illustration of NMF applied to a data matrix X that consists of two submatrices X_A and X_B . The matrices W_A and W_B are the representation matrices corresponding to X_A and X_B , respectively, and H the common dictionary matrix.

updating W and H :

$$W^{(k)} \leftarrow W^{(k-1)} \odot \frac{XH^{(k-1)\top}}{W^{(k-1)}H^{(k-1)}H^{(k-1)\top}}$$

$$H^{(k)} \leftarrow H^{(k-1)} \odot \frac{W^{(k)\top}X}{W^{(k)\top}W^{(k)}H^{(k-1)}}$$

The multiplicative updates algorithm is commonly used due to its ease of implementation, the absence of the need for user-defined hyperparameters, and desirable monotonicity properties (Lee and Seung, 2001).

We comment here that the recent work (Gillis et al., 2021) has developed a distributionally robust algorithm for multi-objective NMF using multiplicative updates. The derivation of their MU scheme follows similarly to ours (in Section 6.3), and in fact, it would be interesting future work to incorporate both our fairness objective as well as their robust framework.

5 Fairer-NMF formulation

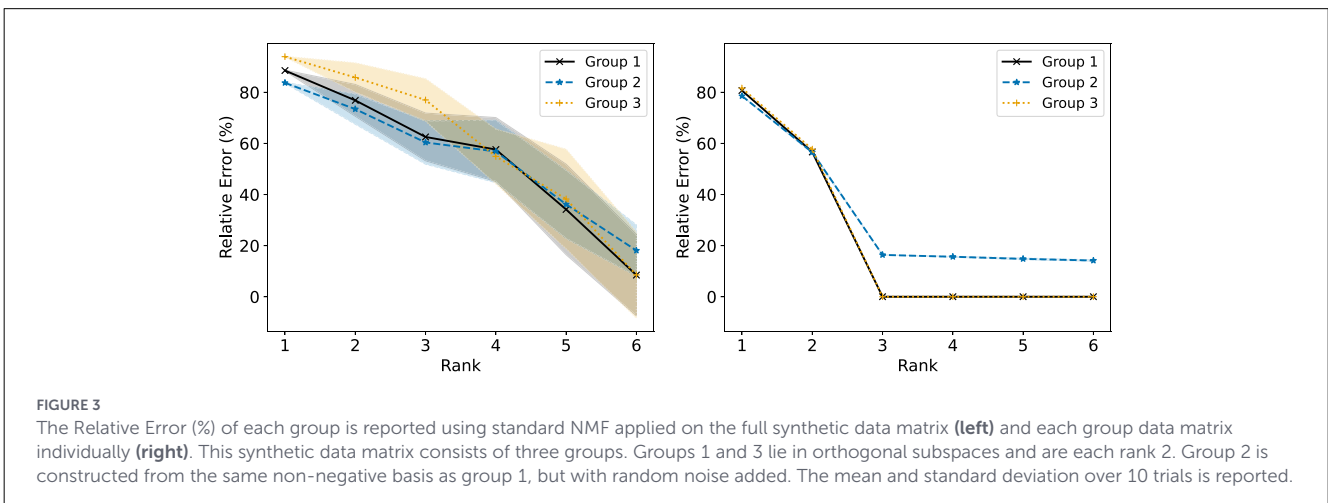
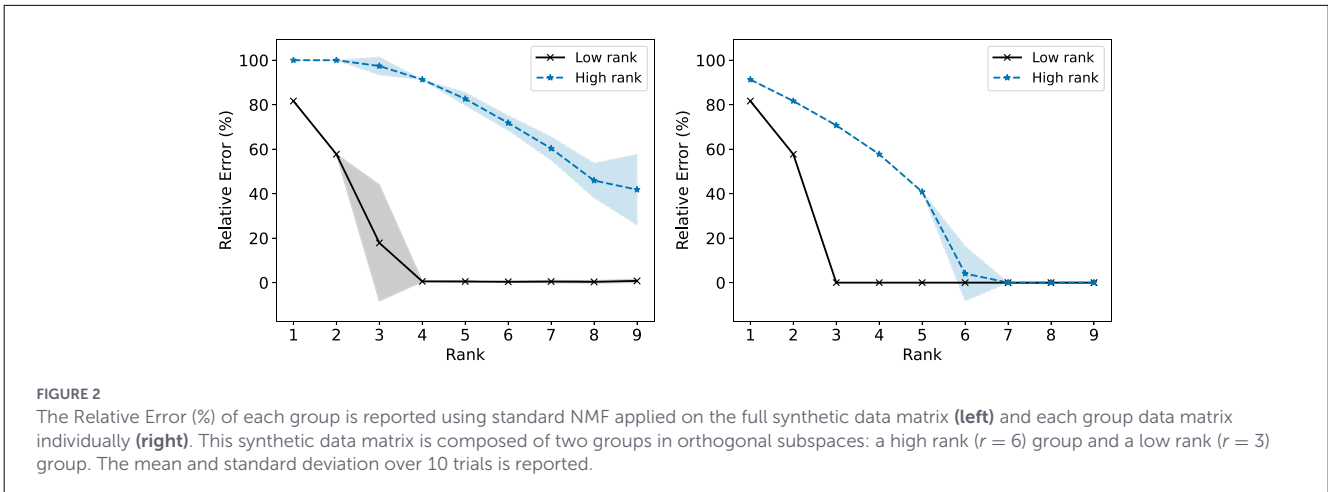
In this section, we provide a technical presentation of standard NMF, highlight characteristics of its objective function at the group level, and propose our approach, Fairer-NMF.

5.1 Standard NMF at the group level

Suppose a dataset consists of two mutually exclusive groups A (with size $|A| = m_1$) and B (with size $|B| = m_2$). For example, these groups could be divided based on a protected attribute in the data. We can write the NMF of a data matrix $X \in \mathbb{R}_{\geq 0}^{m \times n}$ as,

$$X = \begin{bmatrix} X_A \\ X_B \end{bmatrix} \approx \begin{bmatrix} W_A \\ W_B \end{bmatrix} H \tag{3}$$

where $W_A \in \mathbb{R}_{\geq 0}^{m_1 \times r}$ is the representation matrix corresponding to X_A , $W_B \in \mathbb{R}_{\geq 0}^{m_2 \times r}$ is the representation matrix corresponding to X_B , and $H \in \mathbb{R}_{\geq 0}^{r \times n}$ is the common dictionary matrix. An illustration of the decomposition is given in Figure 1.



In the standard NMF, defined in Equation 2, we have:

$$\begin{aligned} & \operatorname{argmin}_{W \in \mathbb{R}_{\geq 0}^{m \times r}, H \in \mathbb{R}_{\geq 0}^{r \times n}} \|X - WH\|^2 \\ &= \operatorname{argmin}_{\substack{W_A \in \mathbb{R}_{\geq 0}^{m_1 \times r}, W_B \in \mathbb{R}_{\geq 0}^{m_2 \times r} \\ H \in \mathbb{R}_{\geq 0}^{r \times n}}} (\|X_A - W_A H\|^2 + \|X_B - W_B H\|^2). \end{aligned}$$

Note that both groups are weighted equally and we seek to minimize the sum of the reconstruction error of each group in the joint decomposition. The problem can be written for L mutually exclusive groups,

$$\operatorname{argmin}_{W \in \mathbb{R}_{\geq 0}^{m \times r}, H \in \mathbb{R}_{\geq 0}^{r \times n}} \|X - WH\|^2 = \operatorname{argmin}_{\substack{W_\ell \in \mathbb{R}_{\geq 0}^{m_\ell \times r}, \forall \ell \in \{1, \dots, L\} \\ H \in \mathbb{R}_{\geq 0}^{r \times n}}} \sum_{\ell=1}^L \|X_\ell - W_\ell H\|^2. \tag{4}$$

This standard objective function is designed to perform well on average. It seeks an overall low reconstruction error which disregards the size and complexity of each data group. For example, in the case of an imbalanced dataset where $|A| \gg |B|$ an overall

low reconstruction error does not guarantee that the reconstruction error restricted to group B is low. Additionally, the standard objective function does not pay attention to the complexity of the data groups.

We consider now some illustrative experiments. Figures 2, 3 show two different experiments with synthetic data to highlight two different situations where standard NMF can produce “unfair” results. In all of these experiments, there exist multiple groups such that we wish to obtain a factorization that works well for all groups simultaneously. In Figure 2, one of these groups is lower rank than the other. In this case, we see that the decompositions perform much better on the lower rank group ($r = 3$) than the higher rank group ($r = 6$). In particular, the reconstructions for rank 6 and above have large error for the high rank group, even though this group is rank 6. In Figure 3, we have three different groups of same size and two of them (groups 1 and 2) lie in approximately the same data subspace. Here, groups 1 and 2 typically have better reconstruction than group 3 for low rank decompositions. While all groups have the same size and magnitude, sharing a similar basis introduces an imbalance in size (thus magnitude) in the full dataset. This is a also common scenario in real-world settings that is important to consider. In both Figures 1, 2, we see that good low rank reconstructions of the dataset are possible when standard

NMF is applied to each group individually. The full details of these experiments, including the details for generating the synthetic data, are in Section 7.2.

Minimizing the maximum reconstruction error may seem desirable, but this approach favors optimizing for the group with inherently higher complexity when the group sizes and magnitudes are all the same. Similarly, this approach favors optimizing for the group with the largest size and magnitude when the rank is the same for all groups. In the next section, we present a fairness criterion that takes into account the size and complexity of the data groups.

5.2 Fairer-NMF objective

In our consideration of a fairer NMF, we work under a min-max fairness framework with a criterion similar to that of Fair-PCA (Samadi et al., 2018). We consider the following definition of reconstruction loss of a group.

Definition 5.1 (Relative reconstruction loss). *Suppose we have a group data matrix $X_\ell \in \mathbb{R}_{\geq 0}^{m_\ell \times n}$, desired rank $r \in \mathbb{N}$, and E_ℓ representing the error obtained by replacing X_ℓ with a r -rank NMF approximation. For a given $W_\ell \in \mathbb{R}^{m_\ell \times r}$ and $H \in \mathbb{R}_{\geq 0}^{r \times n}$, we define the relative reconstruction loss of X_ℓ as,*

$$\frac{\|X_\ell - W_\ell H\| - E_\ell}{\|X_\ell\|}.$$

Remark 5.2. *A reasonable choice for E_ℓ is to take*

$$E_\ell = \mathbb{E}_{X_\ell^*} \|X_\ell - X_\ell^*\|$$

where the expectation is taken over the X_ℓ^* obtained from a specific randomized implementation of rank- r NMF on X_ℓ .

Remark 5.3. *Note here and throughout that we assume a priori knowledge of the L population subgroups. This allows us to focus on the fairness objective here, and while it is reasonable in many settings, we remark on possible relaxations of this assumption in Section 8.*

We provide details on the numerical approximation in Section 6. We again highlight that finding an optimal non-negative rank- r NMF of a matrix $X_\ell \in \mathbb{R}_{\geq 0}^{m_\ell \times n}$ is NP-hard (Vavasis, 2010). Taking the expectation over a randomized NMF implementation allows E_ℓ to compensate both for the underlying dimensionality of the group and for how difficult it is for the NMF algorithm to find a good representation for the group.

Roughly speaking, the reconstruction loss is the difference between how well group X_ℓ is reconstructed via a standard NMF model trained on the entire data X and a standard NMF model trained on X_ℓ only. We further note that we normalize by the Frobenius norm of the group matrix to mitigate for the size of the group and the actual varying magnitudes of $\|X_\ell\|$.

We emphasize that our goal is to learn a common non-negative low-rank NMF model for all groups, rather than separate NMF models for each group. In Fairer-NMF, we seek to minimize the maximum of the average reconstruction loss across L different groups:

$$\min_{W \in \mathbb{R}_{\geq 0}^{m \times r}, H \in \mathbb{R}_{\geq 0}^{r \times n}} \max_{\ell \in \{1, \dots, L\}} \left\{ \frac{\|X_\ell - W_\ell H\| - E_\ell}{\|X_\ell\|} \right\} \quad (5)$$

First, we note that the second term in the objective function E_ℓ is a fixed pre-computed constant for a given group ℓ as defined in Remark 5.2. This optimization problem seeks to learn a common NMF model that minimizes the maximum deviation from the group-wise “optimal” NMF approximation across all groups.

Second, we note that the loss function in Equation 5 is incomplete. We wish to select H that minimizes the loss:

$$\max_{\ell \in \{1, \dots, L\}} \left\{ \frac{\|X_\ell - W_\ell H\| - E_\ell}{\|X_\ell\|} \right\}.$$

However, in general, W is under-specified. We can freely perturb W_ℓ for any group that does not attain the maximum loss and achieves another solution of Equation 5, provided the loss for group ℓ does not grow too large. This is because the group representation matrices W_ℓ all act independently of each other. To resolve this issue, we choose

$$\min_{W_\ell \in \mathbb{R}_{\geq 0}^{m_\ell \times r}} \|X_\ell - W_\ell H\|$$

for all ℓ independently of each other, which is equivalent to choosing

$$\min_{W \in \mathbb{R}_{\geq 0}^{m \times r}} \|X - WH\|. \quad (6)$$

We remark the optimal W for a fixed H may give different losses for different groups even if H is the exact minimum given by Equation 5. While this may result in one group having a lower loss than another, this inequality is “free” in the sense that it does not come at the expense of another group. Therefore, in Fairer-NMF we seek to minimize both Equations 5, 6, prioritizing the first over the second.

Referring back to Section 3.2, Buet-Golfouse and Utyagulov (2022) consider fairness criteria that aim to reduce disparity across group-wise average costs through a penalized learning approach. Their disparity penalty is defined as the difference between each group’s average loss and the overall average loss across all groups—a constant term shared by all groups. In contrast, we formulate fairness as minimizing the maximum group-wise disparity, where each group’s performance is compared against the standard NMF objective evaluated on that group alone. This distinction leads to different fairness criteria and trade-offs. Additionally, we solve a constrained optimization problem with hard non-negativity constraints, whereas Buet-Golfouse and Utyagulov (2022) incorporate non-negativity through penalty terms in an unconstrained formulation.

6 Algorithms

We present two algorithms for solving the Fairer-NMF problem formulation.

6.1 Estimating E_ℓ

As discussed in Remark 5.2, for a group ℓ , a reasonable choice for the estimate of the optimal rank- r error, E_ℓ , is to take the

```

Input data matrix  $\mathbf{X}_\ell \in \mathbb{R}_{\geq 0}^{m_\ell \times n}$  of group  $\ell$ ; desired dimension  $r \in \mathbb{N}$ ; number of estimates  $T \geq 1$ 
 $E \leftarrow 0$ 
for  $t \leftarrow 1$  to  $T$  do
     $\mathbf{W}_\ell, \mathbf{H}_\ell \leftarrow \underset{\substack{\mathbf{W} \in \mathbb{R}_{\geq 0}^{m_\ell \times r}, \mathbf{H} \in \mathbb{R}_{\geq 0}^{r \times n}}{\operatorname{argmin}}}{\|\mathbf{X}_\ell - \mathbf{W}\mathbf{H}\|}$  ▷ Approximate using standard NMF
     $E \leftarrow E + \|\mathbf{X}_\ell - \mathbf{W}_\ell \mathbf{H}_\ell\|$ 
end for
 $E_\ell \leftarrow E/T$ 
Output rank- $r$  error  $E_\ell$ 
    
```

Algorithm 1. Estimating E_ℓ .

expectation over the error obtained from a specific randomized implementation of rank- r NMF for the group. This leads to a natural algorithm for estimating E_ℓ . It suffices to sample the single group NMF reconstruction T times and take the average. This is described in Algorithm 1.

A consequence of this choice in algorithm is that the relative reconstruction loss for a group may be negative. Using a specific randomized algorithm for NMF and taking the expectation of the reconstruction error provides an upper bound for the group reconstruction error, not a lower bound. When a factorization for a group is then obtained using a different algorithm, such as Fairer-NMF, there is no reason in general to expect this error to always be worse than our estimate for E_ℓ . While counterintuitive, this is not a problem for the Fairer-NMF algorithm. By directly choosing E_ℓ using reconstructions for just the single group, a negative loss will only occur when the presence of other groups does not degrade the reconstruction for a single group.

6.2 Alternating minimization (AM) Scheme

The optimization problem in Equation 5 is non-convex with respect to \mathbf{H} and \mathbf{W}_ℓ for all ℓ . However, it is convex with respect to one of the factor matrices while all others are held fixed. Further, the corresponding constraint sets are convex. This allows us to solve the problem using an AM approach on a multi-convex problem as outlined in Algorithm 2. The AM scheme solves a convex problem in each minimization step, ensuring that there is a global minimum. Solving for $\mathbf{H}^{(k)}$ and $\mathbf{W}^{(k)}$ in Algorithm 2 is similar to solving a standard NMF problem using the AM approach.

As remarked in Equation 4, the update rule of $\mathbf{W}^{(k)}$ in Algorithm 2 is equivalent to updating the representation matrix of each group ℓ as

$$\mathbf{W}_\ell^{(k)} = \underset{\substack{\mathbf{W}_\ell \in \mathbb{R}_{\geq 0}^{m_\ell \times r}}{\operatorname{argmin}}}{\|\mathbf{X}_\ell - \mathbf{W}_\ell \mathbf{H}^{(k)}\|}.$$

Consider the function f defined as

$$f(\mathbf{W}^{(k)}, \mathbf{H}^{(k)}) = \max_{\ell \in \{1, \dots, L\}} \frac{\|\mathbf{X}_\ell - \mathbf{W}_\ell^{(k)} \mathbf{H}^{(k)}\| - E_\ell}{\|\mathbf{X}_\ell\|}, \quad (7)$$

where $\mathbf{H}^{(k)}$ and $\mathbf{W}^{(k)}$ defined in Algorithm 2. Indeed,

$$\mathbf{H}^{(k)} \leftarrow \underset{\substack{\mathbf{H} \in \mathbb{R}_{\geq 0}^{r \times d}}{\operatorname{argmin}}}{f(\mathbf{W}^{(k)}, \mathbf{H})}$$

is a convex optimization problem. Then, we have that the loss function f is non-increasing $f(\mathbf{W}^{(k)}, \mathbf{H}^{(k)}) \leq f(\mathbf{W}^{(k-1)}, \mathbf{H}^{(k-1)})$ where equality is achieved at a stationary point. Thus, by iterating the updates of $\mathbf{H}^{(k)}$ and $\mathbf{W}^{(k)}$, we obtain a sequence of estimates whose loss values converge.

The two optimization functions in Algorithm 2 both fall under restricted classes of convex programs that admit specialized solvers. The problem

$$\min_{\substack{\mathbf{H} \in \mathbb{R}_{\geq 0}^{r \times d}}} \max_{\ell \in \{1, \dots, L\}} \frac{\|\mathbf{X}_\ell - \mathbf{W}_\ell^{(k-1)} \mathbf{H}\| - E_\ell}{\|\mathbf{X}_\ell\|}$$

is equivalent to

$$\begin{aligned} & \min_{\substack{t \\ \mathbf{H} \in \mathbb{R}_{\geq 0}^{r \times d} \\ t \in \mathbb{R}}} t \\ & \text{subj. to } \|\mathbf{X}_\ell - \mathbf{W}_\ell^{(k-1)} \mathbf{H}\| \leq t \|\mathbf{X}_\ell\| + E_\ell \quad \forall \ell \in \{1, \dots, L\}, \end{aligned}$$

which is a second-order cone program (SOCP). The minimization problem for \mathbf{W} is equivalent to

$$\min_{\substack{\mathbf{W} \in \mathbb{R}_{\geq 0}^{n \times r}}{\|\mathbf{X} - \mathbf{W}\mathbf{H}^{(k)}\|^2},$$

which is a non-negative least squares (NNLS) problem, a specific type of quadratic program (QP).

6.3 Multiplicative updates (MU) scheme

In addition to the AM scheme, we also adapt the multiplicative updates (MU) scheme for the Fairer-NMF problem formulation. Consider an equivalent form to the loss function in Equation 7:

$$\begin{aligned} f(\mathbf{W}^{(k)}, \mathbf{H}^{(k)}) &= \max_{\substack{\mathbf{c} \in \mathbb{R}^L \\ \|\mathbf{c}\|_1 = 1}} g(\mathbf{W}^{(k)}, \mathbf{H}^{(k)}, \mathbf{c}) \\ &:= \sum_{\ell=1}^L \mathbf{c}_\ell \frac{\|\mathbf{X}_\ell - \mathbf{W}_\ell^{(k)} \mathbf{H}^{(k)}\| - E_\ell}{\|\mathbf{X}_\ell\|}. \end{aligned}$$

Let $\mathbf{c}^{(k)}$ be our estimate of the maximizer \mathbf{c} and consider an alternating approach by maximizing the function $g(\mathbf{W}, \mathbf{H}, \mathbf{c})$ in \mathbf{c} and minimizing in \mathbf{W} and \mathbf{H} . The maximizer $\mathbf{c}^{(k)}$ at the k -th iteration is simply \mathbf{e}_{ℓ_*} (the ℓ_* -th standard basis vector) where,

$$\ell_* = \operatorname{argmax}_{\ell \in \{1, \dots, L\}} \frac{\|\mathbf{X}_\ell - \mathbf{W}_\ell^{(k-1)} \mathbf{H}^{(k-1)}\| - E_\ell}{\|\mathbf{X}_\ell\|}.$$

Input data matrix $X = \begin{bmatrix} X_1 \in \mathbb{R}_{\geq 0}^{m_1 \times n} \\ \vdots \\ X_L \in \mathbb{R}_{\geq 0}^{m_L \times n} \end{bmatrix}$ with L groups; desired dimension $r \in \mathbb{N}$

Compute E_ℓ for each group ℓ ▷ e.g., through [Algorithm 1](#)

Randomly initialize $W^{(0)} = \begin{bmatrix} W_1^{(0)} \in \mathbb{R}_{\geq 0}^{m_1 \times r} \\ \vdots \\ W_L^{(0)} \in \mathbb{R}_{\geq 0}^{m_L \times r} \end{bmatrix}$

$k \leftarrow 0$

while not converged **do**

$k \leftarrow k + 1$

$H^{(k)} \leftarrow \operatorname{argmin}_{H \in \mathbb{R}_{\geq 0}^{r \times d}} \max_{\ell \in \{1, \dots, L\}} \frac{\|X_\ell - W_\ell^{(k-1)} H\| - E_\ell}{\|X_\ell\|}$

$W^{(k)} = \begin{bmatrix} W_1^{(k)} \\ \vdots \\ W_L^{(k)} \end{bmatrix} \leftarrow \operatorname{argmin}_{W \in \mathbb{R}_{\geq 0}^{n \times r}} \|X - WH^{(k)}\|$

end while

Output coefficient matrix $W^{(k)} = \begin{bmatrix} W_1^{(k)} \in \mathbb{R}_{\geq 0}^{m_1 \times r} \\ \vdots \\ W_L^{(k)} \in \mathbb{R}_{\geq 0}^{m_L \times r} \end{bmatrix}$; dictionary matrix $H^{(k)} \in \mathbb{R}_{\geq 0}^{r \times n}$

Algorithm 2. Fairer-NMF: alternating minimization scheme.

Start with $c^{(0)} = \mathbf{0}$. By setting $c^{(k)} = \frac{k-1}{k}c^{(k-1)} + \frac{1}{k}e_{\ell_*}$ we update with a decreasing step size while ensuring $\|c^{(k)}\|_1 = 1$. This is desirable as just setting $c^{(k)} = e_{\ell_*}$ can result in too much oscillation in the largest loss group and therefore poor convergence. This is demonstrated in [Figures 4, 5](#), using the synthetic data discussed in detail in [Section 7.2](#). Exactly optimizing c results in an algorithm which does not converge to a low loss solution.

For fixed $c^{(k)}$ and $W^{(k-1)}$, we minimize g in H by selecting

$$\operatorname{argmin}_{H \in \mathbb{R}_{\geq 0}^{r \times d}} \sum_{\ell=1}^L c_\ell^{(k)} \frac{\|X_\ell - W_\ell^{(k-1)} H\|}{\|X_\ell\|}. \tag{8}$$

A related but easier problem to solve is

$$\operatorname{argmin}_{H \in \mathbb{R}_{\geq 0}^{r \times d}} \sum_{\ell=1}^L (c_\ell^{(k)})^2 \frac{\|X_\ell - W_\ell^{(k-1)} H\|^2}{\|X_\ell\|^2} = \operatorname{argmin}_{H \in \mathbb{R}_{\geq 0}^{r \times d}} \|\tilde{X} - \tilde{W}H\|^2 \tag{9}$$

for the following two block matrices:

$$\tilde{X} = \begin{bmatrix} c_1^{(k)} X_1 / \|X_1\| \in \mathbb{R}_{\geq 0}^{m_1 \times n} \\ \vdots \\ c_L^{(k)} X_L / \|X_L\| \in \mathbb{R}_{\geq 0}^{m_L \times n} \end{bmatrix},$$

$$\tilde{W} = \begin{bmatrix} c_1^{(k)} W_1^{(k-1)} / \|X_1\| \in \mathbb{R}_{\geq 0}^{m_1 \times r} \\ \vdots \\ c_L^{(k)} W_L^{(k-1)} / \|X_L\| \in \mathbb{R}_{\geq 0}^{m_L \times r} \end{bmatrix}.$$

The solution of [Equation 9](#) attains a value of $g(W^{(k)}, H, c^{(k)})$ that is at most \sqrt{L} that of [Equation 8](#) that can be considered as a good choice of minimizer. Thus, we can select $H^{(k)}$ according to the standard multiplicative update for \tilde{X} with \tilde{W} . Once we've

obtained $H^{(k)}$, we can use the multiplicative update for the single group NMF problem $X_\ell \approx W_\ell H$ to obtain $W_\ell^{(k)}$ for each group ℓ . This procedure is described in [Algorithm 3](#).

We make two concluding remarks about this algorithm:

1. Enforcing $\|c^{(k)}\|_1 = 1$ is actually unnecessary, as the multiplicative update is invariant under scalar multiplication for \tilde{X} and \tilde{W} . Therefore it suffices to just set $c^{(k)} = c^{(k-1)} + e_{\ell_*}$.
2. While a learning rate is empirically important for convergence, the exact rate chosen was selected for the simplicity of its implementation and to avoid introducing a hyperparameter which needs to be tuned. It is likely that other update rules would yield similarly good performance.

7 Numerical experiments

In this section, we evaluate the performance of the AM scheme ([Algorithm 2](#)) and the MU scheme ([Algorithm 3](#)) for Fairer-NMF. We compare them with standard NMF on both synthetic benchmark datasets and real-world datasets, including survey and text data, which are common use cases for NMF. Our code is publicly available.¹

7.1 Reporting metrics and implementations

For all the experiments, we compute the relative reconstruction error given in [Definition 4.1](#) and the relative reconstruction loss

¹ <https://github.com/ErinGeorge/Fairer-NMF>

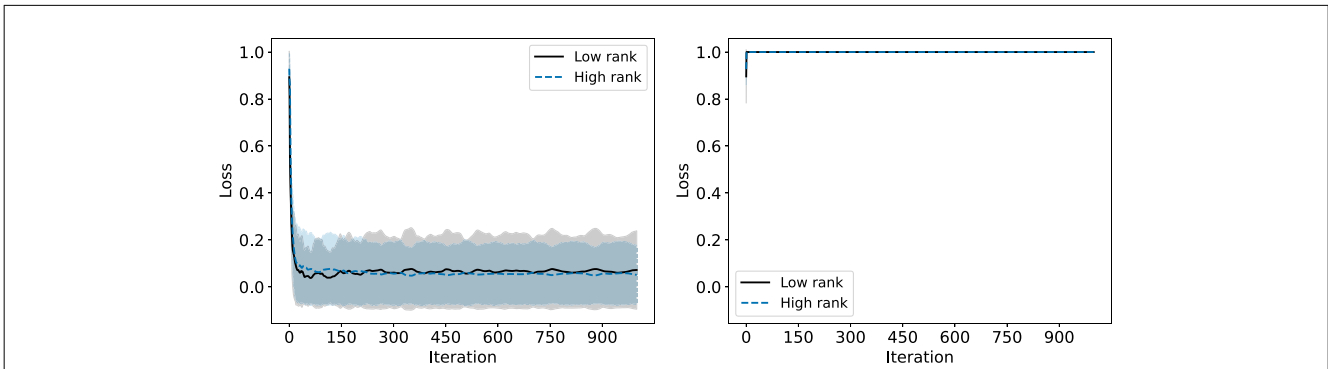


FIGURE 4 Convergence of the multiplicative update rule for Fairer-NMF on a synthetic data matrix with two groups. The multiplicative update rule is compared when using the decreasing step size for **c** (left) and when exactly optimizing **c** (right). The relative reconstruction loss of each group is reported per iteration with the mean and standard deviation taken over 100 trials.

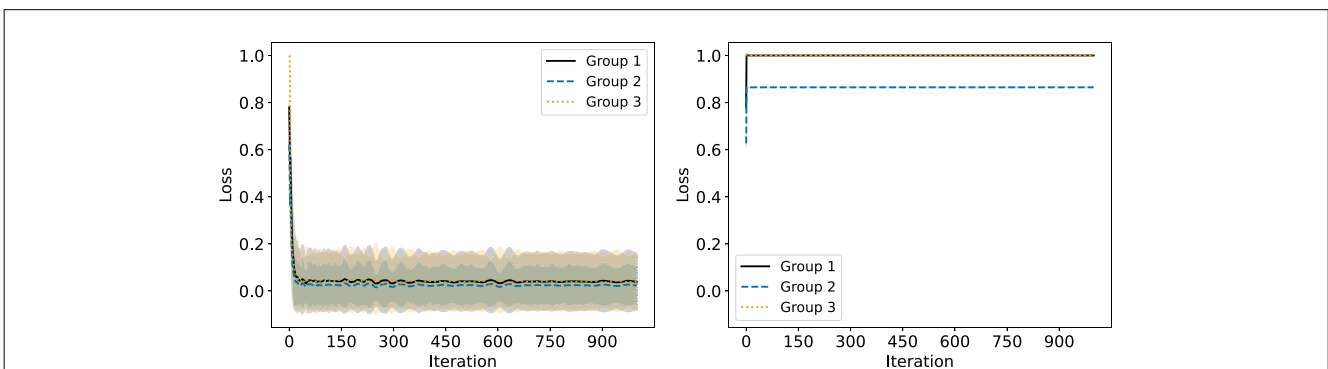


FIGURE 5 Convergence of the multiplicative update rule for Fairer-NMF on a synthetic data matrix with three groups. The multiplicative update rule is compared when using the decreasing step size for **c** (left) and when exactly optimizing **c** (right). The relative reconstruction loss of each group is reported per iteration with the mean and standard deviation taken over 100 trials.

given in Definition 5.1. In the figures, “Relative Error (%)” is the relative reconstruction error scaled by 100. We report the mean and standard deviation (given by the shaded region around the mean) over 10 trials. In each trial, we re-initialize the algorithm with a random initialization and estimate E_ℓ (defined in Remark 5.2) using Algorithm 1 with 5 runs ($T = 5$). We use the acronym R-Error to denote the average relative reconstruction error, and R-Loss for the average relative reconstruction loss.

For the AM scheme implementation, we use the open-source package CVXPY (Diamond and Boyd, 2016), which supports a number of different specialized convex program solvers. The specific solvers we use are ECOS (Domahidi et al., 2013) and SCS (O’Donoghue et al., 2016) for the SOCP to find \mathbf{H} and OSQP (Stellato et al., 2020) for the QP to find \mathbf{W} . For the SOCP, we default to ECOS and switch to SCS only if ECOS fails. Failures of ECOS do occur, but only for large problem sizes.

For both AM and MU schemes, we iterate until the change in each group’s reconstruction error in a single iteration is no more than 10^{-4} times the current reconstruction error. That is when the following condition is met for all groups ℓ :

$$\frac{\|\mathbf{X}_\ell - \mathbf{W}_\ell^{(k)} \mathbf{H}^{(k)}\| - \|\mathbf{X}_\ell - \mathbf{W}_\ell^{(k-1)} \mathbf{H}^{(k-1)}\|}{\|\mathbf{X}_\ell - \mathbf{W}_\ell^{(k)} \mathbf{H}^{(k)}\|} < 10^{-4}. \quad (10)$$

This stopping criterion works when no group is able to be reconstructed exactly. In the synthetic datasets we describe later in this section, this is not true. For these datasets, we stop if for every group ℓ either Equation 10 holds or the R-Error is less than 0.1%. The latter only happens for high ranks in our synthetic datasets, when the data can be perfectly reconstructed with a factorization at a given rank.

For all datasets, before running standard NMF or Fairer-NMF, we normalize the features to have unit ℓ_2 -norm.

7.2 Synthetic datasets

For the synthetic datasets, we generate group ℓ data matrix $\mathbf{X}_\ell \in \mathbb{R}_{\geq 0}^{500 \times 12}$ as $\mathbf{X}_\ell = \mathbf{W}_\ell \mathbf{H}_\ell$ where $\mathbf{W}_\ell \in \mathbb{R}_{\geq 0}^{500 \times r_\ell}$ and $\mathbf{H}_\ell \in \mathbb{R}_{\geq 0}^{r_\ell \times 12}$. We sample the rows of \mathbf{W}_ℓ independently and uniformly from the set $\{\mathbf{e}_1^T, \dots, \mathbf{e}_{r_\ell}^T\}$ (here each standard basis vector $\mathbf{e}_i \in \mathbb{R}^{r_\ell}$). That is, we select the rows of \mathbf{X}_ℓ to be independently and uniformly randomly chosen rows of \mathbf{H}_ℓ .

7.2.1 Synthetic dataset 1: different group ranks

For the first synthetic dataset, we have two groups and take r_ℓ and \mathbf{H}_ℓ to differ for the two groups:

Input data matrix $X = \begin{bmatrix} X_1 \in \mathbb{R}_{\geq 0}^{m_1 \times n} \\ \vdots \\ X_L \in \mathbb{R}_{\geq 0}^{m_L \times n} \end{bmatrix}$ with L groups; desired dimension $r \in \mathbb{N}$

Compute E_ℓ for each group ℓ ▷ e.g., through Algorithm 1

Randomly initialize $H^{(0)} \in \mathbb{R}_{\geq 0}^{r \times n}$, $W^{(0)} = \begin{bmatrix} W_1^{(0)} \in \mathbb{R}_{\geq 0}^{m_1 \times r} \\ \vdots \\ W_L^{(0)} \in \mathbb{R}_{\geq 0}^{m_L \times r} \end{bmatrix}$

Initialize $c^{(0)} = \theta \in \mathbb{R}^L$

$k \leftarrow 0$

while not converged **do**

$k \leftarrow k + 1$

$\ell_* \leftarrow \operatorname{argmax}_{\ell \in \{1, \dots, L\}} \frac{\|X_\ell - W_\ell^{(k-1)} H^{(k-1)}\| - E_\ell}{\|X_\ell\|}$

$c^{(k)} \leftarrow c^{(k-1)} + e_{\ell_*}$ ▷ e_{ℓ_*} is ℓ_* -th standard basis vector

$\tilde{X} \leftarrow \begin{bmatrix} c_1^{(k)} X_1 / \|X_1\| \in \mathbb{R}_{\geq 0}^{m_1 \times n} \\ \vdots \\ c_L^{(k)} X_L / \|X_L\| \in \mathbb{R}_{\geq 0}^{m_L \times n} \end{bmatrix}$, $\tilde{W} \leftarrow \begin{bmatrix} c_1^{(k)} W_1^{(k-1)} / \|X_1\| \in \mathbb{R}_{\geq 0}^{m_1 \times r} \\ \vdots \\ c_L^{(k)} W_L^{(k-1)} / \|X_L\| \in \mathbb{R}_{\geq 0}^{m_L \times r} \end{bmatrix}$

$H^{(k)} \leftarrow H^{(k-1)} \odot \frac{\tilde{W}^T \tilde{X}}{\tilde{W}^T \tilde{W} H^{(k-1)}}$

$W^{(k)} = \begin{bmatrix} W_1^{(k)} \\ \vdots \\ W_L^{(k)} \end{bmatrix} \leftarrow W^{(k-1)} \odot \frac{X H^{(k)T}}{W^{(k-1)} H^{(k)} H^{(k)T}}$

end while

Output coefficient matrix $W^{(k)} = \begin{bmatrix} W_1^{(k)} \in \mathbb{R}_{\geq 0}^{m_1 \times r} \\ \vdots \\ W_L^{(k)} \in \mathbb{R}_{\geq 0}^{m_L \times r} \end{bmatrix}$; dictionary matrix $H^{(k)} \in \mathbb{R}_{\geq 0}^{r \times n}$

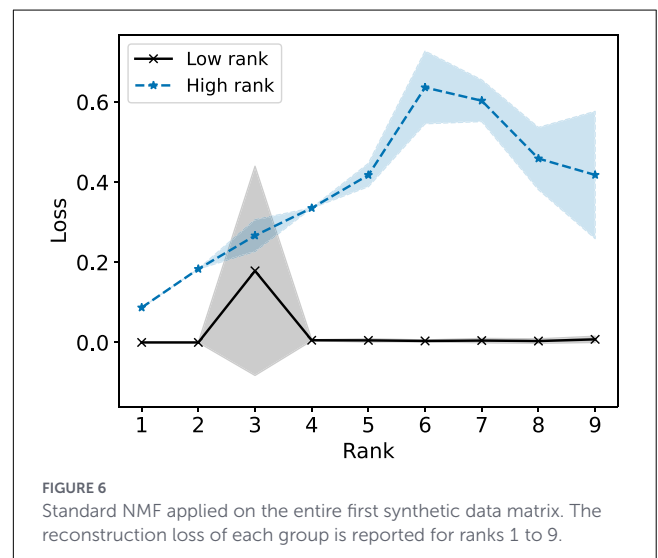
Algorithm 3. Fairer-NMF: multiplicative updates scheme.

- Group 1 (High rank) : $r_1 = 6, H_1 = (e_1 \ e_2 \ e_3 \ e_4 \ e_5 \ e_6)^T$.
- Group 2 (Low rank): $r_2 = 3, H_2 = (e_7 + e_{10} \ e_8 + e_{11} \ e_9 + e_{12})^T$.

Here each standard basis vector $e_i \in \mathbb{R}^{12}$. This construction results in the two groups being orthogonal to each other.

As discussed in Section 5.1, Figure 2 shows that standard NMF exhibits a discrepancy in the R-Error among groups with the same size but that differ in complexity. In Figure 6, for all ranks, we observe a much higher loss for the high rank group ($r = 6$) compared to the low rank group ($r = 3$). This is of course not surprising, given that NMF minimizes total error, which is more efficiently done by minimizing the error of the low rank group. The loss can be interpreted as the difference between the reconstruction error the group would incur by being part of the population and the error the group would have incurred if the model was run on that group alone. Thus, groups with the higher loss are “sacrificing” more by being part of the population.

In Figure 7 for Fairer-NMF (MU and AM), we see how the high-rank group ($r = 6$) still incur larger R-Error than the low-rank group ($r = 3$) for ranks 1 through 5. Starting from rank 6, however, the Fairer-NMF reconstructions have similar R-Error for both groups. This example, designed to be extreme, also highlights that the fairer formulation has the ability to *increase* error for some individuals or groups. In fact, this is not surprising, since especially



when groups are quite different, the factorization needs more rank to explain all groups, and when that rank is fixed, some groups will have to experience increased error for others to experience decreased error. The hope of course, is that the former is small for each individual while the latter may improve things significantly for

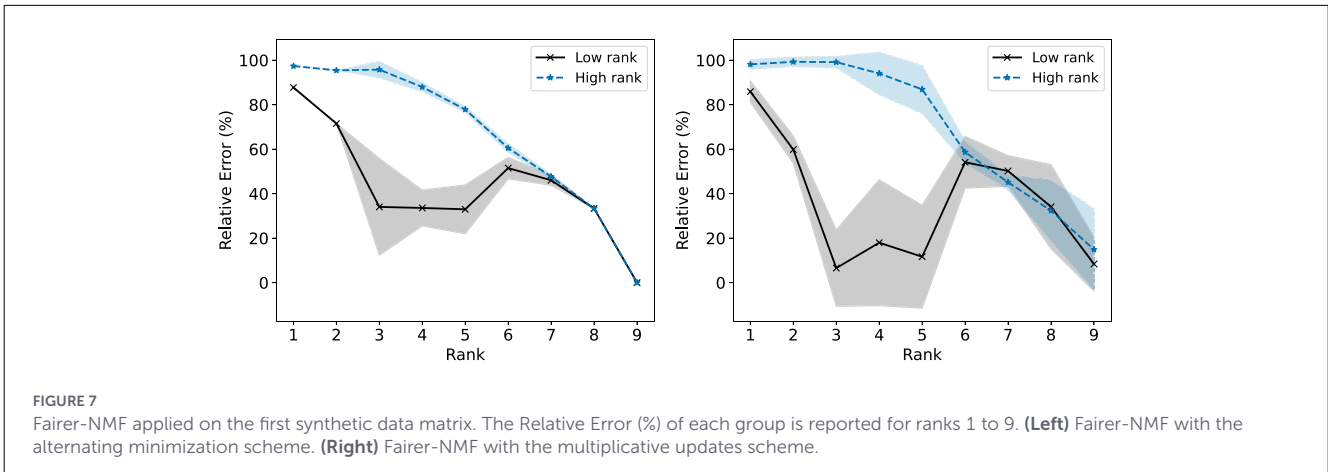


FIGURE 7 Fairer-NMF applied on the first synthetic data matrix. The Relative Error (%) of each group is reported for ranks 1 to 9. **(Left)** Fairer-NMF with the alternating minimization scheme. **(Right)** Fairer-NMF with the multiplicative updates scheme.

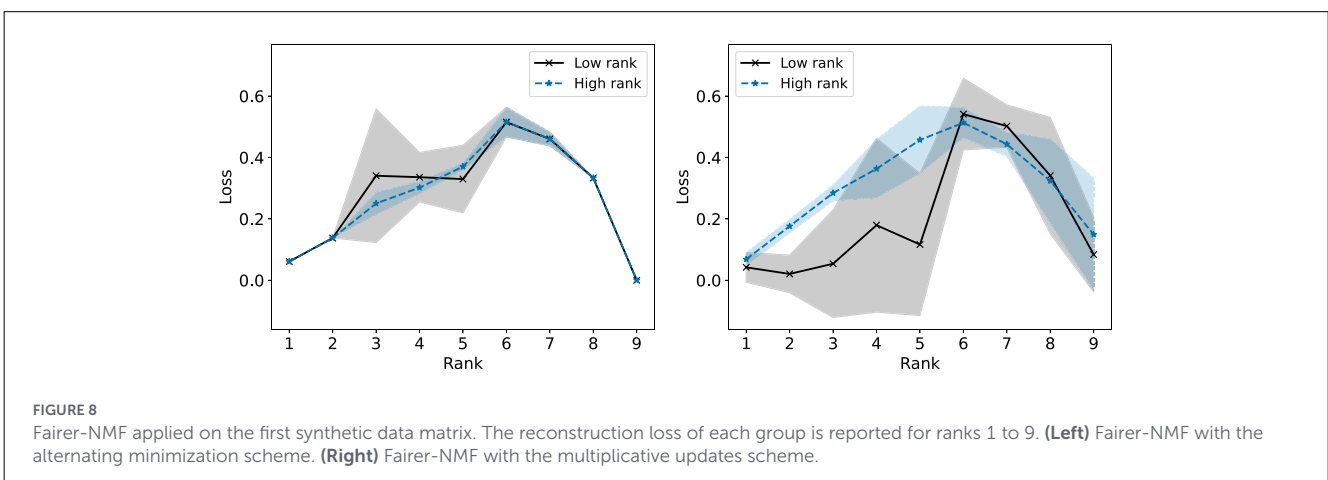


FIGURE 8 Fairer-NMF applied on the first synthetic data matrix. The reconstruction loss of each group is reported for ranks 1 to 9. **(Left)** Fairer-NMF with the alternating minimization scheme. **(Right)** Fairer-NMF with the multiplicative updates scheme.

others. If the groups are incredibly different with little in common, one may of course opt to simply treat them as entirely separate populations. All this being said, as mentioned in the introduction and discussed more in Section 8, what is fair is highly application dependent and algorithm selection should always be done with care. We also observe in Figure 8 that all groups have a comparable R-Loss, although the MU scheme is less effective at finding a minimum that equalizes the loss between the two groups.

7.2.2 Synthetic dataset 2: overlapping subspace structure

For the second synthetic dataset, we take $r_\ell = 3$ for all groups. We take H_ℓ to differ for the three groups:

- Groups 1 and 2: $H_1 = H_2 = (e_1 + e_4 \ e_2 + e_5 \ e_3 + e_6)^\top$.
- Group 3: $H_3 = (e_7 + e_{10} \ e_8 + e_{11} \ e_9 + e_{12})^\top$.

For group 2 in this dataset, we then perturb X_2 by adding an error term to each coordinate as an independently sampled Gaussian random variable with mean 0 and variance 1/100. We then truncate any negative component of the resulting matrix to

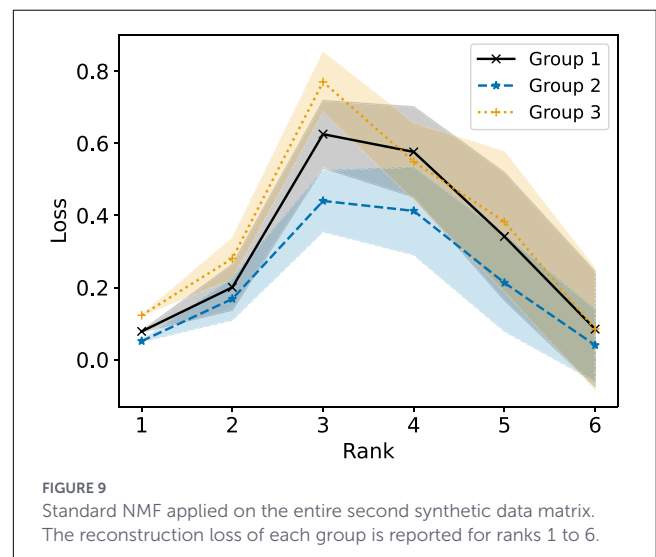


FIGURE 9 Standard NMF applied on the entire second synthetic data matrix. The reconstruction loss of each group is reported for ranks 1 to 6.

0. As a result, group 3 is orthogonal to groups 1 and 2, while groups 1 and 2 have a similar low rank structure.

In Section 5.1, we saw from Figure 3 that standard NMF with rank at most 3 on the full data matrix results in group 3 having higher R-Error than the other two groups in the second synthetic

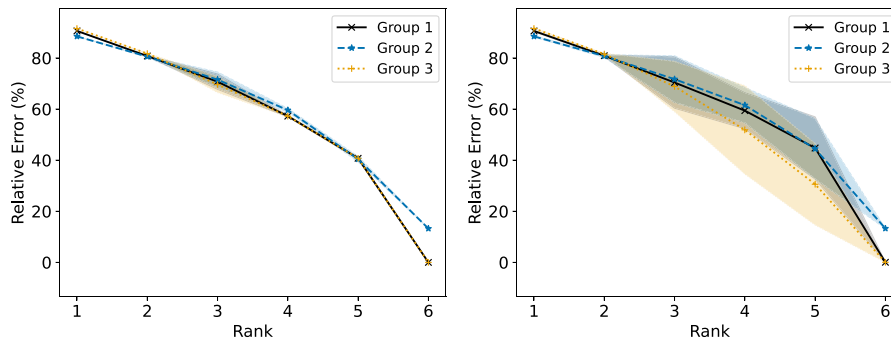


FIGURE 10 Fairer-NMF applied on the second synthetic data matrix. The Relative Error (%) of each group is reported for ranks 1 to 6. **(Left)** Fairer-NMF with the alternating minimization scheme. **(Right)** Fairer-NMF with the multiplicative updates scheme.

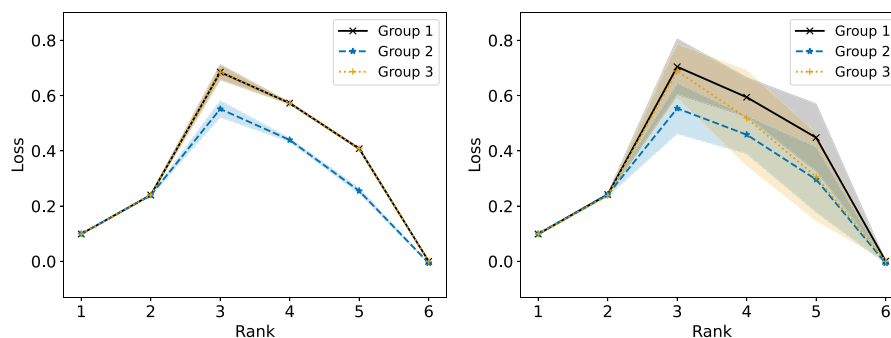


FIGURE 11 Fairer-NMF applied on the second synthetic data matrix. The reconstruction loss of each group is reported for ranks 1 to 6. **(Left)** Fairer-NMF with the alternating minimization scheme. **(Right)** Fairer-NMF with the multiplicative updates scheme.

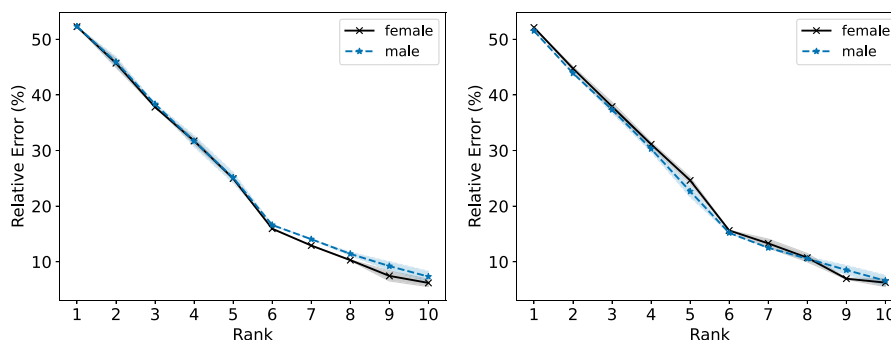


FIGURE 12 The Relative Error (%) of each group in the Heart Disease dataset is reported for ranks 1 to 10. **(Left)** Standard NMF applied on the entire Heart Disease data matrix. **(Right)** Standard NMF applied on the Heart Disease data matrix of each group individually.

dataset. In [Figure 9](#), we see that this is reflected in the loss function, where group 3 has a higher loss for these ranks. We show the results of Fairer-NMF on this dataset in [Figures 10, 11](#). We can see that Fairer-NMF is able to find a solution where the R-Error for all groups is nearly equal for all ranks in the dataset and where the R-Loss for groups 1 and 2 are equal. However, the MU scheme again exhibits much higher variance than the AM scheme for this problem. Furthermore, for the higher rank reconstructions the MU

scheme is not able to consistently achieve parity of the R-Error and R-Loss between groups 1 and 3.

7.3 Heart disease dataset

The heart disease dataset ([Janosi et al., 1989](#)) is a dataset designed for medical research to predict whether a patient has heart

disease or not based on various medical attributes. The dataset is commonly used in machine learning research to evaluate a model's performance in classifying the presence and absence of the disease. The most complete and commonly used subset of the dataset in machine learning research is the Cleveland database. The database consists of 303 samples and 13 attributes that are clinical parameters obtained from the patients such as sex, age, resting blood pressure, and serum cholesterol.

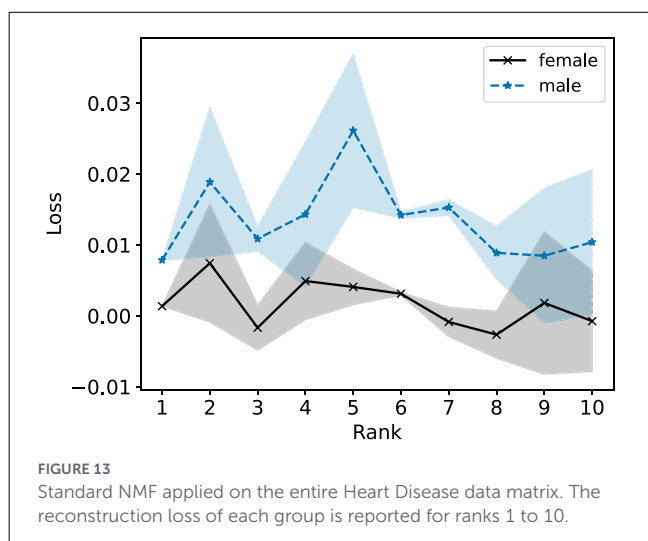
We seek to investigate the performance of standard NMF in representing the population stratified by patient sex (reported only as male or female in this dataset). We conduct this analysis in an unsupervised setting, excluding the binary target variable that indicates disease presence or absence. We also omit the sex attribute to perform our analysis on the two populations. The numerical features in the dataset are non-negative and the categorical features are recorded as integers. There are 201 individuals in the female group and 96 in the male group.

In the right plot of Figure 12, we observe that the male population generally incurs a slightly lower R-Error than the female population for the NMF models with ranks 1–5. In the left plot of Figure 12, NMF achieves similar R-Error for both groups for ranks

1–5 and lower R-Error for the female group compared to the male group. In Figure 13, we observe overall higher R-Loss for the male group compared to the female group which indicates that NMF inadvertently favored the female group.

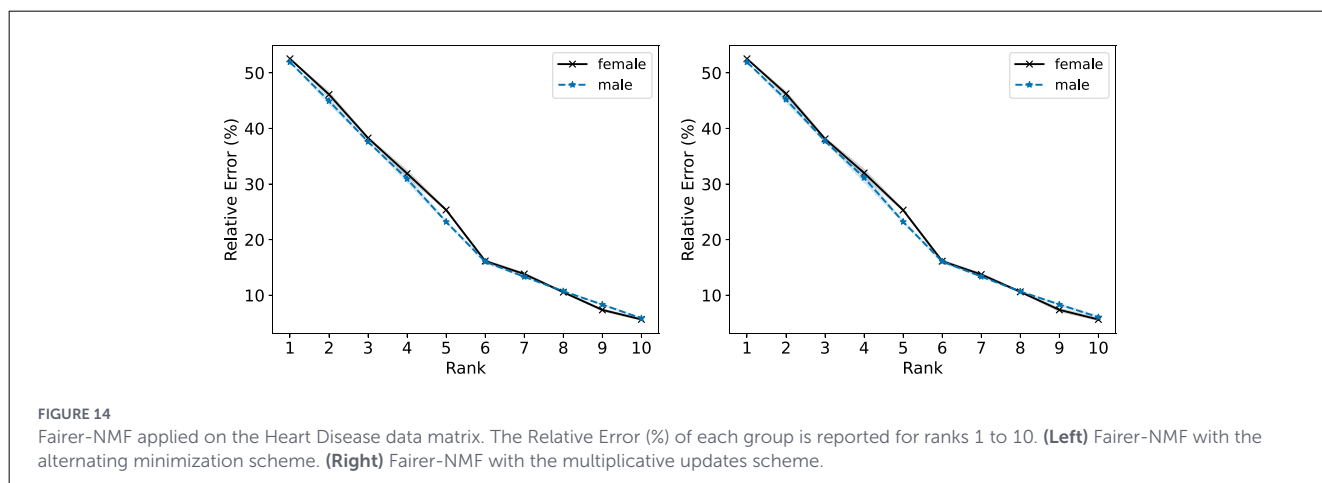
In Figure 14 (left and right), we observe the Fairer-NMF R-Error values are similar to the R-Error values when NMF is applied to each group individually (right plot of Figure 12). As observed in Figure 15, overall Fairer-NMF (MU and AM) achieves a similar loss for both populations which in this application may or may not be “fair”. With the fairness criterion considered in Fairer-NMF, some patients will incur a higher reconstruction error compared to that achieved with a standard NMF model (e.g., for ranks 1–5).

We highlight that a possibility discussed in Section 6.1 occurs for this dataset: it is possible for Fairer-NMF to produce reconstructions with negative loss. We see in Figures 13, 15 that this does happen at some of the ranks for at least one of the groups. Indeed, in Figure 15 we see that the loss for a rank-10 Fairer-NMF is consistently below 0. This shows that sometimes in practice Fairer-NMF is able to find a better reconstruction for each group when applied to the full dataset than standard NMF can even when applied to just a single group. In this case, the “cost” of representing both groups together as opposed to individually is negligible when using Fairer-NMF.



7.4 20Newsgroups dataset

The 20newsgroups dataset (Rennie, 1995) is a popular benchmark dataset containing documents gathered from 20 newsgroups that are partitioned into 6 major subjects. We sample 1,500 documents from the entire dataset with the number of samples from each subject is proportional to the size of the subject in the entire dataset. We cast all letters to lowercase and remove special characters as part of the pre-processing of the data. The TFIDF vectorizer with the English stop words list from the NLTK package is applied to transform the text data into a matrix. After obtaining the data matrix of the entire dataset, the matrix is partitioned into 6 groups according to the 6 subjects present in the original dataset. The sizes of the groups are: Computer 389, Sale 78, Recreation 316, Politics 209, Religion 193, Scientific 315.



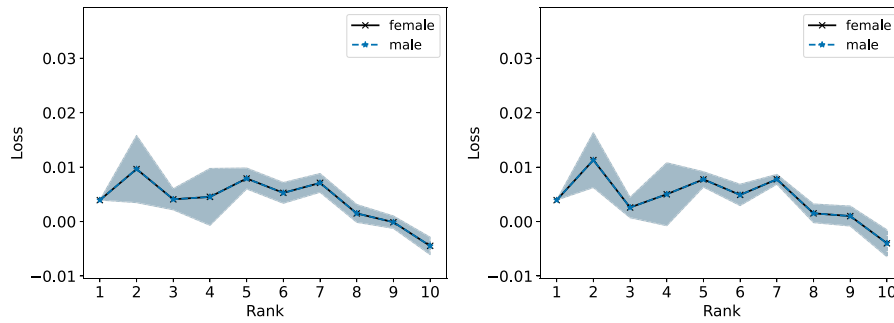


FIGURE 15 Fairer-NMF applied on the Heart Disease data matrix. The reconstruction loss of each group in the dataset is reported for ranks 1 to 10. **(Left)** Fairer-NMF with the alternating minimization scheme. **(Right)** Fairer-NMF with the multiplicative updates scheme.

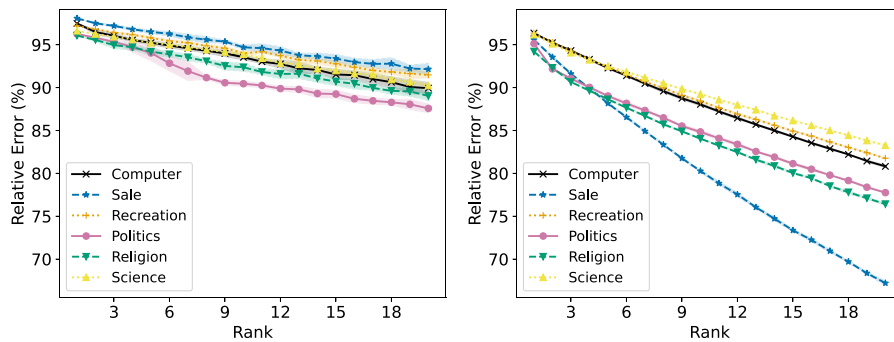


FIGURE 16 The Relative Error (%) of each group in the 20 Newsgroups dataset is reported for ranks 1 to 20. **(Left)** Standard NMF applied on the entire 20 Newsgroups data matrix. **(Right)** Standard NMF applied on the 20 Newsgroups data matrix of each group individually.

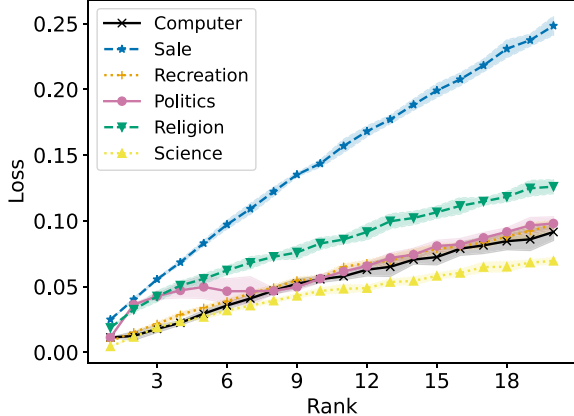


FIGURE 17 Standard NMF applied on the entire 20 Newsgroups data matrix. The reconstruction loss of each group is reported for ranks 1 to 20.

range of R-Error is more tightly clustered around 90%. Notably, the “Sale” group has the lowest R-Error when an NMF model is trained on each group individually, but the highest when one is trained on all groups at once. This results in it having the highest R-Loss, as shown in Figure 17.

Fairer-NMF is able to resolve this discrepancy. Figure 18 (left and right) show that the six different groups have reconstruction errors that correspond to the individual group reconstruction errors. The “Sale” group is the group with the lowest R-Error, and the rest of the groups appear in the order they do in the right plot of Figure 16. Accordingly, the loss of each group under Fairer-NMF (MU and AM), as shown in Figure 19, are all very similar.

Figures 18, 19 show another counterintuitive phenomenon. While R-Error for each group decreases as the number of ranks in the decomposition increases, R-Loss increases as the rank goes up. This is because the reconstruction error of each group decreases in rank much faster when decomposed individually than when decomposed together.

Figure 16 shows the R-Error of each group with NMF, both applied to the full dataset at once and each group individually. When NMF is performed on each group individually, the R-Error at rank 20 ranges from around 70% to 85%, depending on the group. When one NMF reconstruction is obtained for the full dataset, the

7.5 Algorithm comparisons

We propose two different algorithms for Fairer-NMF: the alternating minimization method (Algorithm 2) and the multiplicative updates method (Algorithm 3). In Figures 8, 11,

15, 19, both algorithms are run on three different datasets: the first synthetic dataset, the second synthetic dataset, the heart disease dataset, and the 20Newsgroups dataset, respectively. For the heart disease and 20Newsgroups datasets, both algorithms perform equally well. However, for the synthetic dataset, the alternating minimization method is more consistent in finding low loss solutions for the optimization problem. This discrepancy is mirrored in Section 6, where we only show a non-increasing result for the alternating minimization method.

However, an important consideration when choosing between these two methods is the computation cost of each. The alternating minimization involves solving one SOCP and one NNLS problem with each iteration, which is very expensive. On the other hand, the multiplicative updates scheme only requires a few matrix multiplications in each iteration. Figure 20 shows how long both algorithms took to reach convergence for the two real-world datasets² when run on a 12 core 3.50GHz Intel i9-9920X CPU. For the larger dataset (the 20Newsgroups dataset), the alternating minimization scheme is substantially slower than the multiplicative updates scheme. A single Fairer-NMF decomposition with the alternating minimization method can easily take over an hour to perform, whereas the longest time for convergence with the multiplicative updates method across all datasets and ranks is 129 seconds.

To conclude this comparison, we refer back to Figures 6, 9. Standard NMF applied to the second synthetic dataset sees the same variance as the MU rule for Fairer-NMF, and on the first synthetic dataset the variance can also be high. Furthermore, the MU rule is typically able to produce a lower mean loss for the highest loss group than standard NMF is typically capable of. Even though MU rule is not as consistent in finding low loss solutions as the AM algorithm, it still compares favorably to standard NMF on all datasets. As the MU algorithm is significantly faster than the AM algorithm, we expect it to be the preferred choice in most instances. The AM algorithm should be used for critical applications, for small datasets, or when computation cost is not a concern. In the case where high-quality reconstructions are critical but computation cost is a concern, we note that the only source of randomness in the MU algorithm is the initialization. The MU rule can be modified by adjusting the update rule for \mathbf{c} to decay more slowly or by starting at a different initialization for the matrices.

8 Discussion

We remark here on the title of the manuscript, and the notion that our proposed framework, and indeed any machine learning framework, is very unlikely to ever be completely “fair”. On the other hand, there is certainly a need to make ML algorithms *fairer* to help practitioners identify inequities and provide alternative methods that offer a fairer outcome for some applications. Our objective in Equation 5, for example, asks that the maximum reconstruction loss across all population groups be minimized. In many contexts, as motivated in Section 5, this results in fairer

outcomes. Indeed, in many settings, populations consist of majority groups and minority groups, and because typical models minimize average or overall error, minority groups will typically have a higher reconstruction error than the majority. Further, standard NMF does not take into account the complexity of the groups.

We also comment here on the assumption that the L population subgroups are known a priori. In many settings, this will be the case; for example, when the groups are defined by a specific population demographic within the dataset itself, this is a reasonable way of assigning subgroups. In other settings, the groups may be more complex, or such information may simply be unavailable. We briefly remark on two possible avenues to address these settings, emphasizing again that we view the contribution of this manuscript as the proposal of and derivation of a fairer NMF formulation and not the identification of relevant subgroups (which we believe itself could warrant an entire study).

One possible way to learn the subgroups is by using NMF or another clustering method itself. Indeed, NMF is by its nature designed to learn topics that separate variables and/or the population. One could employ simple thresholding of the learned topic values (i.e. the values in W), or apply a clustering method to the data or topic representations to divide the population according to topic strengths. It would of course be of interest to investigate whether such pre-processing into learned subgroups improves fairness when the groups are known a priori and can be compared. A second alternative would be to apply our proposed fairer NMF method iteratively; that is, run standard NMF and divide the population into groups according to reconstruction error values (a group then consists of data points sharing similar error values). Then those groups can be used, and this process could even be iteratively refined. Of course, when there are no known groups, the notion of fairness itself changes and is highly application dependent. This is indeed why there are many notions of fairness to begin with, including individual fairness, which would align with extremely fine grained group structure.

The objective we propose takes into account the size and complexity of the groups and achieves fairer outcomes in these settings. However, some drawbacks need to be considered, as will be the case for any method that attempts to mitigate fairness. First, we assume the population groups are known a priori. This can likely be overcome by learning the groups on the fly through cross-validation of reconstruction errors and is a future direction of research. The next concern of course is that it may not always be desirable to minimize the maximum reconstruction loss. Indeed, through this fairness mitigation, some groups and therefore individuals may receive a higher reconstruction error than they would have without the “fairer” approach. In settings like medical applications, where these tools are used to predict, for example, the likelihood of a patient having a disease, this may no longer seem fairer. It is thus clear that the notion of fairness itself is highly application-dependent, and great care should be taken when mitigating—or not—in learning methods. A valuable direction for future work is to propose alternative NMF formulations under different fairness criteria and study their effectiveness across various settings and applications. Additionally, comparative evaluation of different fair GLRM frameworks (e.g., Buet-Golfouse and Utyagulov, 2022) would be valuable, particularly in understanding the trade-offs between different fairness criteria and optimization approaches.

² We do not report times for the synthetic datasets, due to the different choice in stopping condition.

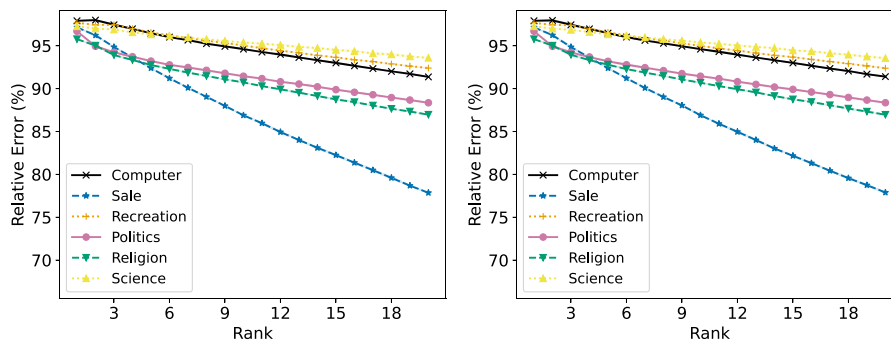


FIGURE 18 Fairer-NMF applied on the 20 Newsgroups data matrix. The reconstruction loss of each group in the dataset is reported for ranks 1 to 20. **(Left)** Fairer-NMF with the alternating minimization scheme. **(Right)** Fairer-NMF with the multiplicative updates scheme.

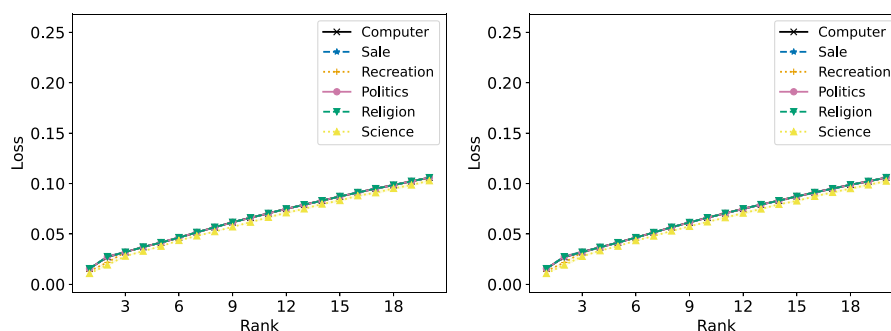


FIGURE 19 Fairer-NMF applied on the 20 Newsgroups data matrix. The reconstruction loss of each group in the dataset is reported for ranks 1 to 20. **(Left)** Fairer-NMF with the alternating minimization scheme. **(Right)** Fairer-NMF with the multiplicative updates scheme.

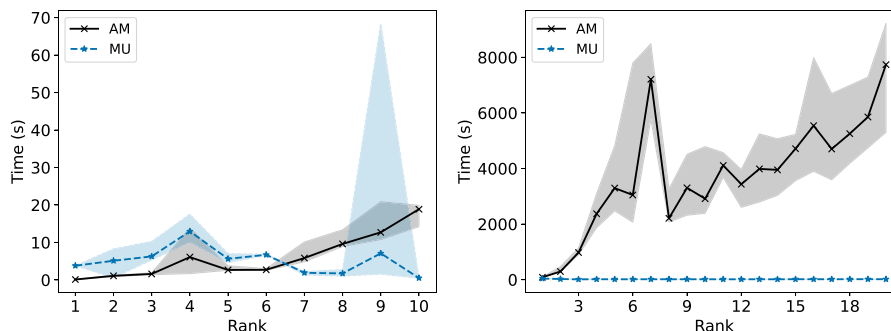


FIGURE 20 Time in seconds required for Fairer-NMF with alternating minimization (AM) scheme and with multiplicative updates (MU) scheme to converge for the Heart Diseases dataset **(left)** and 20Newsgroups dataset **(right)**. The convergence criterion is given by Equation 10. We report the median and interquartile range over 10 trials.

9 Conclusion

NMF is a widely used topic modeling technique in various domains, particularly when interpretability and trust are essential. We believe that examining the fairness of NMF is a valuable contribution to the field and an important step toward tackling key issues related to bias and fairness. In this work, we presented an alternative NMF objective that seeks a non-negative

low-rank model that provides equitable reconstruction loss across different groups. The goal is to learn a common NMF model for all groups under the min-max fairness framework which seeks to minimize the maximum of the average reconstruction loss across groups. We proposed an alternating minimization algorithm and a multiplicative updates algorithm. Numerically, the latter demonstrated reduced computational time compared to a CVXPY (Diamond and Boyd, 2016) implementation of

the AM algorithm while still achieving similar performance. We showcased on synthetic and real datasets how standard NMF could lead to biased outcomes and discussed the overall performance of Fairer-NMF.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://archive.ics.uci.edu/dataset/45/heart+disease>; https://scikit-learn.org/stable/datasets/real_world.html.

Author contributions

LK: Validation, Supervision, Writing – review & editing, Investigation, Writing – original draft. EG: Writing – original draft, Investigation, Writing – review & editing, Methodology, Validation. DN: Conceptualization, Funding acquisition, Writing – original draft, Supervision, Writing – review & editing. HG: Validation, Investigation, Writing – review & editing. NJ: Investigation, Validation, Writing – review & editing. AL: Validation, Writing – review & editing, Investigation.

Funding

The author(s) declared that financial support was received for this work and/or its publication. EG is partially supported by the UCLA Racial Justice seed grant, UCLA Dissertation Year Award, and the NSF Graduate Research Fellowship under grant DGE 2034835. LK and DN are partially supported by the Dunn Family Endowed Chair fund. All authors were partially supported by NSF DMS 2408912. This material is based upon work supported by

the National Science Foundation under Grant No. DMS-1928930 and by the Alfred P. Sloan Foundation under grant G-2021-16778, while the authors EG and DN were in residence at the Simons Laufer Mathematical Sciences Institute (formerly MSRI) in Berkeley, California, during the Fall 2023 semester.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Backurs, A., Indyk, P., Onak, K., Schieber, B., Vakilian, A., and Wagner, T. (2019). "Scalable fair clustering," in *International Conference on Machine Learning* (New York: PMLR), 405–413.
- Barocas, S., Hardt, M., and Narayanan, A. (2023). *Fairness and Machine Learning: Limitations and Opportunities*. Cambridge, MA: MIT Press.
- Berry, M. W., and Browne, M. (2005). Email surveillance using non-negative matrix factorization. *Comput. Mathem. Organiz. Theory* 11, 249–264. doi: 10.1007/s10588-005-5380-5
- Bertsekas, D. P. (1997). Nonlinear programming. *J. Operat. Res. Soc.* 48, 334–334. doi: 10.1057/palgrave.jors.2600425
- Buet-Golfouse, F., and Utyagulov, I. (2022). "Towards fair unsupervised learning," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1399–1409.
- CBS News (2023). *ChatGPT and Large Language Model Bias*. Available online at: <https://www.cbsnews.com/news/chatgpt-large-language-model-bias-60-minutes-2023-03-05/> (Accessed October 30, 2024).
- Chen, X., Fain, B., Lyu, L., and Munagala, K. (2019). "Proportionally fair clustering," in *International Conference on Machine Learning* (New York: PMLR), 1032–1041.
- Chhabra, A., Masalkovaitė, K., and Mohapatra, P. (2021). An overview of fairness in clustering. *IEEE Access* 9, 130698–130720. doi: 10.1109/ACCESS.2021.3114099
- Chierichetti, F., Kumar, R., Lattanzi, S., and Vassilvitskii, S. (2017). "Fair clustering through fairlets," in *Advances in Neural Information Processing Systems*, vol. 30, eds. I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al. (Curran Associates, Inc.). Available online at: https://proceedings.neurips.cc/paper_file/s/paper/2017/file/978fce5bcc4eccc88ad48ce3914124a2-Paper.pdf (Accessed November 10, 2024).
- Cichocki, A., Zdunek, R., Phan, A. H., and Amari, S. (2009). *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. Hoboken, NJ: John Wiley & Sons.
- Diamond, S., and Boyd, S. (2016). CVXPY: A python-embedded modeling language for convex optimization. *J. Mach. Learn. Res.* 17, 1–5.
- Domahidi, A., Chu, E., and Boyd, S. (2013). "ECOS: An SOCP solver for embedded systems," in *2013 European Control Conference (ECC)*, 3071–3076. doi: 10.23919/ECC.2013.6669541
- Ghadiri, M., Samadi, S., and Vempala, S. (2021). "Socially fair k -means clustering," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 438–448.
- Gillis, N., Leplat, V., Tan, V. Y., and Tan, V. (2021). Distributionally robust and multi-objective nonnegative matrix factorization. *IEEE Trans. Pattern Analysis Mach. Intell.* 44, 4052–4064. doi: 10.1109/TPAMI.2021.3058693
- Hamamoto, R., Takasawa, K., Machino, H., Kobayashi, K., Takahashi, S., Bolatkan, A., et al. (2022). Application of non-negative matrix factorization in oncology:

- one approach for establishing precision medicine. *Brief. Bioinform.* 23:bbac246. doi: 10.1093/bib/bbac246
- Hassaine, A., Canoy, D., Solares, J. R. A., Zhu, Y., Rao, S., Li, Y., et al. (2020). Learning multimorbidity patterns from electronic health records using non-negative matrix factorisation. *J. Biomed. Inform.* 112:103606. doi: 10.1016/j.jbi.2020.103606
- Janosi, A., Steinbrunn, W., Pfisterer, M., and Detrano, R. (1989). *Heart Disease*. UCI Machine Learning Repository. University of California, Irvine.
- Johnson, L., Kassab, L., Liu, J., Needell, D., and Shapiro, M. (2024). *Towards Understanding Neurological Manifestations of Lyme Disease Through a Machine Learning Approach with Patient Registry Data*. Baltimore, MD: PharmaSUG.
- Joshi, S., Gunasekar, S., Sontag, D., and Joydeep, G. (2016). "Identifiable phenotyping using constrained non-negative matrix factorization," in *Machine Learning for Healthcare Conference* (New York: PMLR), 17–41.
- Kim, D., Sra, S., and Dhillon, I. S. (2008). Fast projection-based methods for the least squares nonnegative matrix approximation problem. *Stat. Anal. Data Min.* 1, 38–51. doi: 10.1002/sam.104
- Kim, H., and Park, H. (2008). Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM J. Matrix Anal. A.*, 30, 713–730. doi: 10.1137/07069239X
- Lang, K. (1995). "Newsweeder: Learning to filter netnews," in *Proceedings of the 12th International Conference on Machine Learning*, eds. A. Prieditis, and S. Russell, 331–339.
- Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791. doi: 10.1038/44565
- Lee, D. D., and Seung, H. S. (2001). "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems, Vol. 13*, eds. T. Leen, T. Dietterich, and V. Tresp (MIT Press), 556–562. Available online at: https://proceedings.neurips.cc/paper_files/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf (Accessed November 10, 2024).
- Li, Y., Chen, H., Fu, Z., Ge, Y., and Zhang, Y. (2021). "User-oriented fairness in recommendation," in *Proceedings of the Web Conference 2021*, 624–632.
- Lin, C.-J. (2007). Projected gradient methods for nonnegative matrix factorization. *Neural Comput.* 19, 2756–2779. doi: 10.1162/neco.2007.19.10.2756
- Liu, S., Ge, Y., Xu, S., Zhang, Y., and Marian, A. (2022). "Fairness-aware federated matrix factorization," in *Proceedings of the 16th ACM Conference on Recommender Systems* (New York, NY: Association for Computing Machinery), 168–178.
- Mahabadi, S., and Vakilian, A. (2020). "Individual fairness for k -clustering," in *International Conference on Machine Learning* (New York: PMLR), 6586–6596.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Comp. Surv. (CSUR)* 54, 1–35. doi: 10.1145/3457607
- O'Donoghue, B., Chu, E., Parikh, N., and Boyd, S. (2016). Conic optimization via operator splitting and homogeneous self-dual embedding. *J. Optimizat. Theory Appl.* 169, 1042–1068. doi: 10.1007/s10957-016-0892-3
- Ongweso, E. (2019). *Racial Bias in AI Isn't Getting Better and Neither are Researchers Excuses*. Available online at: <https://www.vice.com/en/article/racial-bias-in-ai-isnt-getting-better-and-neither-are-researchers-excuses/> (Accessed October 30, 2024).
- Paatero, P., and Tapper, U. (1994). Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5, 111–126. doi: 10.1002/env.3170050203
- Rajapakse, M., Tan, J., and Rajapakse, J. (2004). "Color channel encoding with NMF for face recognition," in *2004 International Conference on Image Processing, 2004. ICIP'04* (Singapore: IEEE), 2007–2010.
- Samadi, S., Tantipongpipat, U., Morgenstern, J. H., Singh, M., and Vempala, S. (2018). "The price of fair PCA: One extra dimension," in *Advances in Neural Information Processing Systems, vol. 31*, eds. S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc.). Available online at: https://proceedings.neurips.cc/paper_files/paper/2018/file/cc4af25fa9d2d5c953496579b75f6f6c-Paper.pdf (Accessed November 10, 2024).
- Saxena, N. A. (2019). "Perceptions of fairness," in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (New York, NY: Association for Computing Machinery), 537–538. doi: 10.1145/3306618.3314314
- Stellato, B., Banjac, G., Goulart, P., Bemporad, A., and Boyd, S. (2020). OSQP: an operator splitting solver for quadratic programs. *Mathem. Prog. Comput.* 12, 637–672. doi: 10.1007/s12532-020-00179-2
- Tang, Z., Zhang, X., and Zhang, S. (2013). Robust perceptual image hashing based on ring partition and NMF. *IEEE Trans. Knowl. Data Eng.* 26, 711–724. doi: 10.1109/TKDE.2013.45
- Tantipongpipat, U., Samadi, S., Singh, M., Morgenstern, J. H., and Vempala, S. (2019). "Multi-criteria dimensionality reduction with applications to fairness," in *Advances in Neural Information Processing Systems, vol. 32*, eds. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc.). Available online at: https://proceedings.neurips.cc/paper_files/paper/2019/file/2201611d7a08ffd97e3e8c6b667a1bc-Paper.pdf (Accessed November 10, 2024).
- Togashi, R., and Abe, K. (2022). Fair matrix factorisation for large-scale recommender systems. *arXiv [preprint] arXiv:2209.04394*. doi: 10.48550/arXiv.2209.04394
- Truong, K. (2020). *This Image of a White Barack Obama is AI's Racial Bias Problem in a Nutshell*.
- Vavasis, S. A. (2010). On the complexity of nonnegative matrix factorization. *SIAM J. Optimizat.* 20, 1364–1377. doi: 10.1137/070709967
- Vendrow, J., Haddock, J., Needell, D., and Johnson, L. (2020). Feature selection on Lyme disease patient survey data. *Algorithms* 13:334. doi: 10.3390/a13120334
- Wang, Y., Ma, W., Zhang, M., Liu, Y., and Ma, S. (2023). A survey on the fairness of recommender systems. *ACM Trans. Inform. Syst.* 41, 1–43. doi: 10.1145/3547333
- Xu, W., Liu, X., and Gong, Y. (2003). "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval* (New York, NY: Association for Computing Machinery), 267–273. doi: 10.1145/860435.860485
- Zhu, S., Wang, Y., and Wu, Y. (2011). "Health care fraud detection using nonnegative matrix factorization," in *2011 6th International Conference on Computer Science & Education (ICCSE)* (Singapore: IEEE), 499–503.
- Zhu, Z., Wang, J., and Caverlee, J. (2020). "Measuring and mitigating item under-recommendation bias in personalized ranking systems," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY: Association for Computing Machinery), 449–458. doi: 10.1145/3397271.3401177