UNIVERSITY OF CALIFORNIA

Los Angeles

On Fairness and Interpretability in Matrix and Tensor Methods for Stratified Data

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Mathematics

by

Zerrin Vural

2026

ABSTRACT OF THE DISSERTATION

On Fairness and Interpretability in Matrix and Tensor Methods for Stratified Data

by

Zerrin Vural
Doctor of Philosophy in Mathematics
University of California, Los Angeles, 2026
Professor Deanna Needell, Chair

This dissertation develops and analyzes methods for interpretable and fair machine learning that account for stratified data. We investigate two settings in which data may exhibit meaningful differences across groups or sources. First, we propose and analyze Stratified Non-Negative Tensor Factorization, a tensor factorization method that incorporates stratum-level information while learning globally shared topics and stratum-dependent shifts. We develop an efficient multiplicative-update algorithm, incorporate total-variation regularization for image denoising, and demonstrate its effectiveness on synthetic and real-world datasets. Second, we investigate fairness in matrix completion via nuclear norm minimization. We show empirically that minority or structurally distinct groups can experience disproportionately higher reconstruction error under a globally shared low-rank model. We formalize statistical parity and equal opportunity for matrix completion and evaluate fairness-metric regularization and a group-aware formulation for addressing reconstruction disparities. Numerical experiments on synthetic and real datasets demonstrate the trade-offs between parity-based fairness and group-wise reconstruction accuracy. Together, these works demonstrate the value of accounting for local, group-specific structure alongside globally shared information,

and develops approaches for incorporating this balance into matrix and tensor methods toward improved fairness and interpretability.

The dissertation of Zerrin Vural is approved.

Andrea Bertozzi

Guido Montúfar

Hayden Schaeffer

Deanna Needell, Committee Chair

University of California, Los Angeles

2026

To my parents and my brother.

TABLE OF CONTENTS

1 Introduction **1**

1.1 Introduction to Non-negative Matrix and Tensor Factorization 3

1.1.1 Non-negative Matrix Factorization 4

1.1.2 Non-negative Tensor Factorization 8

1.1.3 Factorization for Stratified Data 10

1.2 Introduction to Matrix Completion 14

1.2.1 Low-Rank Matrix Completion 15

1.2.2 Nuclear Norm Minimization 16

1.2.3 Matrix Completion with Heterogeneous Data 17

1.2.4 Fairness in Matrix Completion 18

2 Stratified Non-negative Tensor Factorization **20**

2.1 Introduction 20

2.2 Proposed Method 22

2.2.1 Regularization 23

2.3 Experimental Results 25

2.3.1 20 Newsgroups 26

2.3.2 Olivetti Faces 28

2.3.3 Noisy, watermarked MNIST 31

2.4 Discussion 32

2.5 Conclusion 33

3 On the Fairness of Matrix Completion 34

3.1 Introduction 34

3.1.1 Contribution 36

3.1.2 Organization 36

3.2 Related Works 37

3.2.1 Fairness Metrics in Machine Learning 37

3.2.2 Fairness Notions in Recommender Systems 38

3.2.3 Fairness in Unsupervised Learning 40

3.2.4 Matrix Estimation with Group Structure 40

3.2.5 Impossibility Results in Fairness 41

3.3 Background 42

3.3.1 Notation 42

3.3.2 Matrix Completion 43

3.3.3 Fairness Metrics for Matrix Completion 46

3.4 Methods 47

3.4.1 Fairness Metric Regularization 48

3.4.2 Group-Aware Matrix Completion 49

3.5 Numerical Experiments 51

3.5.1 Experimental Setup 51

3.5.2 Synthetic Datasets 53

3.5.3 MovieLens 100K 61

3.5.4 Fairness Metric Regularization Experiments 62

3.5.5 Group-Aware Matrix Completion Experiments 71

3.5.6 Conclusion 76

4 Conclusion **79**

References **81**

LIST OF FIGURES

1.1 NMF applied to facial image data. Each image is approximated by a non-negative combination of learned dictionary elements, which can correspond to localized parts of a face. The weights of the dictionary elements determine the contribution of each part to an individual image. Adapted from Figure 1 of [LS99]. The notation in this illustration follows the column-wise convention of the original work, in which W contains basis images and H contains the corresponding feature weights. Elsewhere in this dissertation, observations are represented by rows of the data matrix, with W containing the encodings and H the basis elements. Lee DD, Seung HS. "Learning the parts of objects by non-negative matrix factorization." *Nature*, 401, 788–791 (1999), Springer Nature. Reproduced with permission from Springer Nature. . . 5

1.2 The same face approximated using vector quantization (top), NMF (middle), and principal component analysis (bottom). The learned dictionary is shown in the center column and the corresponding coefficients are shown to the right. NMF produces a parts-based representation, whereas vector quantization and PCA produce more holistic representations. Adapted from Figure 1 of [LS99]. Lee DD, Seung HS. "Learning the parts of objects by non-negative matrix factorization." *Nature*, 401, 788–791 (1999), Springer Nature. Reproduced with permission from Springer Nature. 6

1.3 The NMF factorization $A \approx WH$ for image data, with rows of A indexing images and columns indexing pixel locations. The rows of H represent learned components, while the rows of W contain the corresponding feature weights for each image. 7

1.4 The non-negative CP decomposition of a three-mode tensor as a sum of r rank-one components, each formed by the outer product of one non-negative vector from each mode. 10

1.5 The NMF factorization $A \approx WH$ for topic modeling, with rows of A indexing documents and columns indexing words. The rows of H represent learned topics, while the rows of W contain the corresponding topic weights for each document. 11

1.6 The Stratified-NMF decomposition where i indexes strata. The global topics matrix H is shared across all strata, while the coding matrices $W^{(i)}$ and stratum-level features $v^{(i)}$ are stratum dependent. 12

1.7 Stratified-NTF, where i indexes strata. The data tensors $A^{(i)}$ are approximated using stratum-level features specific to each stratum, and topics shared across strata. The stratum-level features are formed by taking the outer product with a vector of ones in the mode indexing samples, reflecting the modeling assumption that each feature is shared uniformly across samples within its stratum. 13

2.1 Loss plots for NTF, Stratified-NMF, and Stratified-NTF run for 1000 iterations on the Olivetti faces dataset. The rapid convergence shows that multiplicative updates are able to efficiently optimize the objective, matching existing methods. 29

2.2 Reconstruction comparison of NTF, Stratified-NMF, and Stratified-NTF on the first (top) and fourth (bottom) images from the first stratum. The NTF and Stratified-NTF reconstructions appear qualitatively similar, while the Stratified-NMF reconstructions fail to capture the different viewing angle between the two images due to differences in parameterization. 30

2.3 Reconstruction comparison of TV-regularized Stratified-NTF on noisy watermarked data across different regularization parameters λ. Each column shows the first image of the first stratum (top) and of the second stratum (bottom). We observe that stronger regularization decreases noise in the resulting reconstructions. 33

3.1 Illustration of synthetic data construction for the two group setting. The matrix $\mathbf{M}$ is partitioned into submatrices $\mathbf{M}_A$ and $\mathbf{M}_B$, each generated from its own factor matrices $\mathbf{U}_g$ and $\mathbf{V}_g$. The group-specific ranks are r_A and r_B, while the full matrix has rank at most $r_A + r_B$. 54

3.2 Reconstruction accuracy of vanilla nuclear norm minimization on the Shared V dataset as a function of masking rate for varying minority group size. The group-wise RMSE for groups A and B remains nearly identical across minority-group fractions, indicating negligible disparity in reconstruction accuracy. The underlying matrix has rank $r = 4$. 56

3.3 Reconstruction accuracy of vanilla nuclear norm minimization on the Shifted Means dataset as a function of masking rate for varying minority group size. The group-wise RMSE for groups A and B remains similar except at high masking rates, with little dependence on minority-group fraction. The underlying matrix has rank $r = 5$. 57

3.4 Reconstruction accuracy of vanilla nuclear norm minimization on the Perturbed V dataset as a function of masking rate for varying minority group size. The underlying matrix has rank $r = 8$, with $r_A = r_B = 4$. 58

3.5 Reconstruction accuracy of vanilla nuclear norm minimization on the Extended Minority V dataset as a function of masking rate for varying minority group size. The underlying matrix has rank $r = 6$, with $r_A = r_B = 4$ and $k = 2$. 59

3.6 Reconstruction accuracy of vanilla nuclear norm minimization on the Approximately Orthogonal dataset as a function of masking rate for varying minority group size. The underlying matrix has rank $r = 8$, with $r_A = r_B = 4$. 60

3.7 Reconstruction accuracy of vanilla nuclear norm minimization on the Orthogonal Groups dataset as a function of masking rate for varying minority group size. The underlying matrix has rank $r = 8$, with $r_A = r_B = 4$. 61

3.8 Reconstruction accuracy of vanilla nuclear norm minimization on the MovieLens 100K dataset as a function of masking rate for varying female-user fractions. Group A denotes female users and group B denotes male users. 62

3.9 Effect of the fairness-penalty weight on synthetic datasets for statistical parity (top) and equal opportunity regularization (bottom). The plots show group-wise RMSE as a function of the penalty weight at a fixed masking rate of 0.7. 64

3.10 Group-wise RMSE as a function of masking rate. Methods are presented from left to right as vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. The minority-group fraction is fixed at 20%. For the Shifted Means dataset, $\lambda_{\text{SP}} = 10$ is used because $\lambda_{\text{SP}} = 100$ produced an unstable solution. For all other synthetic datasets, $\lambda_{\text{SP}} = \lambda_{\text{EO}} = 100$. 65

3.11 Group-wise RMSE as a function of masking rate. Methods are presented from left to right as vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. The minority-group fraction is fixed at 20%. For both the Perturbed V and Extended Minority V datasets, $\lambda_{\text{SP}} = \lambda_{\text{EO}} = 100$. 66

3.12 Group-wise RMSE as a function of masking rate. Methods are presented from left to right as vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. The minority-group fraction is fixed at 20%. For both the Approximately Orthogonal and Orthogonal Groups datasets, $\lambda_{\text{SP}} = \lambda_{\text{EO}} = 100$. 67

3.13 Group-wise RMSE as a function of the fairness-penalty weight on the MovieLens 100K dataset for statistical parity and equal opportunity regularization. The minority-group fraction is fixed at 20% and the masking rate is fixed at 0.7. The selected penalty weights are $\lambda_{\text{SP}} = 100$ and $\lambda_{\text{EO}} = 25$. 68

3.14 Group-wise RMSE as a function of masking rate. Methods are presented from left to right as vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. The minority-group fraction is fixed at 20%, with $\lambda_{\text{SP}} = 100$ and $\lambda_{\text{EO}} = 25$. Both fairness metric regularizers reduce minority-group RMSE and narrow the group-wise error gap relative to vanilla nuclear norm minimization, with a small increase in majority-group RMSE. . . 69

3.15 Absolute group-wise RMSE gap, $|\text{RMSE}_A - \text{RMSE}_B|$, as a function of masking rate on the MovieLens 100K dataset under vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. The minority-group fraction is fixed at 20%, with $\lambda_{\text{SP}} = 100$ and $\lambda_{\text{EO}} = 25$. 70

3.16 Group-wise RMSE and the absolute difference in RMSE between groups as a function of masking rate for the Group-Aware objective on the Shared V and Shifted Means datasets, compared with vanilla nuclear norm minimization. The minority-group fraction is fixed at 20%. 72

3.17 Group-wise RMSE and the absolute difference in RMSE between groups as a function of masking rate for the Group-Aware objective on the Perturbed V and Extended Minority V datasets, compared with vanilla nuclear norm minimization. The minority-group fraction is fixed at 20%. 73

3.18 Group-wise RMSE and the absolute difference in RMSE between groups as a function of masking rate for the Group-Aware objective on the Approximately Orthogonal and Orthogonal Groups datasets, compared with vanilla nuclear norm minimization. The minority-group fraction is fixed at 20%. 74

3.19 Group-wise RMSE and the absolute difference in RMSE between groups as a function of masking rate for the Group-Aware objective on the MovieLens 100K dataset, compared with vanilla nuclear norm minimization. The minority-group fraction is fixed at 20%. 75

ACKNOWLEDGMENTS

Thank you to my advisor, Deanna Needell. Learning about this area of mathematics under your guidance has been a pleasure. I especially admire your enthusiasm for finding meaningful applications of mathematics and your positivity amid the inevitable uncertainties of research. I am grateful for the supportive and inclusive nature of your mentorship, both of which have significantly impacted my growth as a researcher. You provide a role model for this field.

Thank you also to Mason Porter for giving me your time and guidance while I navigated this program, particularly early on. In the same spirit, I am thankful to Marcus Roper for his thoughtfulness and support. I am also grateful to Hayden Schaeffer for his willingness to offer advice and perspective, often through casual or unexpected conversations, throughout my time at UCLA. I also want to express my gratitude to Andrea Bertozzi and Guido Montúfar for their time, thoughtful feedback, and support.

The friendships and community I found during graduate school have meant more to me than I can adequately express here. Thank you to Olha Shevchenko for walking to, and through, graduate school with me. I am lucky to have such an incredible and encouraging friend, and I am grateful for all the memories of our adventures together. Thank you to Joshua Enwright for being there for me every step of the way, always with genuine support, and for meeting my wit with your own. Thank you to Tracy Yu for all the late nights spent studying in coffee shops and for being my applied math sister! I am grateful for how much we have been able to relate to and support each other throughout this experience. Special thanks to Raymond Chu; I was fortunate to experience his gift for teaching mathematics at a pivotal point in my PhD. Thank you to Emmy Van Rooy, Annie Lu, Jack Mauro, Mia Zender, Matt Kowalski, Griffin Pinney, James Chapman, Yotam Yaniv, Roman Krutowski, and Natasha Boreyko. You all made up a wonderful community that I felt so lucky to be a part of. The way you so consistently believed in and celebrated me as I reached each

milestone throughout my PhD truly warms my heart any time I reflect back on it.

Mentorship comes in many forms. Among them, I am grateful for the strong collaborators who contributed to this work, and from whom I've learned so much: Alex Sietsema, James Chapman, Yotam Yaniv, and Minxin Zhang. In other areas of academics, I am grateful for the time and conversations I've shared with Elisa Negrini, Colleen Robicheaux, Gabriela Kováčová, Tony Wong, Tom Gannon, and Ariana Chin. Whether through WIM organizing or teaching, I am happy to have met you all. Each of you strengthened my sense of belonging and became people I look up to. Thank you also to my students and mentees. I find myself this far along in my studies because of the many wonderful teachers who inspired me along the way, and now I find that my students are a significant source of inspiration as well. Thank you for reminding me of the joy of sharing mathematics.

Thank you to Marie Yamamoto and Rabiga Arip, for being the most lovely people and roommates, and for expanding my life beyond math in so many meaningful ways. Thank you to Zach Schlamowitz for being an endless source of positivity, understanding, (indecision), and reassurance–you have made these years feel lighter. Thanks to Shiv Akshar Yadavalli for always being one year and several leagues ahead of me academically,so that you can impart your wisdom to me at whatever stage I am going through. Thank you to Daniel Torres Valladares, who has believed in me from the beginning and inspires me with his passion and purpose for becoming a scientist. Thank you also to Raunak Sinha, Karli Tosun, Siddarth Joshi, Elettra McConnell and Rob Claassen, great friends who have kept me grounded outside of my mathematical pursuits. I cherish the laughs we've shared that have made these years so much fun.

Last, thank you to my family. To my mom and dad, I am lucky to know unconditional and unending love. The opportunity to pursue this dream has been an immense privilege, and I could not have achieved it without your support. The greatest reward I have ever received for any achievement is making you both proud, and the assumption you so casually hold that I can achieve anything inspires me to always try. Thank you for giving me a home

anywhere you are in the world, and for supporting me in pursuing all of my interests—what a gift. Thank you to my brother, Turan Vural, who makes my life immeasurably richer. I have been reaping the benefits of having an outstanding older brother, who thinks I am an even more outstanding younger sister, my whole life. I have been incredibly lucky to have an unreasonable number of my favorite people in LA with me during this chapter of my life.

VITA

2021	B.S. Mathematics, The University of Texas at Austin
2021	B.S. Physics, The University of Texas at Austin
2021	B.A. Plan II Honors Liberal Arts, The University of Texas at Austin
2021	MENTOR Fellowship, UCLA
2024	M.A. Mathematics, UCLA
2024	Liggett Teaching Fellowship, UCLA
2022 - 2026	Teaching Assistant, UCLA
2024 - 2026	Graduate Student Researcher, UCLA

CHAPTER 1

Introduction

Machine learning methods are increasingly prevalent in today's data-driven world, enabling the extraction of meaningful information from large and complex datasets and informing decision-making across a wide range of applications. Many machine learning methods rely on structural assumptions, such as low-rankness, sparsity, or non-negativity, to make learning computationally tractable and to obtain useful representations of data. However, the effectiveness of these assumptions can depend on how well the assumed structure reflects the underlying data. This dissertation studies two methods in machine learning applied to settings in which data contain group- or stratum-level structure. A common thread of the two works presented is the challenge of balancing information specific to individual groups within a dataset with information shared globally across the data. This work therefore explores how global models can be enhanced by considering group-specific structure.

The first part of the dissertation considers this question in the context of interpretable representation learning. Non-Negative Matrix Factorization and Non-Negative Tensor Factorization decompose high-dimensional non-negative data into low-rank, non-negative components that admit interpretable parts-based representations. These methods have been widely used to extract interpretable features from large-scale datasets, including topics from text corpora and visual features from image data. Standard Non-Negative Matrix Factorization and Non-Negative Tensor Factorization are formulated to treat all observations as though they arise from a common underlying source. Incorporating stratum-level data structure into the factorization provides an opportunity to capture additional information present

in the data.

Stratified Non-Negative Matrix Factorization addresses source heterogeneity in matrix-valued data by jointly learning global topics shared across strata as well as stratum-dependent shifts. In Chapter 2, we extend this framework to tensor-valued data, introducing Stratified Non-Negative Tensor Factorization. By operating directly on tensors, this factorization preserves the geometric structure of higher-order data, such as the spatial axes of an image, that would otherwise be lost through flattening. We develop an efficient multiplicative-update algorithm and incorporate mode-specific total-variation regularization for image denoising. We demonstrate the method on text and image data, showing that Stratified Non-Negative Tensor Factorization can recover interpretable structure with fewer learnable parameters than its matrix counterpart. These results demonstrate the benefits of leveraging globally shared structure alongside stratum-level variation while preserving the natural structure of tensor-valued data.

In the second work of this dissertation, we investigate fairness in matrix completion via nuclear norm minimization, a method for recovering a low-rank matrix from a small sample of its entries. Implicit in this method is the assumption that a single low-rank model adequately recovers the entire matrix. When the data contain groups that exhibit different underlying structure, however, this assumption may result in disparities in reconstruction accuracy across groups. In particular, smaller or structurally distinct groups may experience larger reconstruction errors.

Chapter 3 investigates the resulting disparities through the lens of accuracy-based fairness. We show empirically that minority or structurally distinct groups can experience disproportionately larger reconstruction errors under a single globally optimal low-rank solution. We use synthetic data constructions that interpolate between groups with shared latent structure and groups with increasingly distinct latent structure, as well as real data from the MovieLens 100K dataset, to identify two mechanisms that contribute to these disparities: population imbalance, in which a group's smaller representation allows majority-group

structure to dominate the global solution, and group-specific structure, in which a group's structure requires more latent dimensions than the global objective allocates to it. Notably, disparities arising from group-specific structure can persist even when groups are equally represented.

These observations motivate the question of how fairness can be incorporated into matrix completion. We formalize two parity-based fairness criteria, statistical parity and equal opportunity, in the matrix completion setting and investigate two approaches for addressing disparities: fairness-metric regularization and a group-aware formulation that separately regularizes the submatrices corresponding to each group. We present numerical experiments on synthetic and real datasets to evaluate the trade-offs associated with enforcing parity-based fairness and its impact on group-wise reconstruction accuracy. Our results demonstrate that enforcing parity does not, in all cases, imply improved accuracy-based fairness, and that the effectiveness of the fairness interventions depends on the source of disparity.

Each chapter is written to be self-contained, with its own introduction, notation, and discussion of related work. The remainder of Chapter 1 provides additional mathematical and contextual background for the subsequent chapters. Chapter 2 is based on joint work with Alexander Sietsema, James Chapman, Yotam Yaniv, and Prof. Deanna Needell and is a version of [SVC24]. Chapter 3 is based on joint work with Dr. Minxin Zhang and Prof. Deanna Needell. Chapter 4 summarizes the contributions of both works and discusses directions for future research.

1.1 Introduction to Non-negative Matrix and Tensor Factorization

Non-negative Matrix Factorization (NMF) has found applications across a broad range of domains, including topic modeling, image and hyperspectral data analysis, audio signal processing, and biological data analysis [Gil20, Dev08]. Across these applications, its low-rank, non-negative factorization often yields components with meaningful, parts-based represen-

tations [LS99, LS00, Gil14]. We begin with a brief introduction to NMF and describe its extension to tensor-valued data. We then introduce the stratified setting, in which observations arise from multiple sources that share common structure while exhibiting source-specific variation.

1.1.1 Non-negative Matrix Factorization

Consider a collection of non-negative observations stored in a matrix $A \in \mathbb{R}^{m \times n}$, where the rows correspond to samples and the columns correspond to variables. NMF seeks a low-rank approximation of A subject to nonnegativity constraints. Because the learned factors are non-negative, each observation is represented as an additive combination of the learned components, with the contribution of each component determined by a non-negative coefficient. This structure can lead to interpretable representations in which individual components correspond to meaningful features or parts of the data.

A particularly illustrative example of the interpretability of NMF is its application to image data. Grayscale and RGB images are naturally represented by non-negative pixel intensities, making them well suited to this method. When NMF is applied to a collection of facial images, the learned components can correspond to localized features such as eyes, noses, mouths, or mustaches. Each facial image is then represented as a non-negative combination of these learned features. Figure 1.1 illustrates NMF for facial image data.

The interpretability associated with nonnegativity can be illustrated by comparing NMF with other representations. Figure 1.2 shows the same facial image approximated using vector quantization (VQ), NMF, and principal component analysis (PCA). Each of these methods factorize data into a matrix of basis elements and a corresponding matrix of coefficients. In vector quantization, the coefficient vector for each image is subject to a unary constraint, so that each image is represented by a single dictionary element. In this example, the dictionary elements are prototypical whole faces rather than individual parts. PCA instead allows each image to be represented as a linear combination of multiple basis elements, resulting

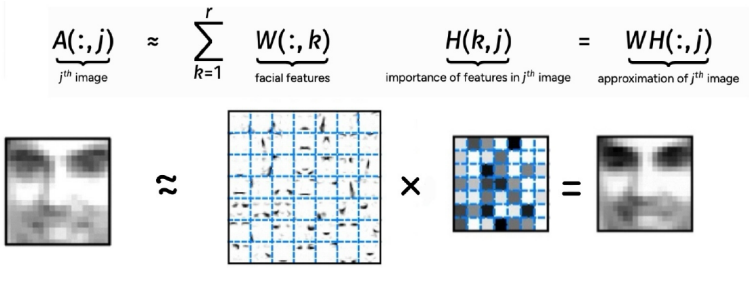


Figure 1.1: NMF applied to facial image data. Each image is approximated by a non-negative combination of learned dictionary elements, which can correspond to localized parts of a face. The weights of the dictionary elements determine the contribution of each part to an individual image. Adapted from Figure 1 of [LS99]. The notation in this illustration follows the column-wise convention of the original work, in which W contains basis images and H contains the corresponding feature weights. Elsewhere in this dissertation, observations are represented by rows of the data matrix, with W containing the encodings and H the basis elements.

Lee DD, Seung HS. "Learning the parts of objects by non-negative matrix factorization." *Nature*, 401, 788–791 (1999), Springer Nature. Reproduced with permission from Springer Nature.

in a more distributed representation. Because PCA permits entries of arbitrary signs, the basis elements can combine through both addition and cancellation, making individual basis elements less intuitive to interpret. In contrast, the nonnegativity constraints in NMF allow multiple components to contribute to a reconstruction, but only through additive combinations. This is results in a parts-based representation, in which individual components can correspond to recognizable parts of a face [LS99].

The NMF problem can be formulated as follows. Given a non-negative data matrix $A \in \mathbb{R}^{m \times n}$, NMF seeks non-negative factor matrices $W \in \mathbb{R}^{m \times r}$ and $H \in \mathbb{R}^{r \times n}$ that minimize the squared Frobenius reconstruction error:

$$\min_{W \geq 0, H \geq 0} \|A - WH\|_F^2 , \tag{1.1.1}$$

where $r \ll \min\{m, n\}$ is the prescribed factorization rank. The matrix H contains the learned basis elements, while W contains the coefficients used to combine basis elements to approximate each observation. We refer to H as the topics matrix and W as the coding matrix throughout this dissertation. Figure 1.3 illustrates this factorization in the image

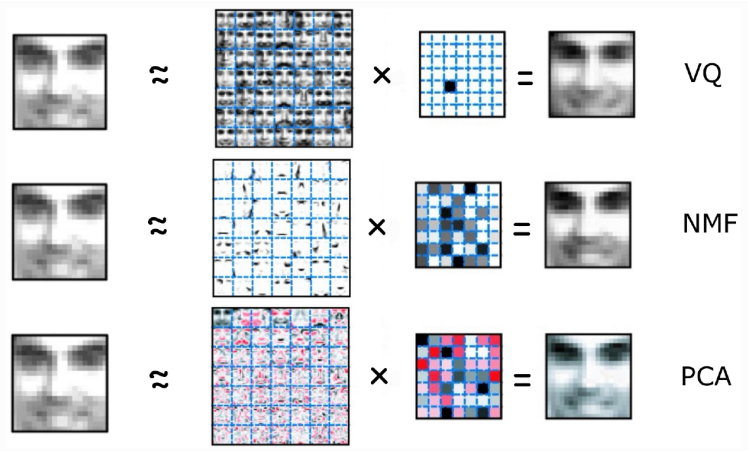


Figure 1.2: The same face approximated using vector quantization (top), NMF (middle), and principal component analysis (bottom). The learned dictionary is shown in the center column and the corresponding coefficients are shown to the right. NMF produces a parts-based representation, whereas vector quantization and PCA produce more holistic representations. Adapted from Figure 1 of [LS99].
Lee DD, Seung HS. "Learning the parts of objects by non-negative matrix factorization." *Nature*, 401, 788–791 (1999), Springer Nature. Reproduced with permission from Springer Nature.

setting. Each row of A corresponds to an image, and each column corresponds to a pixel location. Thus, a row of H represents a learned feature, while a row of W contains the weights of all learned features for a particular image.

The squared Frobenius reconstruction error is one possible choice of reconstruction loss for the NMF objective. For example [LS00] considers both the squared Euclidean loss and the generalized Kullback-Leibler divergence, which can be optimized using similar iterative approaches.

The optimization problem in (1.1.1) is not jointly convex in (W, H), although it is convex in either factor when the other is held fixed. Thus, an alternating minimization approach can be used to seek local minima by fixing one factor while updating the other [Ber97]. Denote the objective function as

$$f(W, H) = \frac{1}{2}\|A - WH\|_F^2.$$

To derive the iterative updates for H, first suppose that W is fixed. The gradient of f with

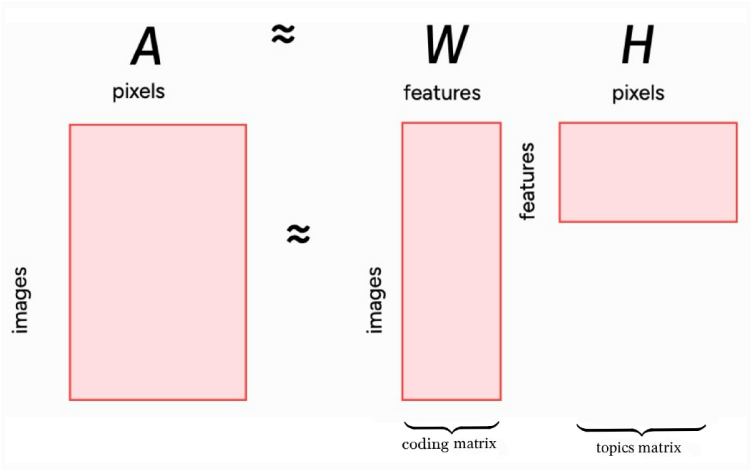


Figure 1.3: The NMF factorization $A \approx WH$ for image data, with rows of A indexing images and columns indexing pixel locations. The rows of H represent learned components, while the rows of W contain the corresponding feature weights for each image.

respect to H is then

$$\nabla_H f(W, H) = W^T WH - W^T A.$$

The standard gradient-descent update for each entry H_{ij} takes the form

$$H_{ij} \leftarrow H_{ij} - \eta_{ij} \left(\nabla_H f(W, H)\right)_{ij},$$

where η_{ij} denotes the step size. However, an additive update does not necessarily preserve the nonnegativity constraint on H. If we choose an entry-dependent step size of the form

$$\eta_{ij} = \frac{H_{ij}}{(W^T WH)_{ij}},$$

and substitute this scaling into the gradient-descent update, we arrive at the following multiplicative-update rule:

$$H_{ij} \leftarrow H_{ij} \frac{(W^T A)_{ij}}{(W^T WH)_{ij}}.$$

This multiplicative update can be viewed as separating the positive and negative components of the gradient and placing the absolute value of the negative component in the

numerator of the multiplicative factor and the positive component in the denominator. This construction can be applied when optimizing W, resulting in the multiplicative updates:

$$W_{ij} \leftarrow W_{ij} \frac{(AH^T)_{ij}}{(WHH^T)_{ij}} \tag{1.1.2}$$

$$H_{ij} \leftarrow H_{ij} \frac{(W^T A)_{ij}}{(W^T WH)_{ij}} \tag{1.1.3}$$

which are applied alternately, with each factor updated while the other is held fixed.

When initialized with non-negative factors, these updates preserve nonnegativity because each entry is multiplied by a ratio of non-negative quantities. Multiplicative updates are also attractive due to their ease of implementation and favorable monotonicity properties. Under the multiplicative update rules, the objective function f is nonincreasing, as follows from a majorization-minimization argument [LS00].

1.1.2 Non-negative Tensor Factorization

A tensor is a d-way array, where d is referred to as the order of the tensor and the different dimensions of a tensor are referred to as modes. Tensors of order $d \geq 3$ are referred to as higher-order tensors. For example, a vector is a 1-way array, while a matrix is a 2-way array, with its rows and columns corresponding to two modes. Higher-order data can be reshaped into lower-order representations, allowing existing lower-order factorization methods to be applied. For example, a tensor-valued observation can be vectorized, or flattened across a single mode, and the resulting vectors arranged as the rows or columns of a matrix. Non-Negative Tensor Factorization (NTF) instead operates directly on higher-order tensor data, extending the low-rank, parts-based representation of NMF while preserving the natural multi-way structure of the data. Retaining this structure allows relationships associated with individual modes to be incorporated directly into the factorization and lends itself to mode-specific modeling and regularization. For example, in image analysis, regularization can be imposed directly along the spatial modes, allowing the factorization to incorporate

spatial structure that would be obscured by vectorization.

For example, a collection of two-dimensional images can be represented as a matrix, with one dimension indexing the vectorized images and the other indexing the pixels, as illustrated in Figure 1.3. Alternatively, the same collection can be represented as a three-mode tensor, with one mode indexing samples and the remaining two modes representing the spatial dimensions of each image. Therefore, tensor factorizations are valuable methods to obtain a low-rank representation of higher-order data.

A common approach to tensor factorization is the CANDECOMP/PARAFAC (CP) decomposition, which represents a tensor as a sum of rank-one tensors [CC70, Har70, KB09]. A rank-one tensor is formed by taking the outer product of one vector from each mode. Thus, for an n-way tensor, a rank-one tensor has the form $\mathcal{A} = u^1 \otimes u^2 \otimes \cdots \otimes u^n$, where $u^k \in \mathbb{R}^{d_k}$. The CP decomposition approximates a tensor as a sum of r rank-one tensors as

$$\mathcal{A} \approx \sum_{j=1}^{r} \otimes_{k=1}^{n} u_j^k.$$

The rank of a tensor is defined as the smallest such r for which this representation is exact. This quantity is also referred to as the CP-rank. More generally, a CP decomposition can use a prescribed number r of rank-one components to approximate a tensor. Figure 1.4 illustrates the CP decomposition in the three-mode case.

NTF imposes elementwise nonnegativity on the components of the CP decomposition. For a non-negative tensor $\mathcal{A} \in \mathbb{R}_+^{d_1 \times \cdots \times d_n}$, NTF seeks non-negative vectors $u_j^k \in \mathbb{R}^{d_k}$ that minimize

$$\min_{\{u_j^k \geq 0\}} \left\| \mathcal{A} - \sum_{j=1}^{r} \otimes_{k=1}^{n} u_j^k \right\|_F^2. \tag{1.1.4}$$

As in NMF, the components of the factorization can be interpreted as features of the data and corresponding feature weights for each sample. The interpretation of these components depends on how the modes of the tensor are assigned to aspects of the data. In the image

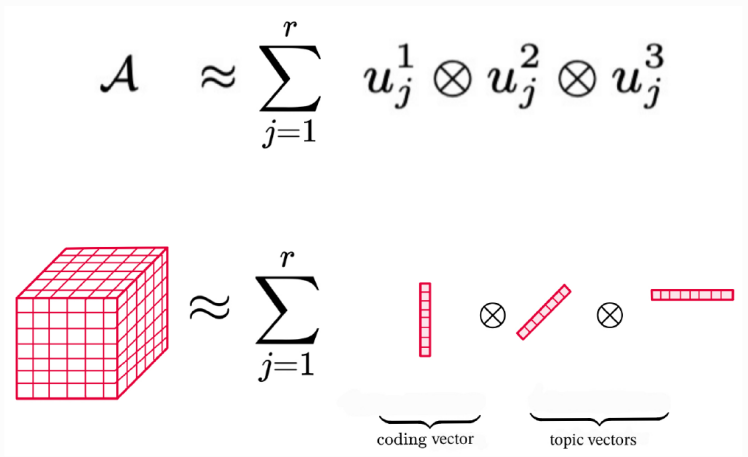


Figure 1.4: The non-negative CP decomposition of a three-mode tensor as a sum of r rank-one components, each formed by the outer product of one non-negative vector from each mode.

and text experiments considered in this dissertation, one mode indexes samples, while the remaining modes represent the variables associated with each observation. For example, if the sample mode has dimension m, then the vector in the factorization corresponding to this mode has dimension m and contains feature weights for each sample. The vectors corresponding to the remaining modes have dimensions determined by the shape of each sample and describe the learned features along the corresponding dimensions.

The non-negative CP problem is again nonconvex, but multiplicative-update rules can be derived by extending the same approach used for NMF [SH05]. As in the matrix setting, these updates preserve nonnegativity while iteratively reducing the reconstruction error.

1.1.3 Factorization for Stratified Data

Many datasets contain observations collected from multiple sources, or *strata*. Stratified factorization provides a framework for modeling such data by learning information shared across strata as well as information specific to each stratum.

To motivate this setting, consider NMF for topic modeling, which aims to identify un-

derlying topics or themes in a collection of documents. For this application, the rows of the data matrix A correspond to documents and the columns correspond to words that appear within the collection of documents. Entry A_{ij} records the number of occurrences of word j in document i. The rows of H represent topics, with each row assigning a non-negative weight to every word in the vocabulary. The words with the largest values in a given row identify the words most strongly associated with that topic and provide a human-interpretable description of its content. The rows of W give the corresponding topic weights for each document, indicating which topics are most strongly represented in each document. Figure 1.5 illustrates this application.

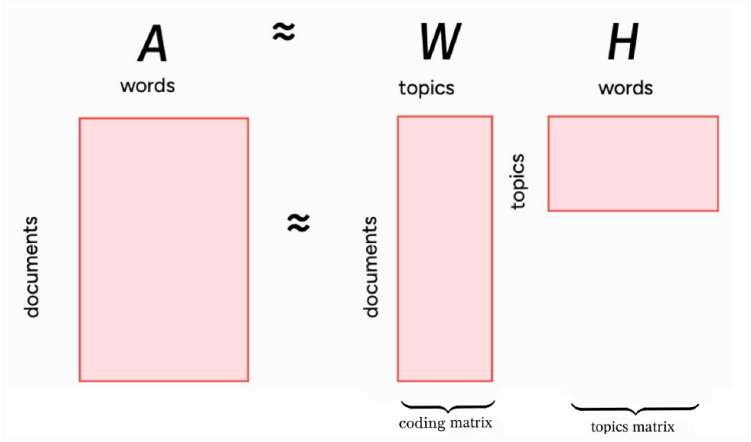


Figure 1.5: The NMF factorization $A \approx WH$ for topic modeling, with rows of A indexing documents and columns indexing words. The rows of H represent learned topics, while the rows of W contain the corresponding topic weights for each document.

Now suppose that the documents are collected from multiple sources, such as different news organizations. Documents from different sources may discuss many of the same topics, while differing in the topics they emphasize and in the language associated with those topics. NTF represents all documents using the same collection of topics and does not explicitly distinguish information shared across sources from information specific to a particular source.

The idea of considering shared and stratum-level information for NMF was introduced in [CYN23]. Suppose the data are partitioned into s strata, indexed by $i = 1, \ldots, s$, with

data matrix $A(i)$ corresponding to stratum i. The Stratified-NMF formulation proposed in [CYN23] solves

$$\min_{v(i)^T, W(i), H \geq 0} \sum_{i=1}^{s} \left\| A(i) - \mathbf{1} v(i)^T - W(i)H \right\|_F^2 . \tag{1.1.5}$$

Here, H is a global topics matrix shared across all strata, while $W(i)$ is a coding matrix learned separately for each stratum. The vector $v(i)$ provides stratum-level information through the term $\mathbf{1}v^{(i)T}$, which assumes the same additive shift to every sample within stratum i. Thus, the model retains a common set of topics while allowing each stratum to exhibit additional source-specific features. We refer to the vectors $v(i)$ as *stratum-level features* and to the rows of H as *global topics*. Figure 1.6 illustrates the Stratified-NMF decomposition for 3 strata.

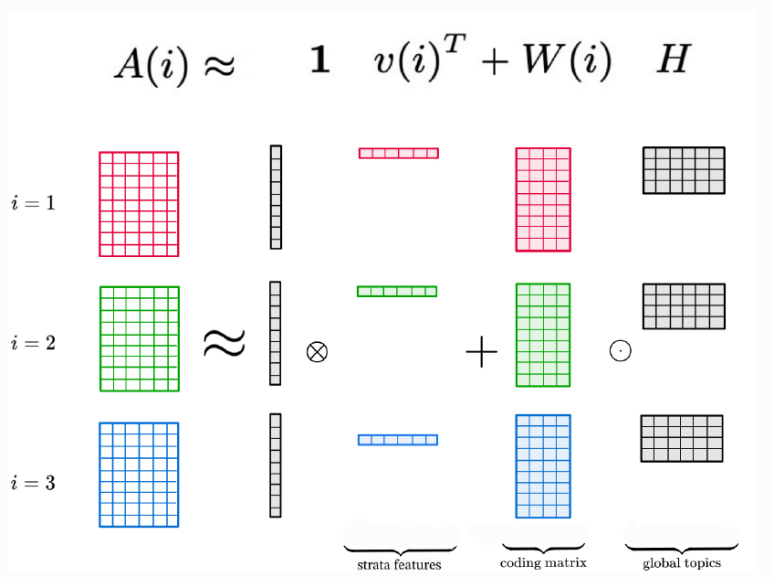


Figure 1.6: The Stratified-NMF decomposition where i indexes strata. The global topics matrix H is shared across all strata, while the coding matrices $W^{(i)}$ and stratum-level features $v^{(i)}$ are stratum dependent.

The same idea can be extended to the tensor setting introduced above. This extension is the focus of Chapter 2, where we develop Stratified-NTF. The resulting decomposition is illustrated in Figure 1.7. As in Stratified-NMF, the sample mode contains a vector of ones, so that the stratum-level features are added uniformly across samples within each stratum.

The remaining modes encode the stratum-level features and global topics through sums of rank-one tensor components.

Unlike Stratified-NMF, where a single rank-one stratum-level feature is learned for each stratum, Stratified-NTF allows each stratum to have a tunable rank $r'(i)$. This additional flexibility is useful because a single rank-one tensor component imposes a stronger structural restriction than a rank-one matrix component. Consequently, a single component may be insufficient to represent a complex stratum-level feature, and multiple rank-one components can be combined to obtain a richer, more interpretable representation.

Chapter 2 derives a multiplicative-update algorithm for this model and for a total-variation-regularized extension. The resulting method is demonstrated on text and image data, where we examine the ability of Stratified-NTF to recover interpretable global and stratum-level structure while preserving the multiway organization of the data.

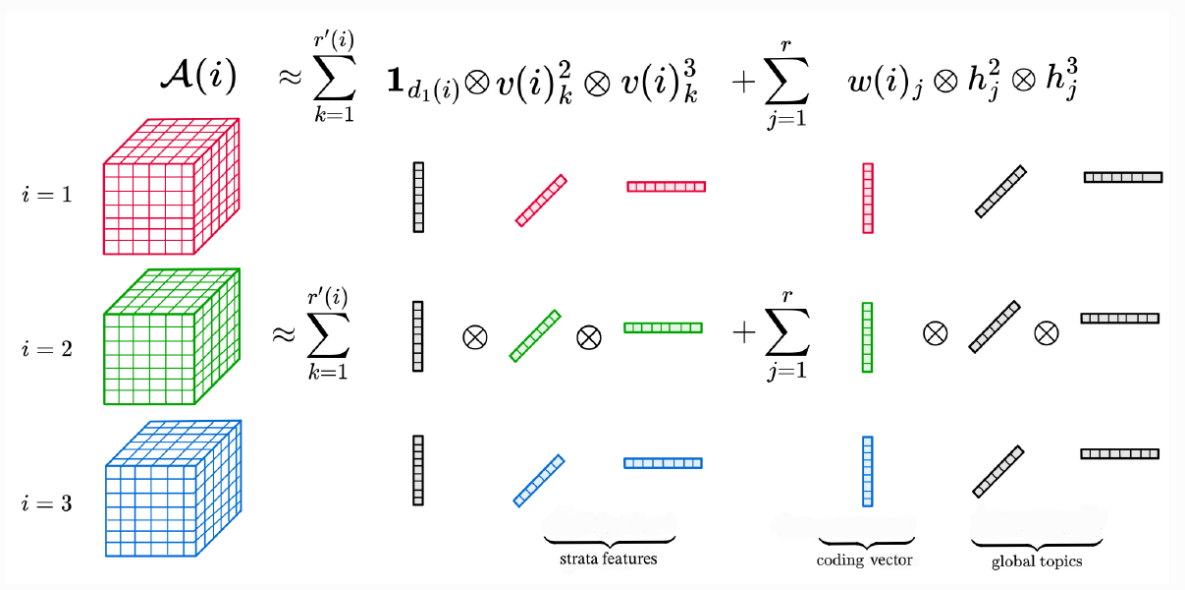


Figure 1.7: Stratified-NTF, where i indexes strata. The data tensors $A^{(i)}$ are approximated using stratum-level features specific to each stratum, and topics shared across strata. The stratum-level features are formed by taking the outer product with a vector of ones in the mode indexing samples, reflecting the modeling assumption that each feature is shared uniformly across samples within its stratum.

1.2 Introduction to Matrix Completion

We now introduce mathematical and contextual background for the second part of this dissertation: fairness in matrix completion.

Many data analysis problems involve matrices whose entries are only partially observed. Matrix completion addresses this problem by seeking to recover an unknown matrix from only a subset of its entries. Given a partially observed matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$, the goal is to construct a matrix $\mathbf{X}$ that agrees with $\mathbf{M}$ on the observed entries and accurately estimates the unobserved ones. This problem arises in a variety of applications, including collaborative filtering, recommender systems, remote sensing, control systems, and computer vision [CP10].

In general, recovering the unobserved entries of a matrix is impossible. However, if the underlying matrix can be assumed to be low rank, or approximately low rank, matrix completion becomes possible. Under suitable conditions on the matrix and the sampling pattern, a relatively small number of observed entries can suffice for accurate, and in some settings exact, recovery.

A useful application for illustrating the matrix completion problem is recommender systems. Consider, for example, a movie recommendation system in which each row of a rating matrix $\mathbf{M}$ corresponds to a user, each column corresponds to a movie, and an entry $\mathbf{M}_{ij}$ records the rating assigned by user i to movie j. Since users typically rate only a small subset of available movies, the resulting matrix contains many unobserved entries. The underlying rating matrix may be assumed to be approximately low rank, since users' preferences can often be described by a relatively small number of underlying factors, such as preferences for different genres or styles of movies. Under this assumption, the observed ratings provide information about the unobserved ratings. If the rating matrix could be completed accurately, a streaming service could use the estimated ratings to provide personalized movie recommendations to its users.

1.2.1 Low-Rank Matrix Completion

The intuition that low-rank structure allows a matrix to be recovered from relatively few observations can be understood through a degrees-of-freedom argument. Consider a matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ of fixed rank r, with singular value decomposition

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T,$$

where $\mathbf{U} \in \mathbb{R}^{m \times r}$ and $\mathbf{V} \in \mathbb{R}^{n \times r}$ have orthonormal columns, and $\mathbf{\Sigma} \in \mathbb{R}^{r \times r}$ is diagonal. The r diagonal entries of $\mathbf{\Sigma}$ contribute r degrees of freedom, while the orthonormal matrices $\mathbf{U}$ and $\mathbf{V}$ contribute $\left(mr - \frac{r(r+1)}{2}\right)$ and $\left(nr - \frac{r(r+1)}{2}\right)$ degrees of freedom, respectively, due to the orthonormality constraints. Thus, a rank-r matrix has $r(m+n-r)$ degrees of freedom. When $r \ll \min\{m, n\}$, this is substantially smaller than the mn entries required to specify an arbitrary $m \times n$ matrix. Thus, a low-rank matrix is determined by substantially fewer parameters than an arbitrary matrix of the same dimensions.

Low rank alone, however, does not guarantee that a matrix can be completed. Consider the rank-one matrix

$$\mathbf{M} = \mathbf{e}_1\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Although $\mathbf{M}$ is low rank, all of its nonzero entries are concentrated in a single row. If even all but one first row is unobserved, the missing value cannot be determined from the remaining entries: any value assigned to that entry produces a matrix of the same rank-one form that is consistent with the other observations. Thus, exact recovery of $\mathbf{M}$ requires every entry of the first row to be observed. Similarly, if an entire row or column is unobserved, matrix completion becomes impossible. In particular, the observation pattern must provide sufficient information about the entries that determine the low-rank matrix.

For this reason, theoretical analyses of matrix completion impose conditions on both the low-rank matrix and the sampling pattern. A common assumption is that the observed entries are sampled uniformly at random, so that particular rows or columns are not systematically under-observed, though recovery guarantees for deterministic and non-uniform sampling patterns have been established [FNP21]. In addition, an incoherence condition requires the singular vectors of $\mathbf{M}$ to be sufficiently spread across their coordinates rather than concentrated on a small subset. Informally, this prevents the information in a low-rank matrix from being concentrated in only a few rows or columns, as in the example above. Under suitable incoherence and sampling conditions, the observed entries can contain sufficient information for recovery [CP10].

1.2.2 Nuclear Norm Minimization

To formulate the matrix completion problem precisely, let $\Omega \subset [m] \times [n]$ denote the set of observed entries. Define the associated sampling operator $P_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ entrywise by

$$[P_\Omega(\mathbf{X})]_{ij} = \begin{cases} \mathbf{X}_{ij}, & (i,j) \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

The observed entries of $\mathbf{M}$ are represented by the projection $P_\Omega(\mathbf{M})$, where unobserved entries are set to zero. Ideally, one would seek the lowest-rank matrix consistent with these observations by solving

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \operatorname{rank}(\mathbf{X}) \qquad \text{subject to} \qquad P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M}). \tag{1.2.1}$$

However, rank minimization is nonconvex and NP-hard problem. A common approach is to replace rank minimization with the following convex relaxation

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \|\mathbf{X}\|_* \qquad \text{subject to} \qquad P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M}). \tag{1.2.2}$$

Recall that the nuclear norm of a matrix is the sum of its singular values, $\|\mathbf{X}\|_* = \sum_{k=1}^{\min\{m,n\}} \sigma_k(\mathbf{X})$. This can also be viewed as the ℓ_1 norm of the vector of singular values of a matrix. Just as ℓ_1 minimization promotes sparse vectors by penalizing the magnitudes of their entries, nuclear norm minimization promotes matrices with a small number of nonzero singular values, which corresponds to low rank.

Candès and Recht established that, for a broad class of low-rank matrices satisfying suitable incoherence assumptions and observed at sufficiently many uniformly sampled entries, nuclear norm minimization exactly recovers the underlying matrix with high probability [CR09]. Subsequent work established sharper recovery guarantees and showed that nuclear norm minimization remains effective when the observed entries are corrupted by noise [CP10]. In particular, under suitable conditions, an $n \times n$ rank-r matrix can be recovered from approximately $nr \log^2(n)$ noisy observations, with reconstruction error proportional to the noise level [CP10]. Thus, nuclear norm minimization provides both a computationally tractable formulation and a theoretically justified approach to recovering low-rank matrices from incomplete observations.

1.2.3 Matrix Completion with Heterogeneous Data

The standard formulation in (1.2.2) seeks a single low-rank representation that agrees with all of the observed entries. However, in many applications, the rows of a data matrix may belong to distinct groups or populations. These populations may differ in their representation in the data, their observation patterns, or the latent structure underlying their entries. For example, if one population is represented by substantially fewer observed entries, the reconstruction of that population may be based on less information than that of other groups in the data. Alternatively, different populations may exhibit latent structures that are not equally well represented by a single global low-rank model. Because standard nuclear norm minimization minimizes the aggregate reconstruction error across all observed entries, it does not explicitly account for how that error is distributed across groups. As a re-

sult, minimizing the overall reconstruction error can yield different reconstruction accuracies across populations.

This observation is particularly relevant in recommender systems and other applications in which matrix completion produces predictions about individuals. Systematically less accurate predictions for one population can result in differences in the quality of information, opportunities, or resources provided to different groups. Such differences may be particularly consequential in domains such as healthcare, education, and financial services. This motivates evaluating and improving the fairness of matrix completion.

1.2.4 Fairness in Matrix Completion

The question of whether a machine learning method performs equitably across populations is part of a broader literature on fairness in machine learning. The appropriate notion of fairness depends on application and context. Therefore, it is difficult to characterize a universal definition of fairness. Instead, different definitions capture different notions of equitable treatment.

One family of criteria considers the distribution of model outcomes across groups. In binary classification, for example, *statistical parity* is a *parity-based* criterion that requires the probability of a positive prediction to be the same across groups, or equivalently, that the prediction be independent of group membership [BHN23]. Other criteria evaluate model predictions conditioned on true outcomes. For example, *equal opportunity* is an *accuracy-based* criterion that requires the true-positive rate to be the same across groups [HPS16b]. The motivating goal of Chapter 3 is to improve fairness in reconstruction error across groups, which we refer to as *error-based fairness*.

For matrix completion, these notions take different forms. Our adaptation of *statistical parity* requires the reconstructed item-wise means to be similar across groups. Our adaptation of *equal opportunity* instead requires the error in the estimated item-wise group means

to be similar across groups. We evaluate *error-based fairness* by comparing the group-wise RMSEs of the completed matrix.

Fairness disparities can arise from different characteristics of the data. *Population imbalance* refers to differences in the number of rows represented by different groups, while *observation bias* refers to differences in the observation patterns across groups. In addition, groups may differ in their latent structure, so that a single low-rank representation may describe some groups more accurately than others. These sources of disparity can affect the reconstruction accuracy experienced by different groups.

Fairness interventions in machine learning can generally be classified as *pre-processing*, *in-processing*, or *post-processing* methods. Pre-processing methods modify the data before model training, for example by reweighting or resampling observations. In-processing methods modify the learning procedure or objective to account for fairness while fitting the model. Post-processing methods modify the model's predictions after training to satisfy a desired fairness criterion.

In this dissertation, we focus on in-processing approaches to error-based fairness in matrix completion. We consider fairness regularization based on statistical parity and equal opportunity, as well as a group-aware nuclear norm objective that assigns separate nuclear norm terms to the submatrices corresponding to each group in addition to the overall nuclear norm. We investigate how these approaches affect overall and group-wise reconstruction accuracy under different sources of disparity, and whether improvements in parity-based or accuracy-based fairness translate into improved error-based fairness.

CHAPTER 2

Stratified Non-negative Tensor Factorization

2.1 Introduction

Unsupervised methods for understanding large, multi-modal datasets have become increasingly important as data complexity has continued to grow. Non-negative matrix factorization (NMF) and non-negative tensor factorization (NTF) for higher-mode data are popular methods for factorizing large-scale data into interpretable, low-rank components [SH05, LS00, LS99]. Given a non-negative data matrix $A \in \mathbb{R}_+^{m \times n}$, NMF decomposes the data into non-negative factor matrices $W \in \mathbb{R}_+^{m \times r}$ and $H \in \mathbb{R}_+^{r \times n}$ that minimize the objective

$$\underset{W,H}{\operatorname{argmin}} \|A - WH\|_F^2. \tag{2.1.1}$$

In this work, the data matrix is organized with rows corresponding to samples and columns corresponding to variables. The topics matrix H gives a correspondence between topics and variables, and the coding matrix W weights how much each topic is represented in the samples. While this decomposition applies directly to data when each sample is structured as a vector, it fails to accommodate data with tensor-like structure without reshaping. For example, flattening an image into a vector removes the spatial structure contained implicitly within the tensor. One approach to address this problem is the Non-negative CANDECOMP/PARAFAC (CP) tensor Decomposition (NCPD), which extends NMF to handle tensor data by approximating a mode-n data tensor $\mathcal{A} \in \mathbb{R}_+^{\prod_{k=1}^n d_k}$ by rank-r components [CC70, Har70]. In particular, NCPD solves the minimization problem

$$\underset{u_j^k}{\operatorname{argmin}} \|\mathcal{A} - \sum_{j=1}^{r} \otimes_{k=1}^{n} u_j^k\|_F^2 \tag{2.1.2}$$

where $u_j^k \in \mathbb{R}_+^{d_k}$ denotes the kth vector used to form the jth rank-one outer product. In this work, we use *rank* to denote the CP-rank [KB09].

In some settings it is desirable to account for data obtained from different sources. For example, data from hospitals across different states reflect separate populations, or experiments run on different equipment may have varying error distributions. In these cases we may wish to use *stratified* methods, which can learn both stratum-dependent information as well as shared information across different sources [HHM53]. The recent Stratified-NMF method extends NMF to consider stratified data that share global topics but differ by positive stratum-dependent shifts, and demonstrates improved performance across data modalities [CYN23].

Given s strata and data matrices for each stratum $A(i),\ 1 \leq i \leq s$, Stratified-NMF minimizes the objective

$$\underset{v(i),W(i),H}{\operatorname{argmin}} \sum_{i=1}^{s} \|A(i) - \mathbf{1}v(i)^T - W(i)H\|_F^2,$$

such that

$$v(i)^T \geq 0,\ W(i) \geq 0,\ H \geq 0,\ \forall 1 \leq i \leq s,$$

where $\mathbf{1}_m \in \mathbb{R}_+^m$ is the vector of all ones, $v(i)$ is the stratum-dependent shift for each stratum, $W(i)$ is the coding matrix for each stratum, and H is the global topics matrix. In this formulation, topics are learned across strata globally, while also allowing for the model to learn characteristics unique to each strata in the vectors $v(i)$ (stratum-level features).

In this paper, we create a Stratified-NTF method that extends Stratified-NMF to tensor data in order to leverage both geometric information contained in tensor data as well as stratification to handle heterogeneous datasets. We derive multiplicative updates to efficiently

optimize the objective as is common of NMF and its variants. In experiments, we demonstrate the effectiveness and interpretability of the method and compare it to existing algorithms. Finally, we derive multiplicative updates for Stratified-NTF with TV-regularization and evaluate it on noisy image data. The code for this paper is made publicly available[1].

2.2 Proposed Method

Given a non-negative set of data tensors $\mathcal{A}(i) \in \mathbb{R}_+^{d_1(i) \times \prod_{k=2}^n d_k}$, we propose the following Stratified-NTF objective:

$$\underset{\mathcal{V}(i), w(i)_j, \mathcal{H}_j}{\operatorname{argmin}} \sum_{i=1}^{s} \|\mathcal{A}(i) - \mathbf{1}_{d_1(i)} \otimes \mathcal{V}(i) - \sum_{j=1}^{r} w(i)_j \otimes \mathcal{H}_j\|_F^2, \tag{2.2.1}$$

where

$$\mathcal{V}(i) = \sum_{j=1}^{r'(i)} \mathcal{V}(i)_j, \quad \mathcal{V}(i)_j = \otimes_{k=2}^n \, v(i)_j^k, \quad \mathcal{H}_j = \otimes_{k=2}^n \, h_j^k,$$

such that

$$v(i)_j^k \in \mathbb{R}_+^{d_k}, \; w(i)_j \in \mathbb{R}_+^{d_1(i)}, \; h_j^k \in \mathbb{R}_+^{d_k}, \; \forall i, j, k.$$

We define $\mathcal{V}(i)_j$ to be the rank-one strata tensors that we sum over to get rank $r'(i)$ stratum-level features $\mathcal{V}(i)$. Note that $\mathcal{V}(i)$ produces strata tensors analogous to the stratum-level features $v(i)$ obtained in Stratified-NMF which learn information common to each sample in the stratum.

We additionally extend the method to include higher-rank decomposition of the stratum-dependent shifts. This is necessary in the tensor setting since rank-one feature information may underfit the data [SH05]. For example, when modeling image data, a single outer product can only generate axis-aligned pixel clusters in the image. By increasing the rank,

[1] Code can be found at https://github.com/alexandersietsema/Stratified-NTF

the stratum-level features can learn sums of such clusters to represent more complex strata specific information.

We derive multiplicative updates to efficiently optimize the objective. Let

$$\mathcal{B}(i) = \mathbf{1}_{d_1(i)} \otimes \mathcal{V}(i) + \sum_{j=1}^{r} w(i)_j \otimes \mathcal{H}_j \tag{2.2.2}$$

be the Stratified-NTF approximation of $\mathcal{A}(i)$. Then the multiplicative updates are

$$v(i)^{\tau}_{lk} \leftarrow v(i)^{\tau}_{lk} \frac{\sum_j \sum_{\alpha \in S^\tau} \mathcal{V}(i)_{l,\alpha} \mathcal{A}(i)_{j,\alpha+ke_\tau}}{\sum_j \sum_{\alpha \in S^\tau} \mathcal{V}(i)_{l,\alpha} \mathcal{B}(i)_{j,\alpha+ke_\tau}} \tag{2.2.3}$$

$$w(i)_{lk} \leftarrow w(i)_{lk} \frac{\sum_{\alpha \in S} \mathcal{H}_{l,\alpha} \mathcal{A}(i)_{k,\alpha}}{\sum_{\alpha \in S} \mathcal{H}_{l,\alpha} \mathcal{B}(i)_{k,\alpha}} \tag{2.2.4}$$

$$h^{\tau}_{lk} \leftarrow h^{\tau}_{lk} \frac{\sum_{ij} \sum_{\alpha \in S^\tau} (w(i)_l \otimes \mathcal{H}_l)_{j,\alpha} \mathcal{A}(i)_{j,\alpha+ke_\tau}}{\sum_{ij} \sum_{\alpha \in S^\tau} (w(i)_l \otimes \mathcal{H}_l)_{j,\alpha} \mathcal{B}(i)_{j,\alpha+ke_\tau}} \tag{2.2.5}$$

where S is the set of length $n-1$ multi-indices, and S^τ is the set of length $n-1$ multi-indices with the τth entry 0. We use the convention that the zeroth index of a vector is defined to be 1 for ease of notation and we use e_j to denote the jth standard basis vector. These updates are written entry-wise for each vector and computed for all l, k, τ indexed as in (2.2.1). The $\mathcal{B}(i)$ tensor is computed based on the existing values as in (2.2.2) in each update step.

2.2.1 Regularization

In this section, we develop a Total Variation (TV) regularized version of Stratified-NTF for image denoising [RO94]. TV regularization is widely utilized in image processing and has been effectively applied in tensor decompositions for promoting smoothness in applications such as hyperspectral unmixing [XQZ19]. Unlike NMF methods, which require flattening images into a single mode, tensor methods retain spatial information which we can directly

apply regularization to. We apply regularization to the global topic parameters $\mathcal{H}$ of form

$$||h_l^\tau||_{TV} = \sum_k |h_{l,k+1}^\tau - h_{l,k}^\tau|. \tag{2.2.6}$$

For application to 3-mode image data, we regularize separately over h^2 and h^3, the spatial modes of $\mathcal{H}$, as

$$\begin{aligned} \underset{\mathcal{V}(i),w(i)_j,\mathcal{H}_j}{\operatorname{argmin}} \sum_{i=1}^{s} \|\mathcal{A}(i) - \mathbf{1}_{d_1(i)} \otimes \mathcal{V}(i) - \sum_{j=1}^{r} w(i)_j \otimes \mathcal{H}_j\|_F^2 \\ + \lambda \sum_l \left(||h_l^2||_{TV} + ||h_l^3||_{TV} \right). \end{aligned} \tag{2.2.7}$$

The gradient of the TV term for fixed τ is

$$\nabla||h_l^\tau||_{TV} = \begin{bmatrix} -\text{sign}(h_{l,1}^\tau - h_{l,0}^\tau) \\ -\text{sign}(h_{l,2}^\tau - h_{l,1}^\tau) + \text{sign}(h_{l,1}^\tau - h_{l,0}^\tau) \\ -\text{sign}(h_{l,3}^\tau - h_{l,2}^\tau) + \text{sign}(h_{l,2}^\tau - h_{l,1}^\tau) \\ \vdots \\ -\text{sign}(h_{l,k+1}^\tau - h_{l,k}^\tau) + \text{sign}(h_{l,k}^\tau - h_{l,k-1}^\tau) \\ \vdots \\ \text{sign}(h_{l,d_\tau}^\tau - h_{l,d_\tau-1}^\tau) \end{bmatrix}. \tag{2.2.8}$$

We define $\nabla^+||h_l^\tau||_{TV}$ and $\nabla^-||h_l^\tau||_{TV}$ to be the vector of positive and negative entries of the gradient, respectively, and zeros elsewhere as

$$\nabla^+||h_l^\tau||_{TV} = \max\{0, \nabla||h_l^\tau||_{TV}\} \tag{2.2.9}$$

$$\nabla^-||h_l^\tau||_{TV} = -\min\{\nabla||h_l^\tau||_{TV}, 0\}, \tag{2.2.10}$$

where max and min are evaluated entry-wise. This results in the regularized multiplicative update equation

$$h_{lk}^{\tau} \leftarrow h_{lk}^{\tau} \frac{\sum_{ij} \sum_{\alpha \in S^{\tau}} (w(i)_l \otimes \mathcal{H}_l)_{j,\alpha} \mathcal{A}(i)_{j,\alpha+ke_\tau} + \lambda \nabla^{-} ||h_l^{\tau}||_{TV}}{\sum_{ij} \sum_{\alpha \in S^{\tau}} (w(i)_l \otimes \mathcal{H}_l)_{j,\alpha} \mathcal{B}(i)_{j,\alpha+ke_\tau} + \lambda \nabla^{+} ||h_l^{\tau}||_{TV}}. \tag{2.2.11}$$

We normalize each h_l^{τ} after each update to stabilize the method. This is a standard technique in ℓ_1-norm regularization for NMF methods [EK04]. The full algorithm is shown in Algorithm 1.

Algorithm 1 Stratified-NTF Multiplicative Update Algorithm

Input: Data $\mathcal{A}(i) \in \mathbb{R}_+^{d_1(i) \times \prod_{k=2}^{n} d_k}$ for $1 \leq i \leq s$, decomposition ranks $r, r'(i)$, number of iterations N, number of strata updates per iteration M, regularization strength λ
Initialize $v(i)_{j'}^{k} \in \mathbb{R}_+^{d_k}, w(i)_j \in \mathbb{R}_+^{d_1(i)}, h_j^k \in \mathbb{R}_+^{d_k}$ for $1 \leq j \leq r,\ 1 \leq j' \leq r'(i)$
for N iterations **do**
    for M iterations **do**
        for τ modes **do**
            Update $v(i)^{\tau}\ \forall i$ in each entry according to (2.2.3)
        end for
    end for
    Update $w(i)_j\ \forall i, j$ in each entry and rank according to (2.2.4)
    for τ modes **do**
        if regularization **then**
            Update $h_j^{\tau}\ \forall j$ in each entry according to (2.2.11)
            Normalize $h_j^{\tau}\ \forall j$
        else
            Update $h_j^{\tau}\ \forall j$ in each entry according to (2.2.5)
        end if
    end for
end for

2.3 Experimental Results

In this section, we demonstrate the effectiveness and interpretability of Stratified-NTF applied to text and image data. When computing the updates (2.2.3), (2.2.4), (2.2.5), we clip each numerator and denominator to machine precision tolerance to avoid zero values for numerical stability [KPA19]. All parameters are randomly initialized from independent and

identically distributed $[0, 1]$ uniform distributions. We compute $\mathcal{V}(i)$ updates twice in each iteration, corresponding to the choice $M = 2$ in Line 4 of Algorithm 1.

2.3.1 20 Newsgroups

We first apply Stratified-NTF to the 20 newsgroups dataset [RL08]. The dataset contains document data from 20 newsgroups of various super-categories and sub-categories as summarized in Table 2.1. We preprocess the data by removing headers, footers, and quotes, and applying TF-IDF vectorization for the top 5000 words, ignoring words which appear in over 95% of the documents. We construct our dataset with strata consisting of the five super-categories, and for each stratum we construct the data tensor with axes specified by

$$\text{sub-category} \times \text{document} \times \text{word}.$$

The learned parameters of the Stratified-NTF method should thus represent the correspondence between each learned topic and each corresponding mode. In particular, $w(i)_j$ learns the correspondence between the jth topic and each sub-category for a super-category i, h_j^2 learns the correspondence between the jth topic and each document, and h_j^3 learns the correspondence between the jth topic and each word. This three-mode representation allows us to characterize topics in finer detail than is possible with existing matrix methods, which have only two modes in which to learn topic information. As the factorization is non-negative, these parameters yield a naturally interpretable linear representation of the data tensor as a product of the sub-category, document, and word correspondences.

In order for the document dimension to align between tensors of different strata, we choose 100 documents for each sub-category. We apply Stratified-NTF to this dataset with strata rank 3 and topic rank 5 for 11 iterations. Because of the natural sparsity of text data, we observe almost immediate convergence.

In Table 2.1, we identify the top three highest correspondence words for each strata

Super-category	Sub-categories		
Computers	Graphics	MS Windows	PC Hardware
	Mac Hardware	X Windows	
Recreation	Autos	Motorcycles	Baseball
	Hockey		
Science	Crypto	Electronics	Medicine
	Space		
Politics	Guns	Middle East	Miscellaneous
Religion	Atheism	Christianity	Miscellaneous

Table 2.1: Super-categories and sub-categories for the 20 Newsgroups dataset.

feature, that is, the words corresponding to the three highest values in each $v(i)_j^3$. We see that the stratum-level features capture relevant topic information for each stratum. For example, the "Recreation" stratum contains features corresponding to the sports-related "Hockey" and "Baseball" sub-categories (the feature "Players", "Year", "Flyers") as well as to the "Motorcycles" sub-category (the feature "Bike", "Miles", "Game"). Similarly, the "science" stratum contains a feature related to the "Crypto" sub-category (the feature "Key", "Private", "Keys").

Similarly, we can identify the highest correspondence documents for each strata feature by finding the largest values in the document mode for $v(i)_j^2$. We see that such documents are indeed topical: for example, for the hockey-related feature ("Players", "Year", "Flyers"), the document corresponding to the largest value of $v(2)_1^2$ in the "Hockey" mode discusses a potential player trade between the Philadelphia Flyers and Quebec Nordiques in the National Hockey League. While this requires human interpretation of the stratum-level features to identify the correct sub-category mode in which to find the related document, this demonstrates that the learned stratum-level features also reflect the content of individual documents. Overall, we observe that for this dataset, the method is able to identify interpretable stratum-level features that are both relevant to the super-category of each stratum, as well as relevant to the sub-categories contained in each stratum's data tensor.

Super-category	1	2	3
Computers	Thanks	Use	Mail
	Files	Apps	Windows
	Window	Tiff	Problem
Recreation	Players	Year	Flyers
	Bike	Miles	Game
	Car	Think	Ya
Science	Key	Private	Keys
	Clipper	Chip	Encryption
	Space	Electronics	Amp
Politics	Armenians	Armenian	Firearms
	Turkish	Gun	Government
	Armenian	Armenians	Gay
Religion	Sin	People	Christian
	God	Jesus	Father
	Objective	Morality	Moral

Table 2.2: The top three most relevant words learned for each strata feature. The three rows within a super-category i correspond to each of $v(i)_1^3, v(i)_2^3, v(i)_3^3$.

2.3.2 Olivetti Faces

In this section, we compare the performance of NTF, Stratified-NMF, and Stratified-NTF on the Olivetti Faces Dataset constructed by AT&T Laboratories Cambridge [SH94] as provided by the scikit-learn package [PVG11]. The Olivetti Faces Dataset consists of 10 images of each of 40 individuals taken from multiple angles all of dimension 64×64. To create our stratified dataset, we stratify over each individual to create 40 data tensors where each stratum is of dimension $10 \times 64 \times 64$.

As a baseline comparison, we use a standard NTF implementation from Tensorly [KPA19] with the loss updated to provide loss values matching (2.1.2). Since NTF is not a stratified method, we input the data tensor concatenated across strata of shape $400 \times 64 \times 64$. For Stratified-NMF, we flatten the data tensor across the last two dimensions to create data matrices for each stratum of shape 10×4096. In order to make fair comparisons across methods, we pick hyperparameters for each method to yield as close to the same total

	Stratified-NTF	NTF	Stratified-NMF
Topic rank	40	186	1
Strata rank	15	-	1
Total params	∼97K	∼98K	∼168K

Table 2.3: Learnable parameters for Stratified-NTF, NTF, and Stratified-NMF on the faces experiment.

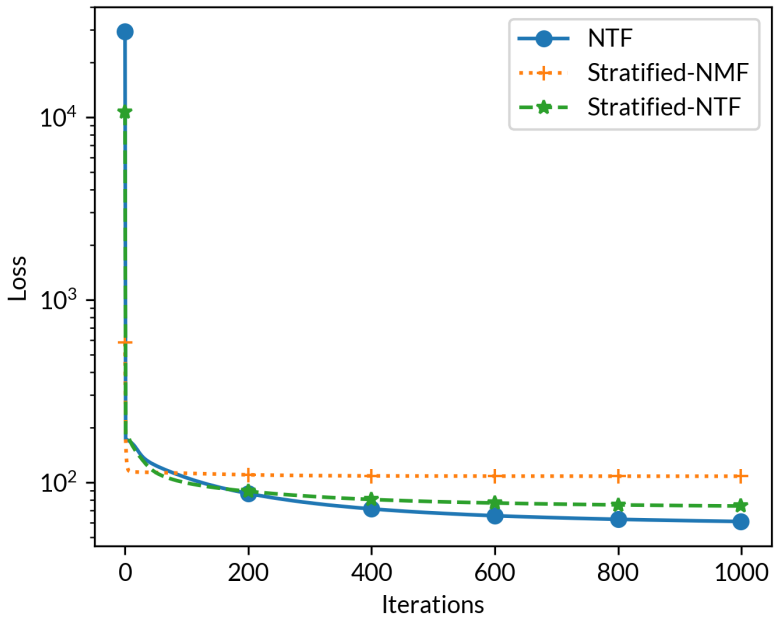
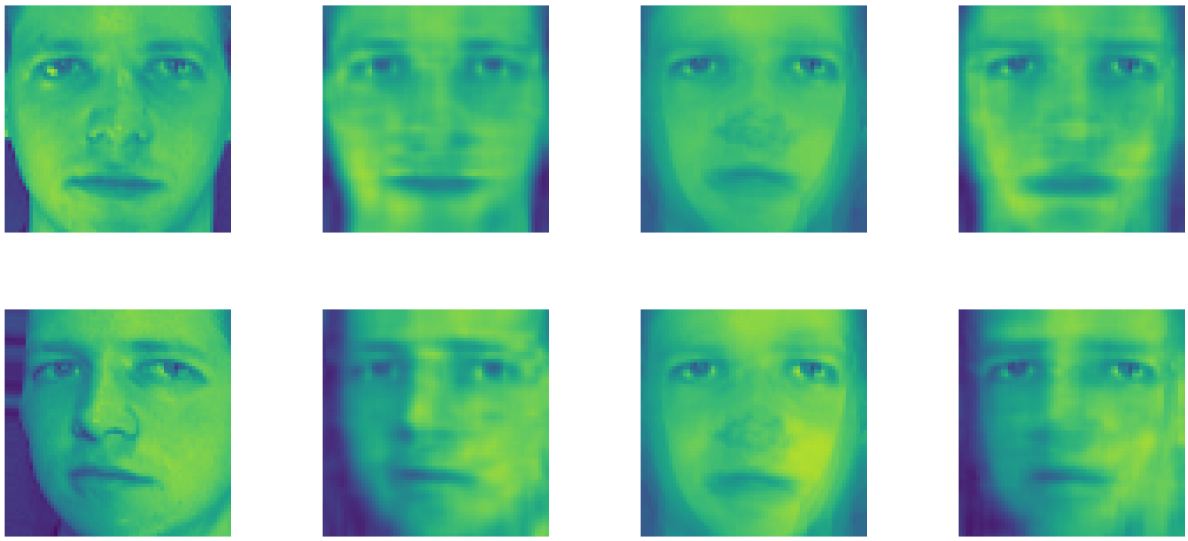


Figure 2.1: Loss plots for NTF, Stratified-NMF, and Stratified-NTF run for 1000 iterations on the Olivetti faces dataset. The rapid convergence shows that multiplicative updates are able to efficiently optimize the objective, matching existing methods.

number of learnable parameters, as shown in Table 2.3. Note that as Stratified-NMF uses flattened data, each rank is an entire image, so the closest possible comparison uses a strata feature rank of 1 (fixed in Stratified-NMF) and a topic rank of 1 for a total of ∼168K parameters.

Figure 2.1 compares the convergence of each method over 1000 iterations. The multiplicative update method for Stratified-NTF described in Algorithm (1) converges on the Olivetti dataset with a final loss of 74.115. We see nearly immediate convergence, which highlights the efficiency of the multiplicative updates. This outperforms Stratified-NMF,

which obtains a final loss of 107.921 on this dataset. We also see that NTF modestly outperforms Stratified-NTF. This is not unexpected: though the total number of parameters is similar for each method, for a particular image, Stratified-NTF has access to only 65 total topics (the global topics and one set of stratum-level features). On the other hand, NTF has access to all 186 topics, so NTF has more flexibility for a given image. However, these performance gains are offset by the added interpretability of learning both global and local information as provided by stratified methods.



(a) Ground truth (b) NTF (c) Stratified-NMF (d) Stratified-NTF

Figure 2.2: Reconstruction comparison of NTF, Stratified-NMF, and Stratified-NTF on the first (top) and fourth (bottom) images from the first stratum. The NTF and Stratified-NTF reconstructions appear qualitatively similar, while the Stratified-NMF reconstructions fail to capture the different viewing angle between the two images due to differences in parameterization.

Figure 2.2 shows a comparison of the resulting reconstructions from each method for the first pair of images of the first individual with different viewing angles. Qualitatively, we see that NTF and Stratified-NTF perform similarly, while the Stratified-NMF reconstruction is far less detailed and displays color banding around the edges of the face. Additionally, recon-

structed images for Stratified-NMF appear front-facing while the tensor methods are able to capture the change in head angle. This highlights the added expressiveness of Stratified-NTF compared to Stratified-NMF, even while using only just over half the same number of parameters. This indicates that the tensor methods may be more memory-efficient for this dataset.

2.3.3 Noisy, watermarked MNIST

To evaluate the performance of Stratified-NTF with TV-regularization, we conduct an experiment on a noisy, watermarked version of the MNIST dataset [LCB10]. The MNIST dataset consists of images of handwritten digits from the Torchvision library [MR10], which we rescale to floating-point values between 0 and 1. To construct our experiment data, we create two strata: the first consists of 100 images of 1's and 100 images of 2's, all with a text watermark overlay of the word "ONE" positioned at the top of the image. The second stratum consists of 100 images of 2's and 100 images of 3's, all with a text watermark overlay of the word "TWO" at the bottom of the image. We add salt-and-pepper noise to all images by randomly setting pixels of each image to 0 and 1, both with probability 0.15. We note that TV regularization is a standard technique for removing salt-and-pepper noise [Rod13, Nik04]. We run TV-regularized Stratified NTF on this dataset across regularization parameters of $\lambda = 0.0, 5.0, 10.0$, each for 100 iterations with strata feature rank 100 and topic rank 100.

As the text watermarks are present uniformly in each image of each stratum, we expect the stratum-level features to identify and separate out the watermarks, effectively performing watermark removal on the dataset. As the strata differ in the contained digits as well as the watermarks, Stratified-NTF must capture both the fine detail of the text as well as the larger forms of the digits. By regularizing only the topic parameters, we expect to denoise the digits while not affecting text learned by the stratum-level features. Unlike the topic parameters, the stratum-level features are added uniformly to each image in each stratum,

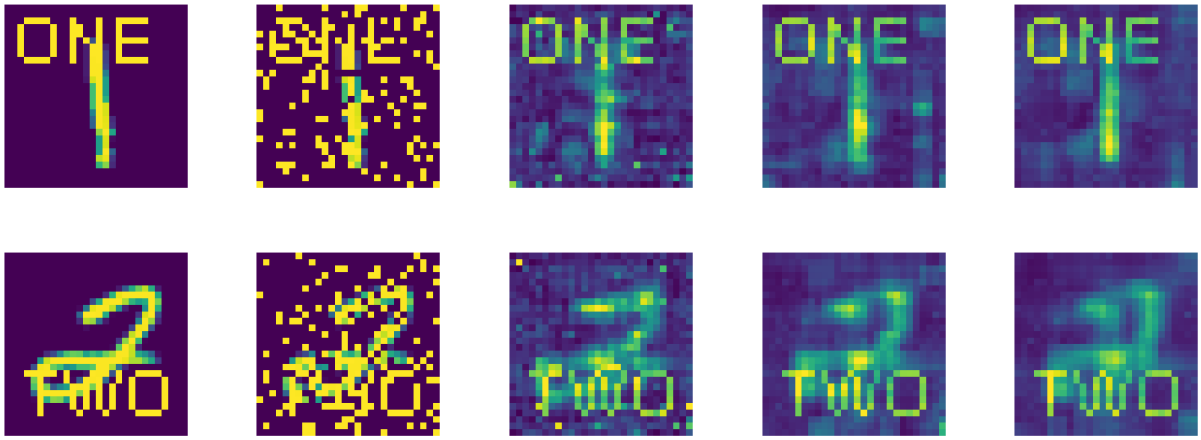
so they cannot overfit to the image-specific noise patterns. Therefore regularization should only be necessary in the $\mathcal{H}_j$ tensors to achieve denoising in all of the learned parameters.

In Fig. 2.3, we compare the reconstructions of the first image in each stratum across noise levels. We observe that increasing the regularization parameter significantly reduces the amount of noise present in the reconstructions, while also preserving the text watermarks. The digit 1 in the first stratum images (top row) is significantly clearer with higher regularization parameters, while the digit 2 in the second stratum (bottom row) is quite blurry. We suspect that this is due to the shape of the digits interacting with the total variation of the topics: the digit 1 is nearly convex, which has a naturally lower total variation than the top and loop of the digit 2, so the model may have an easier time reconstructing the 1 in the presence of the regularization, especially given the high level of corruption in the data.

Overall, while this formulation of TV regularization does not lead to dramatic improvements in reconstruction quality, we are still able to see the desired denoising effects. We note that this is a simple example to demonstrate the flexibility of mode-specific regularization compared to matrix methods like Stratified-NMF; there are a number of spatial and temporal regularization techniques for tensor factorizations which may be introduced [ABC17], as well as classical regularization approaches for NMF which may be extended to the tensor case [SG17].

2.4 Discussion

The extension of NTF and Stratified-NMF to Stratified-NTF allows practitioners to combine the benefits of both methods. Practitioners can now leverage geometric priors in tensor data through an explicit tensor representation and with regularization, while also handling heterogeneous data. This work also opens the door to some important open questions regarding



(a) Watermarked Data (b) Noisy Watermarked Data (c) $\lambda = 0.0$ (d) $\lambda = 5.0$ (e) $\lambda = 10.0$

Figure 2.3: Reconstruction comparison of TV-regularized Stratified-NTF on noisy watermarked data across different regularization parameters λ. Each column shows the first image of the first stratum (top) and of the second stratum (bottom). We observe that stronger regularization decreases noise in the resulting reconstructions.

- Principled understanding of hyperparameter choices
- Investigation of additional regularization methods
- Regularization interacting between stratum-level features and global features
- Proofs of convergence of the multiplicative updates [SH05, CYN23, Lin07].

2.5 Conclusion

In this work, we extend the Stratified-NMF method into a stratified non-negative tensor decomposition. Stratified-NTF can learn both global and local information from multi-modal data from different sources. We develop multiplicative update rules for the method as well as for a total-variation regularized variant. We demonstrate the effectiveness of the standard method on text and image data, and demonstrate the regularized method's effectiveness for noise reduction on noisy, watermarked image data.

CHAPTER 3

On the Fairness of Matrix Completion

3.1 Introduction

Machine learning systems are increasingly used to make predictions and support decisions across a wide range of domains, including healthcare, finance, education, criminal justice, and advertising. As these systems have become more influential, there is increasing recognition that data-driven systems can produce or perpetuate biased or unfair outcomes. Such outcomes can arise from multiple sources, including bias in the data and bias introduced by algorithmic choices. A substantial body of work has developed around fairness in machine learning, proposing a variety of fairness definitions and interventions [BHN23, MMS21]. In this work, we focus on disparities in how machine learning models perform across different populations represented in a dataset. Models are generally optimized for overall performance, which does not necessarily ensure equitable treatment across groups. What is considered *fair* is often application-dependent, and different fairness definitions impose different, and sometimes competing, requirements. Since fairness choices rest with human designers, understanding the trade-offs introduced by fairness objectives is an important area of study.

In this work, we focus on matrix completion, a problem in data analysis that seeks to infer unobserved entries of a matrix from a sparse set of observations. Matrix completion arises in a variety of applications, including imaging, computer vision, genomics, and recommender systems [CP10]. Specifically, we consider matrix completion via nuclear norm minimization, a convex optimization method that can recover an unknown low-rank matrix from a relatively

small number of observed entries. Nuclear norm minimization provides a convex relaxation of rank minimization and is therefore well suited to settings in which the underlying data are reasonably modeled as low rank. However, when data contain distinct populations, a single global low-rank model may not represent all groups equally well. In particular, if groups differ in their observation rates, underlying distributions, noise levels, or relationships among features, the global reconstruction may favor majority group patterns that are more strongly represented in the data and yield less accurate recovery for other groups [KGN24].

These disparities can be consequential when predictions are used for downstream decisions. For example, in healthcare applications, statistical models are used to estimate disease risk or inform treatment decisions. a model that performs well overall may nevertheless produce less accurate predictions for a minority population. Evaluating only aggregate performance may therefore obscure disparities in the quality of predictions across groups and contribute to inequities in patient care.

Such concerns motivate the development of fairness criteria for various machine learning tasks. Fairness in machine learning has been studied in classification and regression settings [ZVR17, ZWS13], and more recent work has considered fairness in recommender systems [BCD19, TA22, YH17, LCX23, WMZ23]. These works study how fairness criteria can be incorporated into predictive models or used to evaluate disparities in model outcomes across groups. Traditionally, matrix completion methods focus primarily on accurately predicting missing entries. Evaluating fairness in matrix completion predictions remains relatively less explored. Recent work has adapted parity- and accuracy-based criteria to define fairness metrics for collaborative filtering [YH17]. This motivates our study of optimization approaches and fairness metrics for matrix completion, and of how different fairness criteria and formulations affect overall and group-wise reconstruction accuracy, particularly for minority groups.

3.1.1 Contribution

This work makes several contributions to the study of fairness in matrix completion. First, we formalize two commonly studied fairness criteria, statistical parity and equal opportunity, in the matrix completion setting. We incorporate these criteria into the nuclear norm objective for matrix completion through regularization, and additionally study the fairness of a group-aware matrix completion objective that separately regularizes the submatrices corresponding to different groups, allowing the model to accommodate group-specific low-rank structure. We then use synthetic and real data experiments to systematically investigate how population imbalance and differences in group-specific structure contribute to disparities in reconstruction accuracy, and how these disparities are affected by different fairness interventions. Our results show that the effectiveness of a fairness intervention depends on both the source of disparity and the fairness criterion being enforced: some interventions reduce disparities in reconstruction accuracy in particular settings, while others have little effect or can even worsen them. More broadly, our findings demonstrate that satisfying a parity-based fairness criterion does not necessarily lead to improved fairness in reconstruction error, highlighting the importance of evaluating fairness objectives in relation to the underlying data structure and the particular notion of fairness being targeted.

3.1.2 Organization

In Section 3.2, we review related work on fairness in machine learning, with an emphasis on fairness in recommender systems. Section 3.3 provides the mathematical background for matrix completion via nuclear norm minimization and formalizes statistical parity and equal opportunity for the matrix completion setting. In Section 3.4, we present two in-processing approaches for incorporating fairness into matrix completion: fairness-metric regularization and a group-aware matrix completion objective. Section 3.5 presents numerical experiments on a range of synthetic and real datasets to evaluate the group-wise reconstruction accuracy

of these approaches under different sources of disparity. Section 3.5.6 concludes with a discussion of the main findings and directions for future work.

3.2 Related Works

3.2.1 Fairness Metrics in Machine Learning

Data used to train machine learning systems can be biased in several, often overlapping ways. A demographic group may be underrepresented in the data (*population imbalance*), a group may have systematically fewer observed labels or ratings (*observation bias*), different groups' data may be corrupted by different levels of noise, or groups may genuinely differ in their underlying distributions [YH17]. Fairness criteria provide different ways to formalize and evaluate disparities that arise from such sources.

Parity-based criteria require that model outputs be statistically independent of a protected attribute. The most common such criterion, *statistical parity* (sometimes refered to as *demographic parity*), requires that for a binary decision $\hat{Y} \in \{0, 1\}$ and binary protected attribute $Z \in \{z, z'\}$,

$$\Pr\{\hat{Y} = 1 \mid Z = z\} = \Pr\{\hat{Y} = 1 \mid Z = z'\}.$$

[HPS16a] note two limitations of this criterion, following [DHP11]: it does not itself guarantee fairness, since it permits accepting qualified individuals in one group while accepting unqualified individuals in another, so long as acceptance rates match; and it can reduce utility when the true outcome Y is genuinely correlated with A, since the ideal predictor $\hat{Y} = Y$ may itself violate statistical parity without being discriminatory.

Other parity-based criteria instead condition on the true outcome. As defined by [HPS16a], a predictor $\hat{Y}$ satisfies *equalized odds* with respect to Z and true outcome Y if $\hat{Y}$ and Z are

independent conditional on Y. In the binary case, this is amounts to requiring

$$\Pr\{\hat{Y} = 1 \mid Z = z, Y = y\} = \Pr\{\hat{Y} = 1 \mid Z = z', Y = y\}, \quad y \in \{0, 1\},$$

i.e. equal true positive and false positive rates across groups.

Equal opportunity is the corresponding relaxation obtained by requiring only equal true-positive rates ($y = 1$). It is generally easier to satisfy than equalized odds because it does not additionally constrain predictions for negative instances. Equal opportunity is described as an *accuracy-based* fairness criterion because it compares a measure of predictive performance across groups.

These criteria illustrate that fairness can be defined with respect to different aspects of a prediction, and that the choice of criterion depends on the goals and assumptions of the application. To our knowledge, statistical parity and equal opportunity have not previously been formulated explicitly for the matrix completion setting. In this work, we formalize these two criteria for matrix completion. We refer to the goal of achieving comparable reconstruction error across groups as *error-based fairness*. Specifically, we evaluate error-based fairness by comparing group-wise reconstruction errors in the completed matrix using the group-wise RMSE defined in (3.5.1). We then investigate whether enforcing parity-based or accuracy-based criteria also improves error-based fairness across demographic groups.

3.2.2 Fairness Notions in Recommender Systems

Fairness in recommender systems can be broadly considered from the perspectives of *user fairness* and *item or provider fairness*. User fairness focuses on whether different groups of users receive comparable recommendation quality or treatment, regardless of sensitive attributes such as race, gender, or age. Item or provider fairness, by contrast, focuses on the fair representation and exposure of items or providers in recommendation results [BSO18]. For example, [TA22] considers the Gini index of average predicted scores as a measure of

inequality in item exposure.

Within user fairness, statistical parity may be appropriate when differences in recommendations across groups are viewed as reflecting bias rather than genuine differences in preferences that a recommender system should preserve. As [YH17] describe, for example, in a course recommendation setting, one may wish to recommend STEM courses to female and male students at equal rates if observed gender differences in historical enrollment are attributed to structural and societal barriers rather than to underlying differences in interest. In such a setting, enforcing statistical parity intentionally prevents the recommendation system from reproducing disparities present in the observed data.

Alternatively, one may focus on having prediction accuracy be comparable across groups. For example, in an educational assessment setting, one may seek to predict students' exam scores with comparable accuracy across demographic groups while still allowing the groups to have different underlying score distributions. This motivates considering accuracy-based fairness criteria, such as *equal opportunity*.

Additional notions of fairness have been considered in matrix-factorization-based collaborative filtering. [YH17] propose four fairness metrics–*value unfairness*, *absolute unfairness*, *underestimation unfairness*, and *overestimation unfairness*–that measure different forms of signed and unsigned prediction error across groups.

The existing literature thus illustrates that fairness in recommendation can concern either the distribution of predictions across groups or the accuracy of those predictions. This distinction is particularly relevant to matrix completion, where demographic groups may differ in their representation in the observed data and in the structure underlying their entries.

3.2.3 Fairness in Unsupervised Learning

In unsupervised learning settings, such as dimensionality reduction and matrix factorization, fairness-aware objectives have been proposed that seek a common representation with comparable reconstruction quality across demographic groups. Samadi et al. [STM18] introduce *Fair PCA*, which seeks a single low-dimensional representation that minimizes the maximum average reconstruction loss across groups. This formulation addresses a potential limitation of standard PCA, whose global objective may place greater emphasis on groups that contribute more to the overall reconstruction loss. Tantipongpipat et al. [TSS20] subsequently generalize this framework through a multi-criteria formulation that considers multiple group-level objectives.

Kassab et al. [KGN24] extend a similar perspective to nonnegative matrix factorization through *Fairer-NMF*. Their approach minimizes the maximum relative reconstruction loss across groups, with the loss normalized to account for differences in group size and intrinsic complexity. This work is particularly relevant to matrix completion because it also demonstrates that disparities in reconstruction accuracy can arise not only from differences in group representation, but also from differences in the complexity of group-specific data. Inspired by these works, we likewise consider a group-aware formulation of the matrix completion objective and evaluate how it affects reconstruction accuracy across groups.

3.2.4 Matrix Estimation with Group Structure

Several works have considered how information about groups or other attributes can be incorporated into matrix estimation. *Side information* refers to information about the rows or columns of a matrix beyond the observed matrix entries, such as categorical group membership or other observed attributes. Herbster et al. [HPT20] study online matrix completion with side information and show that the quality of the side information can affect the complexity of the matrix completion problem. In particular, when side information is predictive

of the underlying factorization, their notion of quasi-dimension can be substantially smaller than the corresponding complexity in the absence of informative side information. This suggests that information associated with groups or other observed attributes can be useful for recovering the underlying structure of a matrix. However, incorporating categorical group membership directly as side information also requires specifying how that information relates to the matrix. If that assumed relationship does not accurately describe the data, the resulting estimator can be biased.

A related approach is proposed by Golubovic et al. [GSA26] through *Group-Aware Matrix Estimation* (GAME), a framework for matrix estimation with multiple, potentially overlapping groups that applies group-specific nuclear norm penalties while jointly estimating the full matrix. This approach uses categorical group information to identify submatrices that may exhibit distinct low-rank structure without assuming a particular mechanism through which group membership generates the observed entries. The resulting objective allows subgroup-specific structure to be preserved while sharing information across groups, and their theoretical and empirical results demonstrate benefits when groups exhibit distinct low-rank structure or uneven observation patterns. Our work considers a related setting in which demographic groups may differ in their representation and underlying structure, but focuses specifically on measuring fairness in reconstruction accuracy between groups rather than solely on minimizing overall reconstruction error.

3.2.5 Impossibility Results in Fairness

Kleinberg et al. [KMR16] show that, for risk scores predicting a binary outcome, three natural fairness conditions–calibration within each group, balance of average scores for the positive class, and balance of average scores for the negative class–cannot generally be satisfied simultaneously. All three conditions can hold only in special cases, such as when the groups have equal base rates or the outcome can be predicted perfectly. Pleiss et al. [PRW17] extend this analysis by showing that even a relaxed error-parity condition, requiring equality

of a weighted combination of group error rates rather than point-wise equality of the rate, is generally incompatible with calibration. In their relaxed setting, the only feasible solution may involve randomly withholding predictions for some individuals, which is an undesirable consequence in practice, particularly in sensitive domains such as healthcare or criminal justice.

These results demonstrate that natural fairness criteria can be mutually incompatible and that imposing one criterion may require sacrificing another desirable property. More broadly, Foulds and Pan [FP20] discuss why parity-based criteria should not be viewed as universally appropriate measures of fairness. The suitability of parity-based criteria depends on the application, the source of observed disparities, and the particular notion of fairness one seeks to achieve.

These considerations are particularly relevant when applying parity-based criteria to matrix completion. Enforcing parity in the distribution of predictions may introduce a trade-off between parity and reconstruction accuracy. This motivates our study of whether parity-based criteria improve error-based fairness in matrix completion and under what conditions such improvements occur.

3.3 Background

3.3.1 Notation

We use boldface upper-case Latin letters to denote matrices. For example, $\mathbf{X} \in \mathbb{R}^{m \times n}$ denotes an $m \times n$ real matrix. Vectors are denoted using boldface lower-case letters, such as $\mathbf{v}$. For a matrix $\mathbf{X}$, the Frobenius norm is denoted by $\|\mathbf{X}\|_F$, and the nuclear norm is denoted by $\|\mathbf{X}\|_*$. We write $\mathbf{1}_n \in \mathbb{R}^n$ for the all-ones vector and denote the index set $[n] := 1, \ldots, n$. Throughout, $\mathbf{M}$ denotes the true matrix, whose entries are only partially observed, and $\mathbf{X}$ denotes the optimization variable in a matrix completion problem. We write $\mathbf{X}^*$ for an estimate of $\mathbf{M}$ obtained by solving such a problem.

In this work, we consider fairness in the comparison of algorithm performance across two mutually exclusive groups defined by a group-specific attribute, although the framework can be extended to comparisons among an arbitrary number of groups. We consider a data matrix $\mathbf{X}$ with m samples and n features, where each sample belongs to either group A or B. We write $\mathcal{G}_A, \mathcal{G}_B \subseteq [m]$ for the index sets of rows belonging to groups A and B, respectively, and let $m_A = |\mathcal{G}_A|$ and $m_B = |\mathcal{G}_B|$ denote respective group sizes. In most of this work, groups A and B form a disjoint partition of the rows, so that $\mathcal{G}_A \cap \mathcal{G}_B = \emptyset \text{and} \mathcal{G}_A \cup \mathcal{G}_B = [m]$, in which case $m_A + m_B = m$. Unless otherwise stated, we assume this partition. Under this assumption, we write $\mathbf{X}$ in block form as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix} \in \mathbb{R}^{m \times n}, \qquad \mathbf{X}_A \in \mathbb{R}^{m_A \times n}, \quad \mathbf{X}_B \in \mathbb{R}^{m_B \times n}.$$

3.3.2 Matrix Completion

Given a partially observed matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$, let $\Omega \subset [m] \times [n]$ be the set of indices of the observed entries in $\mathbf{M}$. The goal of matrix completion is often to recover a low-rank approximation of $\mathbf{M}$ based on available information $\{\mathbf{M}_{ij} : (i,j) \in \Omega\}$. Define the sampling operator $P_\Omega(\cdot)$ by

$$[P_\Omega(\mathbf{X})]_{ij} = \begin{cases} \mathbf{X}_{ij}, & (i,j) \in \Omega \\ 0, & \text{otherwise.} \end{cases} \tag{3.3.1}$$

The matrix completion problem may be formulated as a rank minimization problem of the form

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \text{rank}\,(\mathbf{X}) \quad \text{subject to } P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M}). \tag{3.3.2}$$

However, the problem in (3.3.2) is nonconvex and NP-hard. A popular convex relaxation of (3.3.2) is the nuclear norm minimization problem of the form

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \|\mathbf{X}\|_* \quad \text{subject to } P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M}). \tag{3.3.3}$$

Recall that the nuclear norm is defined as $\|\mathbf{X}\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{X})$, where $\sigma_i(\mathbf{X})$ are the singular values of the matrix $\mathbf{X}$. Thus, the nuclear norm can be thought of as the ℓ_1-norm of the singular values of a matrix. Since minimizing the ℓ_1-norm is well known to promote sparsity, the objective in (3.3.3) encourages solutions $\mathbf{X}^*$ with many zero singular values. As the rank of a matrix is equal to the number of its nonzero singular values, minimizing the nuclear norm therefore promotes low-rank solutions. More formally, the nuclear norm is the convex envelope of the rank function over the unit spectral-norm ball, making it the tightest convex relaxation of rank minimization on that set. Replacing the rank objective with the nuclear norm therefore transforms the original NP-hard optimization problem into a convex optimization problem with linear constraints that can be solved efficiently.

Candès and Plan [CP10] established that (3.3.3) recovers $\mathbf{M}$ exactly under appropriate rank and incoherence conditions and with sufficiently many observations. The convexity of (3.3.3) allows it to be formulated as a semidefinite program (SDP) using the standard characterization of the nuclear norm as the optimal value of a semidefinite optimization problem [Faz02]. For modestly sized problems, this SDP formulation can be solved to high accuracy using interior-point methods. For larger-scale problems, first-order methods are often preferred, including Singular Value Thresholding [CCS10], an iterative soft-thresholding scheme applied to the singular values of the current iterate, and the Soft-Impute algorithm [MHT10]. In this work, we solve the matrix completion problem using CVXPY [DB16] with the SCS solver [OCP16], a first-order operator-splitting method for conic optimization.

In practice, the observed data may contain noise. One can assume noise in the observed

entries of a matrix $\mathbf{Y}$ and model this as

$$\mathbf{Y}_{ij} = \mathbf{M}_{ij} + \mathbf{Z}_{ij}, \quad (i,j) \in \Omega,$$

where $\mathbf{Z}$ represents additive noise restricted to the observed entries. Rather than requiring an exact fit to the observations as in (3.3.3), a noise-tolerant formulation allows for a prescribed level of discrepancy:

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \|\mathbf{X}\|_* \quad \text{subject to } \|P_\Omega(\mathbf{X} - \mathbf{Y})\|_F \leq \delta. \tag{3.3.4}$$

where $\delta > 0$ bounds the magnitude of the noise on the observed entries [CP10]. Equivalently, this may be reformulated in Lagrange form as

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \frac{1}{2}\|P_\Omega(\mathbf{X} - \mathbf{Y})\|_F^2 + \lambda \|\mathbf{X}\|_*, \tag{3.3.5}$$

where $\lambda \geq 0$ is a regularization parameter. Candès and Plan [CP10] showed that nuclear norm minimization provides stable recovery when the observed entries are corrupted by sufficiently small noise, with reconstruction error controlled by the noise level. Throughout this work, we use the equality-constrained formulation (3.3.3). We therefore do not explicitly model observation noise, though the frameworks considered here can be easily adapted to accommodate noisy observations.

In collaborative filtering, matrix completion is also commonly approached through matrix factorization, in which the partially observed user-item interaction matrix is approximated by a product of low-rank factor matrices [KBV09, TA22]. Although our work focuses on nuclear norm minimization, similar fairness considerations could be investigated for matrix-factorization formulations.

3.3.3 Fairness Metrics for Matrix Completion

We now adapt statistical parity and equal opportunity to the matrix completion setting. Suppose each entry $\mathbf{X}_{ij}$ is associated with a binary group-defining attribute $Z \in \{z, z'\}$ according to the group membership of its row. That is, $Z = z$ for $i \in \mathcal{G}_A$ and $Z = z'$ for $i \in \mathcal{G}_B$. We use this attribute to define group-wise averages of the reconstructed entries.

$$\mathbb{E}[\mathbf{X}^*_{ij} \mid Z = z] = \mathbb{E}[\mathbf{X}^*_{ij} \mid Z = z']. \tag{3.3.6}$$

The corresponding empirical condition can be written for each column

$$\frac{1}{|\mathcal{G}_A|} \sum_{i \in \mathcal{G}_A} \mathbf{X}^*_{ij} = \frac{1}{|\mathcal{G}_B|} \sum_{i \in \mathcal{G}_B} \mathbf{X}^*_{ij}, \tag{3.3.7}$$

where j is fixed. Thus, statistical parity requires the reconstructed values to have the same group-wise average for each column.

Similarly, we adapt equal opportunity to the matrix completion setting by considering reconstruction error relative to the underlying matrix $\mathbf{M}$. Specifically, we require the group-wise mean reconstruction error to be equal across groups:

$$\mathbb{E}[\mathbf{X}^*_{ij} \mid Z = z] - \mathbb{E}[\mathbf{M}_{ij} \mid Z = z] = \mathbb{E}[\mathbf{X}^*_{ij} \mid Z = z'] - \mathbb{E}[\mathbf{M}_{ij} \mid Z = z']. \tag{3.3.8}$$

Since $\mathbf{M}$ is only partially observed, its group-wise means must be estimated from the observed entries. Define the set of observed row indices in column j as

$$\Omega_j = \{i \in [m] : (i, j) \in \Omega\},$$

and define the set of columns for which at least one entry is observed in both groups as

$$\mathcal{J} = \{j \in [n] : |\mathcal{G}_A \cap \Omega_j| > 0,\ |\mathcal{G}_B \cap \Omega_j| > 0\}.$$

For each $j \in \mathcal{J}$, the empirical equal opportunity condition is

$$\left(\frac{1}{|\mathcal{G}_A|} \sum_{i \in \mathcal{G}_A} \mathbf{X}^*_{ij} - \frac{1}{|\mathcal{G}_A \cap \Omega_j|} \sum_{i \in \mathcal{G}_A \cap \Omega_j} \mathbf{M}_{ij} \right) = \left(\frac{1}{|\mathcal{G}_B|} \sum_{i \in \mathcal{G}_B} \mathbf{X}^*_{ij} - \frac{1}{|\mathcal{G}_B \cap \Omega_j|} \sum_{i \in \mathcal{G}_B \cap \Omega_j} \mathbf{M}_{ij} \right). \quad (3.3.9)$$

Unlike statistical parity, equal opportunity allows systematic differences in average values between groups while seeking to prevent the reconstruction procedure from introducing systematic differences in mean error. Neither criterion directly imposes any constraint on the magnitude of reconstruction error across groups. We therefore investigate whether enforcing equal opportunity can also improve error-based fairness in matrix completion.

In the fairness literature, statistical parity and equal opportunity are also commonly referred to as *fairness metrics*. In the matrix completion setting, however, satisfying these metrics does not necessarily correspond to our stated measure of fairness: comparable reconstruction accuracy across groups. In Section 3.4, we investigate two approaches for incorporating statistical parity and equal opportunity into matrix completion: fairness-metric regularization, which incorporates these criteria directly into the matrix completion objective, and a group-aware objective that allows each group's low-rank structure to be regularized separately. We examine whether, and under what conditions, enforcing these parity-based criteria affects reconstruction accuracy across demographic groups.

3.4 Methods

Fairness interventions in machine learning are broadly categorized as pre-processing, in-processing, or post-processing methods, depending on whether fairness is addressed through the input data, incorporated directly into the learning objective, or applied to the model's outputs after training, respectively [CH24]. The approaches considered in this work are all in-processing methods, in that they incorporate group information or fairness metrics directly into the matrix completion optimization problem.

We consider two approaches for addressing disparities in reconstruction across demo-

graphic groups. First, we incorporate two parity-based fairness criteria, statistical parity and equal opportunity, directly into the objective through *fairness metric regularization*. Second, *Group-Aware Matrix Completion* incorporates group-specific structure into the matrix completion objective, allowing the reconstruction to account for differences in the underlying structure of each group.

The resulting methods allow us to examine different ways of addressing group-level disparities in matrix completion, and in particular to compare group-aware modeling with approaches that explicitly enforce parity-based fairness criteria.

3.4.1 Fairness Metric Regularization

Section 3.3.3 defines statistical parity (Eq. (3.3.7)) and equal opportunity (Eq. (3.3.9)) in terms of the reconstruction $\mathbf{X}^*$ and the underlying matrix $\mathbf{M}$ for the two-group setting. We now incorporate each criterion as a regularization penalty $\mathcal{F}(\mathbf{X})$ added to the nuclear norm minimization objective:

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \|\mathbf{X}\|_* + \lambda\,\mathcal{F}(\mathbf{X}) \qquad \text{subject to} \qquad P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M}). \tag{3.4.1}$$

For statistical parity and equal opportunity, we denote the corresponding fairness penalties by $\mathcal{F}_{\text{SP}}$ and $\mathcal{F}_{\text{EO}}$, with regularization parameters λ_{SP} and λ_{EO}, respectively. These parameters control the trade-off between minimizing the nuclear norm and encouraging the corresponding fairness criterion.

Statistical Parity

To encourage statistical parity in the completed matrix, we seek to make the average reconstructed values of the two demographic groups similar across the matrix. We aggregate

the per-column discrepancy and penalize deviations from parity. Specifically, we define

$$\mathcal{F}_{\mathrm{SP}}(\mathbf{X}) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\sum_{i\in\mathcal{G}_A}\mathbf{X}_{ij}}{|\mathcal{G}_A|} - \frac{\sum_{i\in\mathcal{G}_B}\mathbf{X}_{ij}}{|\mathcal{G}_B|}\right)^2. \tag{3.4.2}$$

Equal Opportunity

To encourage equal opportunity, we penalize differences in the mean prediction error between groups. Because the entries of $\mathbf{M}$ are only partially observed, we estimate the group-wise means of $\mathbf{M}$ using the observed entries. Using the sets Ω_j and $\mathcal{J}$ defined in Section 3.3.3, we define

$$\mathcal{F}_{\mathrm{EO}}(\mathbf{X}) = \frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}}\left[\left(\frac{\sum_{i\in\mathcal{G}_A}\mathbf{X}_{ij}}{|\mathcal{G}_A|} - \frac{\sum_{i\in\mathcal{G}_A\cap\Omega_j}\mathbf{M}_{ij}}{|\mathcal{G}_A\cap\Omega_j|}\right) - \left(\frac{\sum_{i\in\mathcal{G}_B}\mathbf{X}_{ij}}{|\mathcal{G}_B|} - \frac{\sum_{i\in\mathcal{G}_B\cap\Omega_j}\mathbf{M}_{ij}}{|\mathcal{G}_B\cap\Omega_j|}\right)\right]^2. \tag{3.4.3}$$

Thus, $\mathcal{F}_{\mathrm{EO}}$ penalizes differences between the reconstructed-minus-observed group means across the columns for which both groups have observed entries.

Both $\mathcal{F}_{\mathrm{SP}}$ and $\mathcal{F}_{\mathrm{EO}}$ are convex in $\mathbf{X}$. Consequently, adding either penalty to the objective with a nonnegative weight preserves the convexity of the optimization problem.

We penalize the per-column group gap using an ℓ_2 (squared) penalty rather than an ℓ_1 (absolute-value) penalty for both criteria. An ℓ_1 penalty encourages sparsity and can drive most column-wise gaps to zero while allowing a relatively large gap to remain on a small number of columns. An ℓ_2 penalty instead penalizes large gaps quadratically, discouraging any single column from retaining a large disparity. We therefore use the ℓ_2 penalty throughout as a more conservative choice that discourages solutions in which a small number of columns remain highly disparate despite small average differences across the matrix.

3.4.2 Group-Aware Matrix Completion

Standard nuclear norm minimization assumes that a single low-rank model adequately represents all groups within the data. In practice, however, different demographic groups may exhibit distinct latent structure or varying levels of complexity. When one group is

substantially underrepresented, the shared low-rank model may place greater emphasis on the structure of the majority group, potentially resulting in higher reconstruction error for the minority group.

The *Group-Aware Matrix Estimation* (GAME) framework of [GSA26] extends nuclear norm minimization to matrix estimation with multiple and possibly overlapping groups. GAME applies group-specific nuclear norm penalties to the corresponding group submatrices. In the general GAME formulation, a row may belong to multiple groups, and the overlap among groups allows information to be shared across the corresponding submatrices.

We consider a specialized form of the GAME objective for two mutually exclusive groups. Specifically, we take the set of categories to be $\mathcal{G} = \{A, B, [m]\}$, where $[m]$ denotes the group containing all rows. We enforce agreement with the observed entries as an equality constraint and seek to minimize

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \lambda_A\|\mathbf{X}_A\|_* + \lambda_B\|\mathbf{X}_B\|_* + \lambda_{[m]}\|\mathbf{X}\|_* \quad \text{subject to} \quad P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M}). \tag{3.4.4}$$

Recall that $\mathbf{X}_A$ and $\mathbf{X}_B$ denote the submatrices corresponding to groups A and B, respectively. The group-specific nuclear norms allow the low-rank structure within each group to be regularized separately, while the all-row nuclear norm encourages a low-rank representation shared across the full matrix.

The relative weights determine the balance between group-specific and shared low-rank structure. In the limiting case $\lambda_{[m]} \to \infty$, the all-row nuclear norm dominates, recovering the standard nuclear norm minimization formulation. In contrast, when $\lambda_{[m]} \to 0$, the objective separates into independent nuclear norm minimization problems for the two groups. Intermediate values allow the estimator to balance group-specific structure with structure shared across the groups.

The theoretical analysis of GAME motivates category-specific weights whose dominant dependence is on the number of observations in each category. In particular, Theorem 3.5

of [GSA26] gives a weight scaling of the form

$$\lambda_g \gtrsim \frac{\sigma}{\kappa_{\min}} \sqrt{\frac{N m_g}{m} \frac{\log(m_g + n)}{\min(m_g, n)}} + \frac{R \log(m_g + n)}{\kappa_{\min}}, \tag{3.4.5}$$

where m_g is the number of rows in category g, N is the total number of observed entries $|\Omega|$, σ and R are parameters associated with the subexponential noise model, and $\kappa_{\min}$ denotes the minimum row-overlap multiplicity. Under uniform sampling, the expected number of observations in category g is

$$N_g = \frac{N m_g}{m},$$

Thus, the first term in (3.4.5) has a $\sqrt{N_g}$ dependence on the number of observations in category g. Following this dependence, we use the practical weighting heuristic

$$\lambda_g = \sqrt{N_g}, \qquad g \in \{A, B, [m]\}, \tag{3.4.6}$$

where N_g denotes the number of observed training entries in category g. In our experiments, this quantity is computed directly from the observed training entries.

3.5 Numerical Experiments

3.5.1 Experimental Setup

We analyze fairness in reconstruction accuracy across groups for the two methods introduced in Section 3.4: group-aware matrix completion and fairness metric regularization. We conduct experiments on synthetic low-rank data and on real data from the MovieLens 100K dataset, a widely used benchmark for recommender systems. We continue to consider a binary partition of the data into two mutually exclusive groups, labeled A and B, as described in Section 3.3.1. We designate group A as the group of interest, due to population imbalance and group-specific structure that differs from group B's. In our experiments, we

vary the size of group A, denoted by m_A, as a fraction of the total number of rows m, with $m_A \leq m_B$. When $m_A < m_B$, we refer to A as the minority group. The methods and analysis can be extended naturally to more than two groups and to non-exclusive group assignments.

Following [YH17], we consider population imbalance as one potential source of disparities in reconstruction accuracy by varying the proportion of the focal group in the dataset. We also consider settings in which the groups differ in the complexity of their underlying structure, allowing us to distinguish disparities associated with group representation from those arising from data complexity

We evaluate reconstruction accuracy using the root mean squared error (RMSE), computed separately for each group on the corresponding held-out entries. Let $\Omega_{\text{test}} \subseteq \Omega$ denote the held-out subset of observed entries used for evaluation. For each group $g \in \{A, B\}$,

$$\Omega_{g,\text{test}} = \{(i,j) \in \Omega_{\text{test}} : i \in \mathcal{G}_g\}.$$

Definition 3.5.1 (Group-wise RMSE)**.** *The group-wise RMSE of an estimate* $\mathbf{X}^*$ *against the true matrix* $\mathbf{M}$ *is*

$$\text{RMSE}_g = \sqrt{\frac{1}{|\Omega_{g,\text{test}}|} \sum_{(i,j)\in\Omega_{g,\text{test}}} (\mathbf{X}^*_{ij} - \mathbf{M}_{ij})^2}, \qquad g \in \{A, B\}.$$

We use RMSE rather than relative Frobenius norm error because relative error normalizes by the magnitude of the true entries. In our experiments, groups have entries on the same scale but may differ in their average values. This normalization can therefore make groups with smaller mean values appear to have larger relative error even when their absolute reconstruction quality is comparable. Conversely, when group means are similar, relative error can obscure differences in absolute reconstruction quality. RMSE avoids these effects by measuring reconstruction error in the original units of the data. It also normalizes by the number of held-out entries rather than the magnitude of the corresponding group submatrix,

making it suitable for comparing groups of different sizes.

We control the difficulty of each completion problem by varying the fraction of entries used for training. We denote this fraction by the *observation rate* p and define the corresponding *masking rate* as $q := 1 - p$. Thus, for synthetic data, the masking rate is defined with respect to all entries of the underlying matrix, while for MovieLens it is defined with respect to the naturally observed ratings. In both cases, the masking rate determines the fraction of available entries withheld from training and used for evaluation. We use uniform random masking on the entire matrix, applying the same observation rate to both groups, so that the masking mechanism itself does not introduce observation bias between groups.

We vary the masking rate from 0.1 to 0.8 across experiments. For each masking rate, we perform 20 independent trials and report the mean group-wise RMSE across trials. Synthetic experiments use 100×100 matrices, while MovieLens experiments use 300×300 matrices.

For experiments with vanilla nuclear norm minimization, we vary the representation of the focal group A to study the effect of population imbalance on reconstruction accuracy. We consider minority-group fractions ranging from 10% to 50%.

For method-comparison experiments, the minority-group fraction is fixed at 20% of the rows in each dataset. The training and test masks are drawn uniformly at random at the beginning of each trial and are shared across all methods. For synthetic experiments, the factor matrices used to generate each matrix are also drawn independently at the beginning of each trial. For real-data experiments, the MovieLens matrix is fixed and only the train/test mask varies across trials.

3.5.2 Synthetic Datasets

For the synthetic experiments, we partition the rows of the data matrix into two groups and generate each group from its own low-rank factorization. As illustrated in Figure 3.1,

we write

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_A \\ \mathbf{M}_B \end{bmatrix} = \begin{bmatrix} \mathbf{U}_A \mathbf{V}_A \\ \mathbf{U}_B \mathbf{V}_B \end{bmatrix},$$

where, for each group $g \in \{A, B\}$, the factor matrices $\mathbf{U}_g \in \mathbb{R}^{m_g \times r_g}$ and $\mathbf{V}_g \in \mathbb{R}^{r_g \times n}$ represent the sample and feature factors, respectively, with entries sampled independently from $\mathcal{N}(0, 1)$. The rank of the full matrix r satisfies $r := \text{rank}\,(\mathbf{M}) \leq r_A + r_B$ with equality when the row spaces of $\mathbf{M}_A$ and $\mathbf{M}_B$ are linearly independent. Unless otherwise stated, all synthetic experiments use 100×100 matrices with $r_A = r_B = 4$.

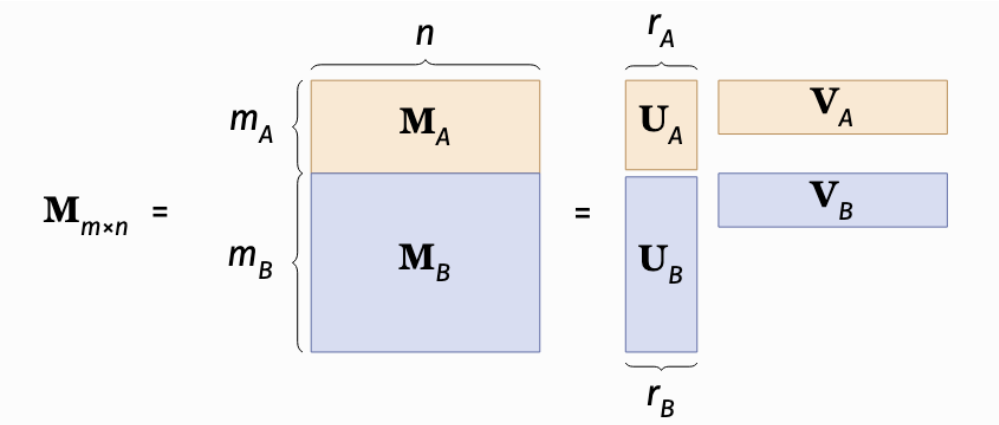


Figure 3.1: Illustration of synthetic data construction for the two group setting. The matrix $\mathbf{M}$ is partitioned into submatrices $\mathbf{M}_A$ and $\mathbf{M}_B$, each generated from its own factor matrices $\mathbf{U}_g$ and $\mathbf{V}_g$. The group-specific ranks are r_A and r_B, while the full matrix has rank at most $r_A + r_B$.

The synthetic datasets are designed to form a progression from shared to increasingly distinct latent structure between groups. Table 3.1 summarizes the key properties of each data model. The *Shared V* dataset represents groups with identical item factor matrices, so both groups share the same latent feature structure. The *Perturbed V* and *Extended Minority V* datasets introduce increasing structural differences through perturbations and additional latent dimensions, respectively. The *Approximately Orthogonal* and *Orthogonal Groups* datasets represent increasingly distinct group-specific latent structures, with the latter enforcing exact orthogonality between the groups.

The *Shifted Means* dataset serves a different purpose. The two groups share the same latent structure but have different mean shifts, providing a setting in which statistical parity can conflict with the underlying group means without introducing substantial differences in low-rank structure. This dataset therefore provides a useful setting for examining the effects of parity-based fairness criteria.

These data models allow us to empirically examine two potential sources of disparities in reconstruction accuracy: *population imbalance*, in which a group's smaller representation may result in less influence on the global solution, and *group-specific structure*, in which differences in latent structure between groups can lead to disparities even when groups are equally represented. We vary the minority-group fraction of the data in experiments with vanilla nuclear norm minimization to assess how the effect of group size on reconstruction disparities varies across data models. We then evaluate the two fairness interventions across datasets.

Table 3.1: Synthetic data models used for experiments.The datasets form a progression from maximally shared to completely independent latent structure between groups. The ranks reported correspond to the setting $r_A = r_B$

Dataset	**Construction**	$\text{rank}\,(\mathbf{M})$
Shared V	$\mathbf{M}_g = \mathbf{U}_g\mathbf{V}$ (shared item factor)	r_A
Shifted Means	$\mathbf{UV}$ with group-wise mean shifts	$r_A + 1$
Perturbed V	$\mathbf{V}_A = \mathbf{V} + \mathbf{E}$, $\mathbf{E}_{ij} \sim \mathcal{N}(0,1)$	$r_A + r_B$
Extended Minority V	$\mathbf{V}_A$ augmented with k additional dimensions	$r_A + k$
Approximately Orthogonal	$\mathbf{V}_A, \mathbf{V}_B$ drawn independently	$r_A + r_B$
Orthogonal Groups	$\text{row}(\mathbf{V}_B) \subseteq \text{row}(\mathbf{M}_A)^\perp$	$r_A + r_B$

3.5.2.1 Shared V

The Shared V dataset uses a common item factor matrix for both groups. Starting from the general construction, we write

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_A \\ \mathbf{M}_B \end{bmatrix} = \begin{bmatrix} \mathbf{U}_A\mathbf{V} \\ \mathbf{U}_B\mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_A \\ \mathbf{U}_B \end{bmatrix} \mathbf{V} = \mathbf{U}\mathbf{V}.$$

The group-specific ranks are set to be equal, $r_A = r_B$, thus the rank of the full matrix satisfies $r \leq r_A$, since the row spaces of both $\mathbf{M}_A$ and $\mathbf{M}_B$ are contained in the row space of the common matrix $\mathbf{V}$. Under the random Gaussian construction used here, $\mathbf{V}$ has rank r_A with probability one, and thus $r = r_A = 4$. This dataset represents a simplistic setting in which the two groups do not have distinct latent structure: they are simply partitions of a common low-rank matrix generated from the shared factor matrix $\mathbf{V}$.

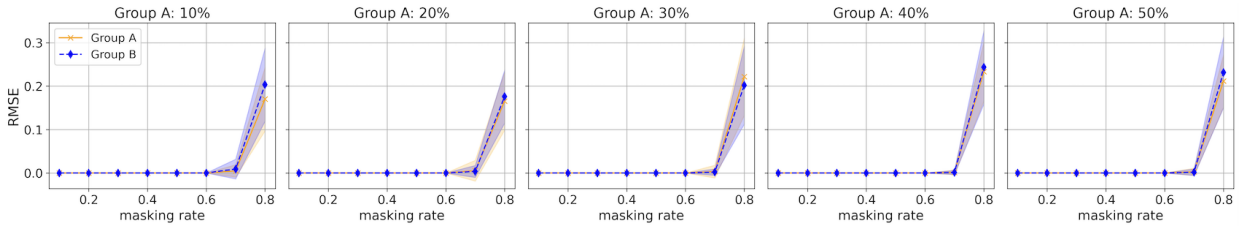


Figure 3.2: Reconstruction accuracy of vanilla nuclear norm minimization on the Shared V dataset as a function of masking rate for varying minority group size. The group-wise RMSE for groups A and B remains nearly identical across minority-group fractions, indicating negligible disparity in reconstruction accuracy. The underlying matrix has rank $r = 4$.

Figure 3.2 shows negligible error disparity across all minority fractions. Since both groups share the same item factor matrix $\mathbf{V}$, the data have no group-specific latent structure. The global low-rank model therefore captures the underlying structure of both groups without a meaningful disparity in reconstruction accuracy.

3.5.2.2 Shifted Means

The Shifted Means dataset uses the same shared latent structure as the Shared V dataset but introduces different group-wise mean shifts. We set the mean shifts to be $\mu_A = 1.0$, $\mu_B = 5.0$. The data can be formulated as

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_A \\ \mathbf{M}_B \end{bmatrix} = \begin{bmatrix} \mathbf{U}_A\mathbf{V} + \mu_A \mathbf{1}_{m_A}\mathbf{1}_n^\top \\ \mathbf{U}_B\mathbf{V} + \mu_B \mathbf{1}_{m_B}\mathbf{1}_n^\top \end{bmatrix} = \mathbf{U}\mathbf{V} + \begin{bmatrix} \mu_A \mathbf{1}_{m_A}\mathbf{1}_n^\top \\ \mu_B \mathbf{1}_{m_B}\mathbf{1}_n^\top \end{bmatrix},$$

where $\mathbf{1}_{m_g} \in \mathbb{R}^{m_g}$ and $\mathbf{1}_n \in \mathbb{R}^n$ denote all-ones vectors of the appropriate dimensions. The group-wise mean-shift term has rank one, so $r \leq r_A + 1$. With $r_A = r_B = 4$ and the randomized factor construction used here, the full matrix has rank $r = 5$ with probability one.

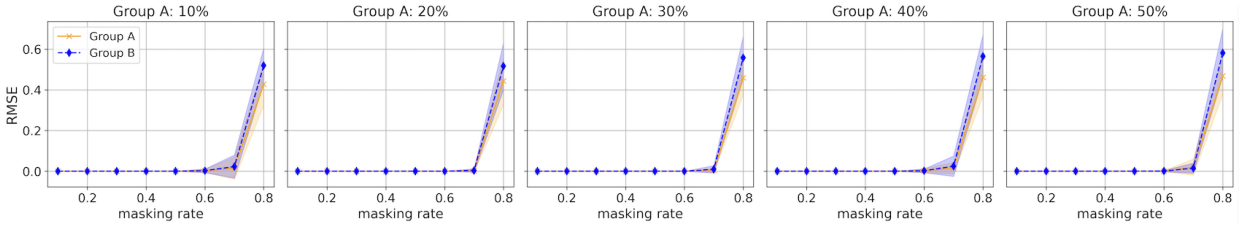


Figure 3.3: Reconstruction accuracy of vanilla nuclear norm minimization on the Shifted Means dataset as a function of masking rate for varying minority group size. The group-wise RMSE for groups A and B remains similar except at high masking rates, with little dependence on minority-group fraction. The underlying matrix has rank $r = 5$.

The mean shifts introduce a systematic difference in the true group means without introducing group-specific latent feature structure. As shown in Figure 3.3, vanilla nuclear norm minimization exhibits reconstruction error disparity only at very high masking rates, with little dependence on minority-group size. This dataset provides a setting for studying parity-based criteria when the groups have genuinely different mean values.

3.5.2.3 Perturbed V

In the Perturbed V dataset, we perturb group A's item factor matrix by additive Gaussian noise. Specifically, let

$$\mathbf{V}_A = \mathbf{V} + \mathbf{E},$$

where $\mathbf{E} \in \mathbb{R}^{r_A \times n}$ has entries sampled independently from $\mathcal{N}(0, 1)$, while group B retains the unperturbed item factor matrix $\mathbf{V}$. The resulting matrix is

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_A \\ \mathbf{M}_B \end{bmatrix} = \begin{bmatrix} \mathbf{U}_A(\mathbf{V} + \mathbf{E}) \\ \mathbf{U}_B\mathbf{V} \end{bmatrix}.$$

We set $r_A = r_B = 4$. Since the two groups generally have distinct item-factor subspaces, the full matrix has rank $r = r_A + r_B = 8$ with probability one. This construction introduces group-specific latent structure while retaining some shared structure through factor matrix $\mathbf{V}$.

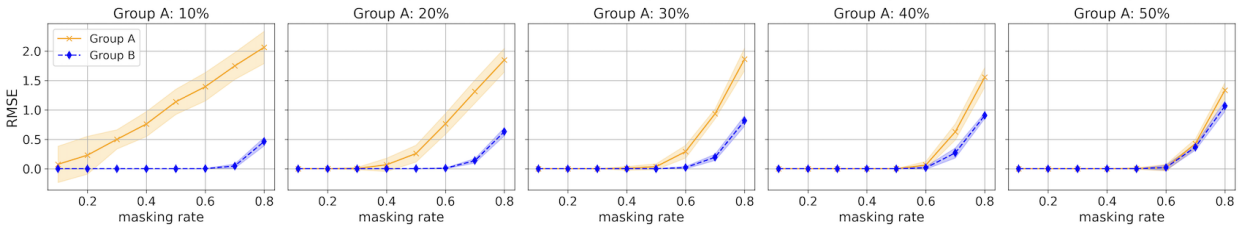


Figure 3.4: Reconstruction accuracy of vanilla nuclear norm minimization on the Perturbed V dataset as a function of masking rate for varying minority group size. The underlying matrix has rank $r = 8$, with $r_A = r_B = 4$.

Figure 3.4 shows a clear disparity in reconstruction accuracy between the two groups. The disparity is most pronounced when group A is a small minority within the data and decreases as the groups approach equal size, suggesting that population imbalance contributes to the disparity. Notably, the gap persists when $m_A = m_B$, indicating that differences in group-specific structure contribute to the disparity in addition to population imbalance.

3.5.2.4 Extended Minority V

In the Extended Minority V dataset, group A has additional latent structure beyond the subspace shared by the two groups. Specifically, the item factors for group A are augmented with k additional dimensions. Define

$$\tilde{\mathbf{V}} = \begin{bmatrix} \mathbf{V} \\ \mathbf{V}_{\text{extra}} \end{bmatrix} \in \mathbb{R}^{(r_A+k)\times n},$$

with $\mathbf{U}_A \in \mathbb{R}^{m_A \times (r_A+k)}$. Then

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_A \\ \mathbf{M}_B \end{bmatrix} = \begin{bmatrix} \mathbf{U}_A \tilde{\mathbf{V}} \\ \mathbf{U}_B \mathbf{V} \end{bmatrix}.$$

We set $r_A = r_B$, so $r \le r_A + k$. With $r_A = r_B = 4$ and $k = 2$, the construction has $r = 6$ with probability one.

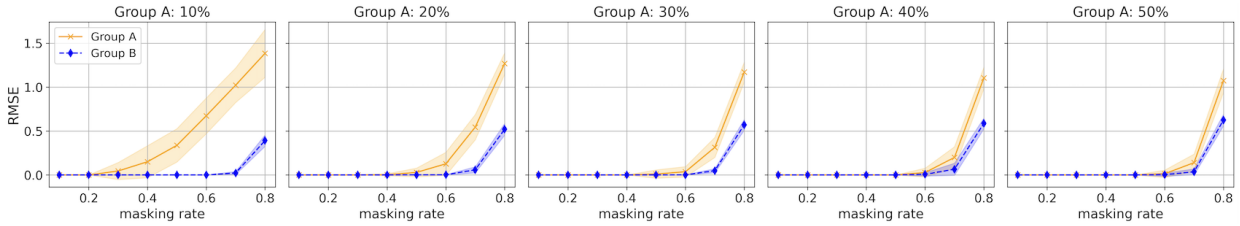


Figure 3.5: Reconstruction accuracy of vanilla nuclear norm minimization on the Extended Minority V dataset as a function of masking rate for varying minority group size. The underlying matrix has rank $r = 6$, with $r_A = r_B = 4$ and $k = 2$.

Figure 3.5 shows a persistent disparity in reconstruction accuracy between the two groups. Group A has more complex latent structure than group B and exhibits higher reconstruction error across the minority-group fractions considered. The error gap therefore reflects the combined effects of population imbalance and differences in group-specific latent structure. Notably, the disparity persists when $m_A = m_B$, showing that group-specific structure alone can produce unequal reconstruction accuracy even when the groups are equally represented.

3.5.2.5 Approximately Orthogonal

In the Approximately Orthogonal dataset, each group has its own independently sampled user and item factor matrices:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_A \\ \mathbf{M}_B \end{bmatrix} = \begin{bmatrix} \mathbf{U}_A \mathbf{V}_A \\ \mathbf{U}_B \mathbf{V}_B \end{bmatrix}.$$

Since $\mathbf{V}_A$ and $\mathbf{V}_B$ are sampled independently from $\mathcal{N}(0, 1)$, their row spaces are approximately orthogonal when $r_g \ll n$. Thus, with $r_A = r_B = 4$, the full matrix has rank $r = r_A + r_B = 8$ with probability one. This dataset represents groups with distinct latent item-factor structure and relatively little overlap between their row spaces.

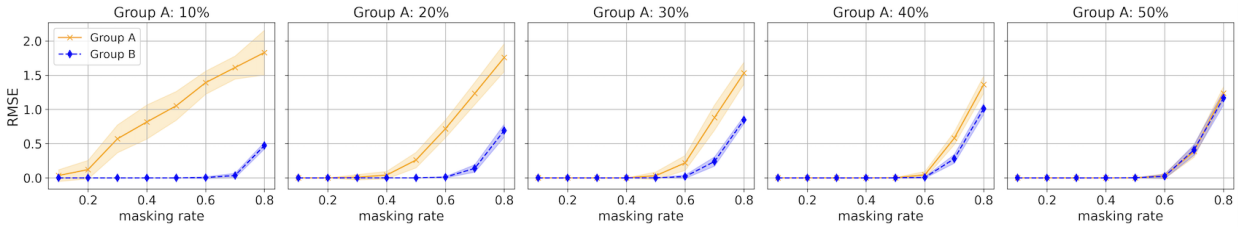


Figure 3.6: Reconstruction accuracy of vanilla nuclear norm minimization on the Approximately Orthogonal dataset as a function of masking rate for varying minority group size. The underlying matrix has rank $r = 8$, with $r_A = r_B = 4$.

Figure 3.6 shows a disparity in reconstruction accuracy that is most pronounced when group A is a small minority and decreases as the groups approach equal size. When $m_A = m_B$, there is no meaningful disparity gap, indicating that neither group disadvantaged by its representation in the global low-rank model when the groups are equally represented. This suggests that population imbalance contributes substantially to the disparity.

3.5.2.6 Orthogonal Groups

The Orthogonal Groups dataset enforces orthogonality between the latent structures of the two groups. Specifically, the item factors for group B are constructed to lie in the

orthogonal complement of the row space of $\mathbf{M}_A$:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_A \\ \mathbf{M}_B \end{bmatrix} = \begin{bmatrix} \mathbf{U}_A \mathbf{V}_A \\ \mathbf{U}_B \mathbf{V}_B \end{bmatrix}, \qquad \text{row}(\mathbf{V}_B) \subseteq \text{row}(\mathbf{M}_A)^{\perp}.$$

The matrix $\mathbf{V}_B$ is constructed by selecting r_B basis vectors from the orthogonal complement of row$(\mathbf{M}_A)$. Consequently, the row spaces of $\mathbf{M}_A$ and $\mathbf{M}_B$ are orthogonal and $\mathbf{M}_A \mathbf{M}_B^\top = \mathbf{0}$. With $r_A = r_B = 4$, the full matrix has rank $r = r_A + r_B = 8$.

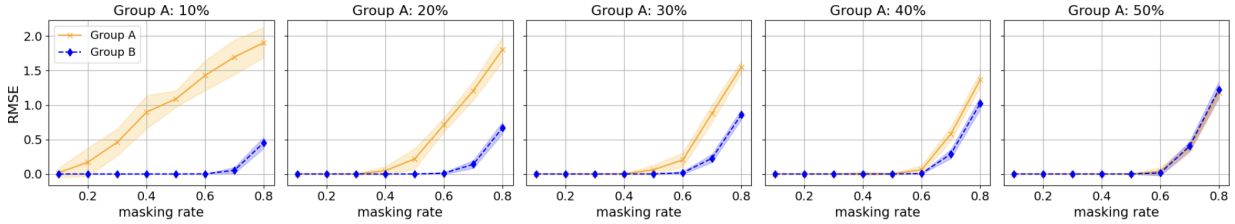


Figure 3.7: Reconstruction accuracy of vanilla nuclear norm minimization on the Orthogonal Groups dataset as a function of masking rate for varying minority group size. The underlying matrix has rank $r = 8$, with $r_A = r_B = 4$.

Figure 3.7 shows a similar dependence on group size, with larger disparities when group A is a smaller fraction of the data, suggesting that population imbalance contributes substantially to the disparity. When $m_A = m_B$, there is no disparity gap, indicating that neither group is strongly disadvantaged by its representation in the global low-rank model when the groups are equally represented.

3.5.3 MovieLens 100K

The MovieLens 100K dataset [HK15] is a widely used benchmark for collaborative filtering containing $100,000$ ratings from 943 users on $1,682$ movies, with ratings on a discrete 1–5 scale. We select a subset of 300 users and 300 items and partitioning users by gender. We designate female users as group A and male users as group B. To study the effect of group size on disparities in reconstruction accuracy, we vary the female fraction from 10%

to 50% of the total user population, as in the synthetic experiments. For each fraction, a new subset of users is selected while maintaining a total of 300 users.

The rating matrix is naturally sparse, with missing rates ranging from 86%–88% for group A and 83%–84% for group B across the different group proportions. Average ratings are also similar across groups, with female users having average ratings ranging from 3.65–3.71 and male users from 3.68–3.69. Thus, the groups have similar rating scales and comparable proportions of observed ratings, although the underlying latent structure of the data is unknown.

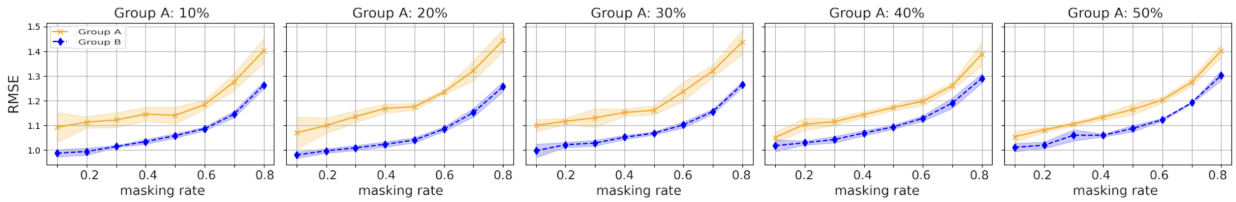


Figure 3.8: Reconstruction accuracy of vanilla nuclear norm minimization on the MovieLens 100K dataset as a function of masking rate for varying female-user fractions. Group A denotes female users and group B denotes male users.

Figure 3.8 shows that group A has higher reconstruction error than group B across the range of group proportions considered. This disparity persists even when female users constitute a majority of the selected users; the error gap does not disappear until the female fraction exceeds approximately 70%. This suggests that, in addition to population imbalance, differences in group-specific rating patterns may contribute to the disparity.

3.5.4 Fairness Metric Regularization Experiments

We evaluate the effects of statistical parity and equal opportunity regularization on reconstruction accuracy across the synthetic and MovieLens datasets. For both criteria, we use the regularized nuclear norm minimization formulation

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \|\mathbf{X}\|_* + \lambda\,\mathcal{F}(\mathbf{X}) \qquad \text{subject to} \qquad P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{M}), \tag{3.5.1}$$

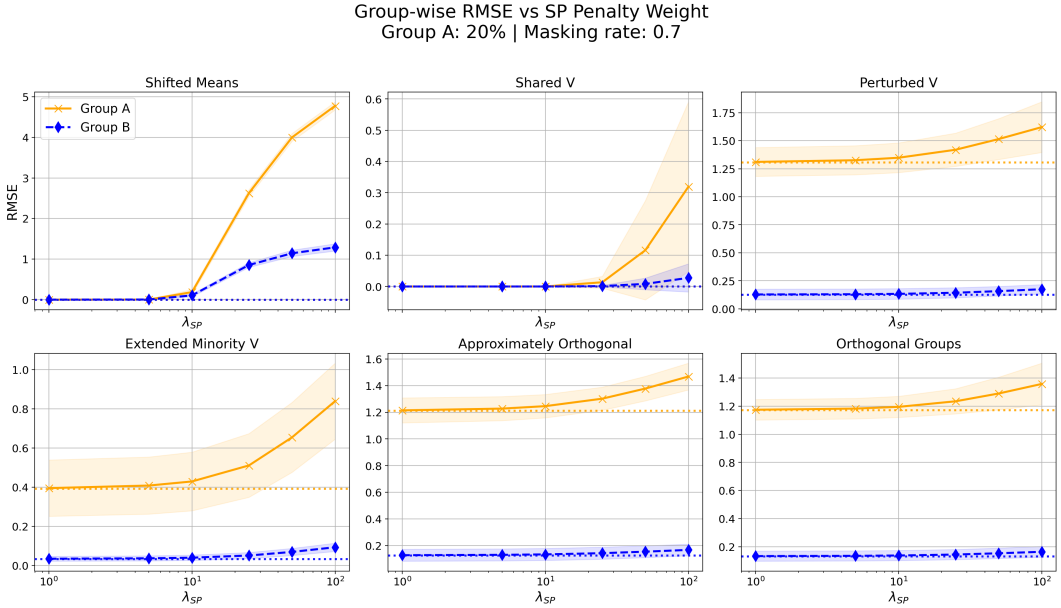
where $\mathcal{F}(\mathbf{X})$ denotes either the statistical parity or equal opportunity penalty defined in Section 3.4.1. Synthetic experiments use 100×100 matrices, while MovieLens experiments use 300×300 matrices, with the minority group comprising 20% of the rows, consistent with the experimental setup described above.

3.5.4.1 Synthetic Data Experiments

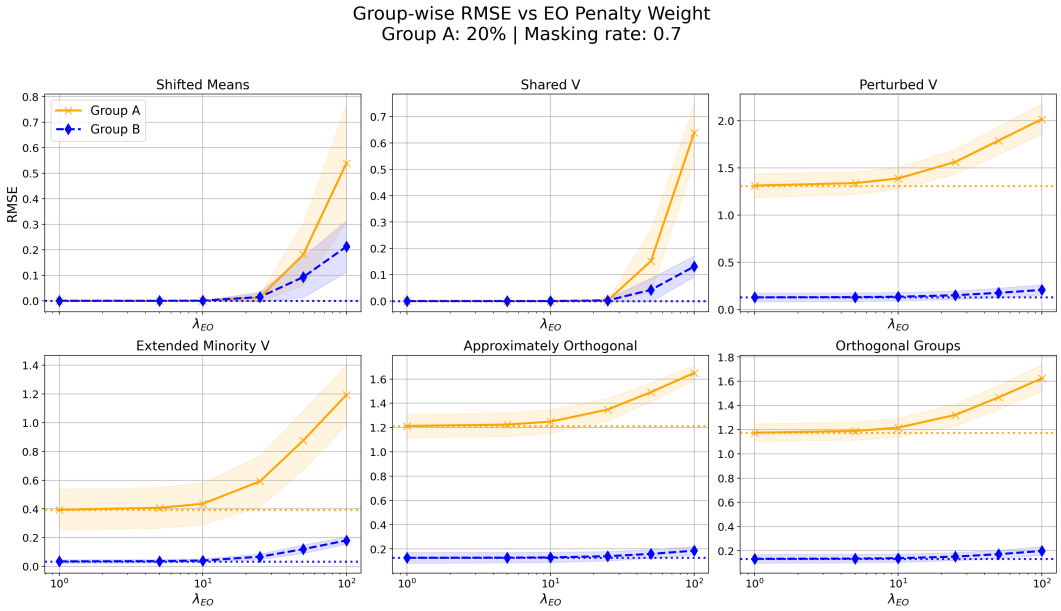
We first examine the effect of the fairness-penalty weight. For synthetic datasets, we consider a range of penalty weights for both statistical parity and equal opportunity, at a fixed masking rate of 0.7, as shown in Figure 3.9. As the penalty weight increases, the reconstruction error of both groups monotonically increases relative to the vanilla nuclear norm minimization baseline. The increase in error is disproportionately larger for group A, resulting in an increased disparity in reconstruction error between the groups. Since the penalty sweeps do not identify a weight that improves reconstruction accuracy, we select $\lambda_{\text{SP}} = 100$ and $\lambda_{\text{EO}} = 100$ for the method-comparison experiments, unless otherwise stated. This choice provides a common penalty weight at which the effects of the fairness penalties are clearly visible and facilitates comparison across datasets.

We compare vanilla nuclear norm minimization with statistical parity and equal opportunity regularization across the six synthetic datasets. The results are reported in Figures 3.10–3.12, with each figure showing the group-wise RMSE as a function of masking rate for a fixed minority-group fraction of 20%.

We first consider the Shared V and Shifted Means datasets, for which vanilla nuclear norm minimization exhibits little or no reconstruction error disparity. Both fairness penalties increase reconstruction error for both groups in these settings. Under statistical parity, the increase is larger for the minority group, introducing a disparity where little or none existed under vanilla nuclear norm minimization. On Shared V, this occurs despite the groups sharing the same underlying latent structure, indicating that the fairness penalties can reduce reconstruction accuracy even when the unregularized solution already provides



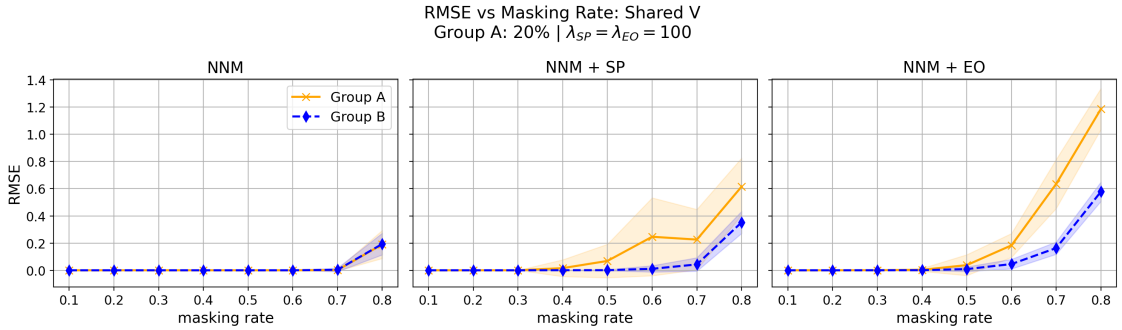
(a) Statistical parity regularization.



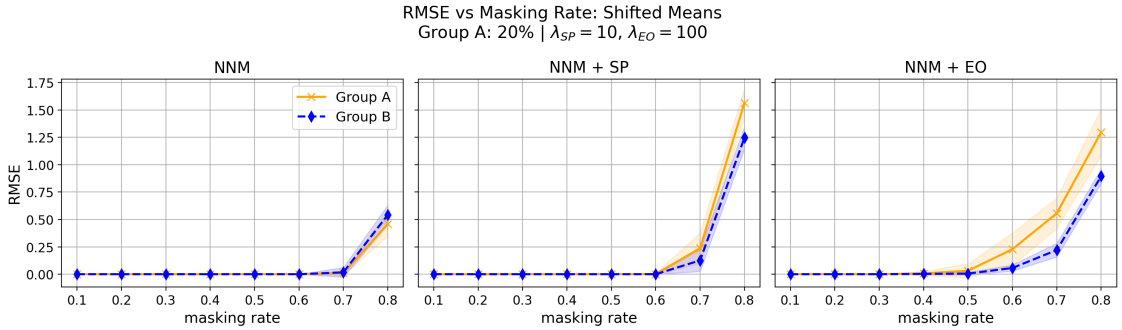
(b) Equal opportunity regularization.

Figure 3.9: Effect of the fairness-penalty weight on synthetic datasets for statistical parity (top) and equal opportunity regularization (bottom). The plots show group-wise RMSE as a function of the penalty weight at a fixed masking rate of 0.7.

comparable accuracy across groups.



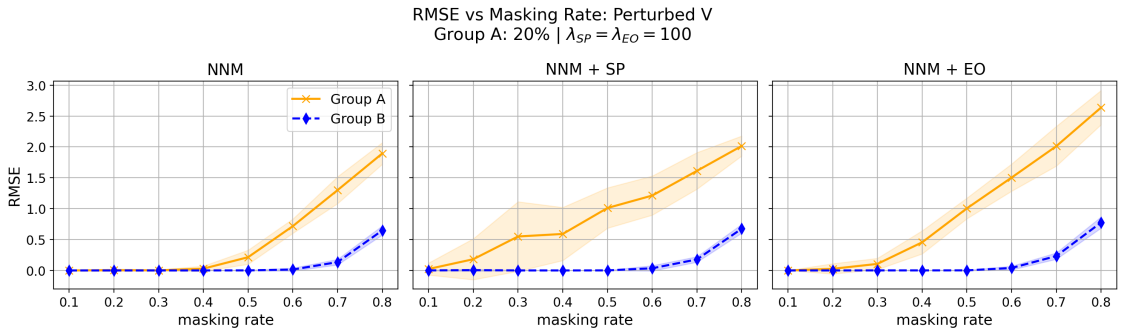
(a) Shared V.



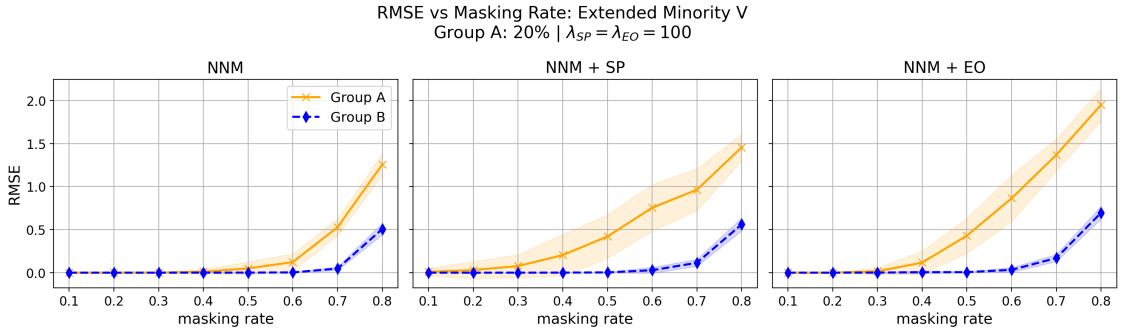
(b) Shifted Means.

Figure 3.10: Group-wise RMSE as a function of masking rate. Methods are presented from left to right as vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. The minority-group fraction is fixed at 20%. For the Shifted Means dataset, $\lambda_{\mathrm{SP}} = 10$ is used because $\lambda_{\mathrm{SP}} = 100$ produced an unstable solution. For all other synthetic datasets, $\lambda_{\mathrm{SP}} = \lambda_{\mathrm{EO}} = 100$.

We next consider the Perturbed V and Extended Minority V datasets, both of which exhibit reconstruction error disparities under vanilla nuclear norm minimization. In the Perturbed V setting, both fairness penalties increase the minority-group RMSE while leaving the majority-group RMSE approximately unchanged, so the existing disparity is not reduced. A similar pattern occurs on Extended Minority V, where both penalties increase the minority-group RMSE. In these settings, enforcing parity in group-level averages therefore does not improve error-based fairness and instead comes at the expense of minority-group reconstruction accuracy.



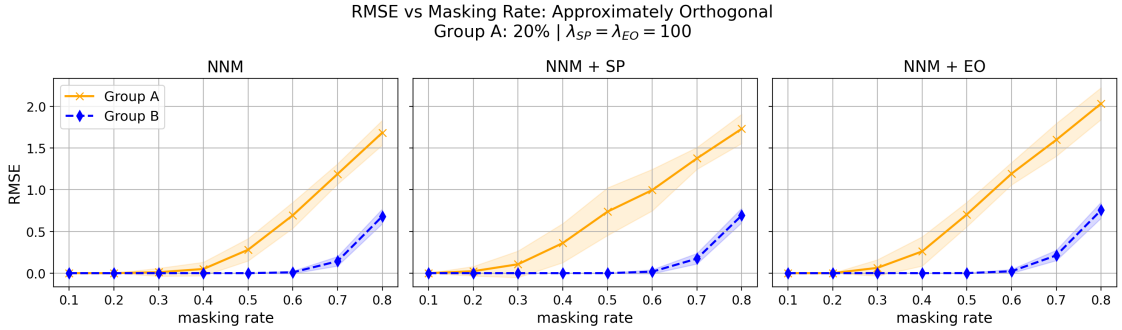
(a) Perturbed V.



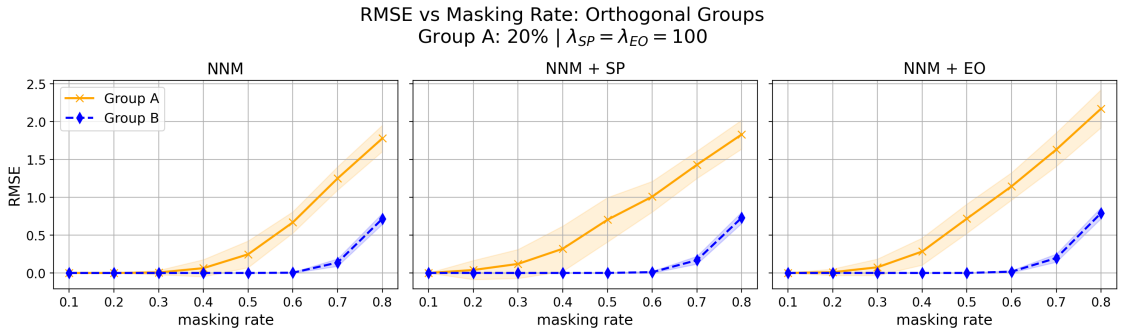
(b) Extended Minority V.

Figure 3.11: Group-wise RMSE as a function of masking rate. Methods are presented from left to right as vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. The minority-group fraction is fixed at 20%. For both the Perturbed V and Extended Minority V datasets, $\lambda_{\mathrm{SP}} = \lambda_{\mathrm{EO}} = 100$.

Finally, we consider the Approximately Orthogonal and Orthogonal Groups datasets, which differ in whether the group-specific latent subspaces are approximately or exactly orthogonal. On Approximately Orthogonal data, statistical parity produces little meaningful change in either group's RMSE, whereas equal opportunity increases reconstruction error for both groups. On Orthogonal Groups, we see a similar trend. Thus, neither fairness criterion improves the reconstruction error disparity in these settings.



(a) Approximately Orthogonal.



(b) Orthogonal Groups.

Figure 3.12: Group-wise RMSE as a function of masking rate. Methods are presented from left to right as vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. The minority-group fraction is fixed at 20%. For both the Approximately Orthogonal and Orthogonal Groups datasets, $\lambda_{\text{SP}} = \lambda_{\text{EO}} = 100$.

Overall, the synthetic experiments show that fairness regularization does not consistently improve reconstruction accuracy across groups. In the Shifted Means dataset, the worsening effect of statistical parity is expected because the two groups are constructed to have different true means, making exact recovery incompatible with statistical parity. This incompatibil-

ity explains the lower statistical parity penalty weight, $\lambda_{\mathrm{SP}} = 10$, required to obtain stable results on the Shifted Means dataset. However, surprisingly, statistical parity also increases reconstruction error in the other synthetic settings, where the groups have similar underlying mean values. In settings where vanilla nuclear norm minimization already provides comparable reconstruction accuracy, imposing statistical parity or equal opportunity can increase reconstruction error for both groups while creating a disparity where little or none previously existed. In settings where vanilla nuclear norm minimization exhibits a disparity, the fairness penalties generally increase the error gap by disproportionately increasing the reconstruction error of the minority group.

3.5.4.2 MovieLens Experiments

The MovieLens results differ from those of the synthetic experiments. We first examine the effect of the fairness-penalty weight on reconstruction. Figure 3.13 shows the group-wise RMSE as the penalty weight increases for both statistical parity and equal opportunity regularization.

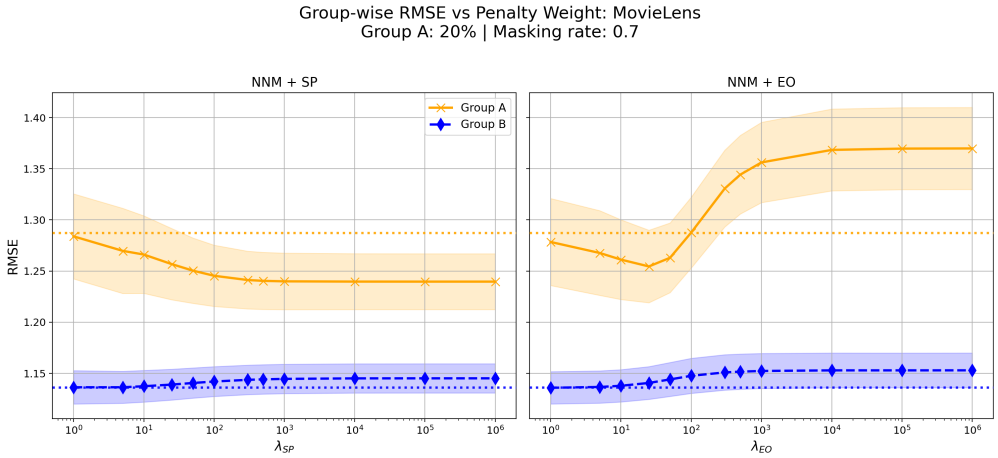


Figure 3.13: Group-wise RMSE as a function of the fairness-penalty weight on the MovieLens 100K dataset for statistical parity and equal opportunity regularization. The minority-group fraction is fixed at 20% and the masking rate is fixed at 0.7. The selected penalty weights are $\lambda_{\mathrm{SP}} = 100$ and $\lambda_{\mathrm{EO}} = 25$.

For equal opportunity, the penalty-weight sweep identifies a range in which the minority-group reconstruction error decreases while the majority-group error increases only slightly, decreasing the error gap between the two groups. Among the penalty weights considered, $\lambda_{\mathrm{EO}} = 25$ provides the largest reduction in the group-wise error gap, and we therefore use this value for the method-comparison experiments. For statistical parity, increasing the penalty weight similarly reduces minority-group RMSE and narrows the group-wise error gap, at the cost of a small increase in majority-group RMSE. We select $\lambda_{\mathrm{SP}} = 100$ for the method-comparison experiments.

Figure 3.14 compares group-wise reconstruction accuracy under vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. Both fairness regularizers slightly reduce the reconstruction error of group A while causing only a small increase in the error of group B. Consequently, both methods narrow the group-wise error gap. Under the error-based fairness criterion considered here, this constitutes an improvement in fairness, although accompanied by a small increase in majority-group error.

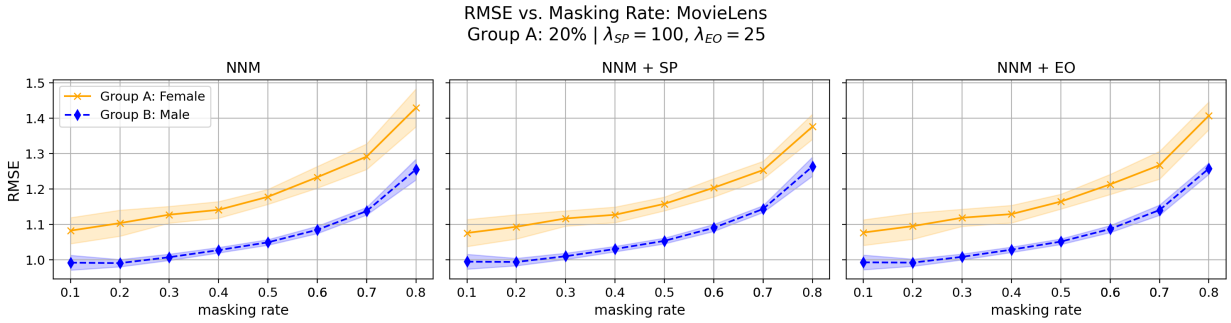


Figure 3.14: Group-wise RMSE as a function of masking rate. Methods are presented from left to right as vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. The minority-group fraction is fixed at 20%, with $\lambda_{\mathrm{SP}} = 100$ and $\lambda_{\mathrm{EO}} = 25$. Both fairness metric regularizers reduce minority-group RMSE and narrow the group-wise error gap relative to vanilla nuclear norm minimization, with a small increase in majority-group RMSE.

Figure 3.15 directly compares the absolute group-wise RMSE gap, $|\mathrm{RMSE}_A - \mathrm{RMSE}_B|$, across the three methods. At a masking rate of 0.8, statistical parity reduces the error gap

by approximately 5% relative to vanilla nuclear norm minimization. Equal opportunity also reduces the group-wise error gap across all masking rates, although the effect is smaller.

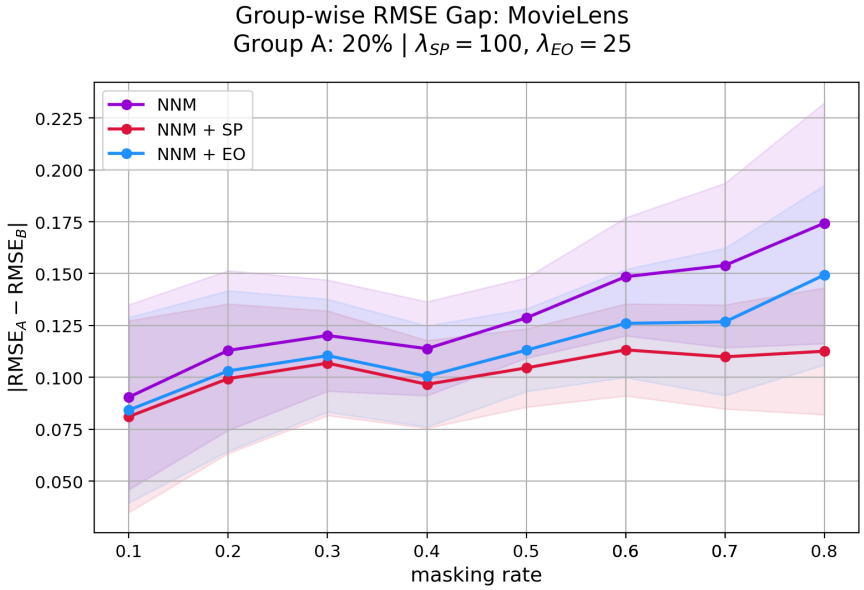


Figure 3.15: Absolute group-wise RMSE gap, $|\text{RMSE}_A - \text{RMSE}_B|$, as a function of masking rate on the MovieLens 100K dataset under vanilla nuclear norm minimization, statistical parity regularization, and equal opportunity regularization. The minority-group fraction is fixed at 20%, with $\lambda_{\text{SP}} = 100$ and $\lambda_{\text{EO}} = 25$.

The behavior of statistical parity on MovieLens differs from that observed across the synthetic datasets. Average ratings are nearly identical across gender groups, with $\mu_A \approx 3.69$ and $\mu_B \approx 3.68$, and the two groups have comparable proportions of observed ratings. Thus, the observed ratings are approximately consistent with statistical parity, and enforcing statistical parity improves error-based fairness on MovieLens. This contrasts with the synthetic datasets that also have similar group means but for which statistical parity does not improve reconstruction-error disparities.

As the underlying latent structure of the MovieLens data is unknown, the group means and observation rates alone cannot fully explain the different effects of fairness regularization across datasets. Nevertheless, the MovieLens experiments demonstrate that parity-based regularization can improve error-based fairness in a real-data setting, suggesting potential

for these approaches and motivating further investigation into the settings in which they can be effective.

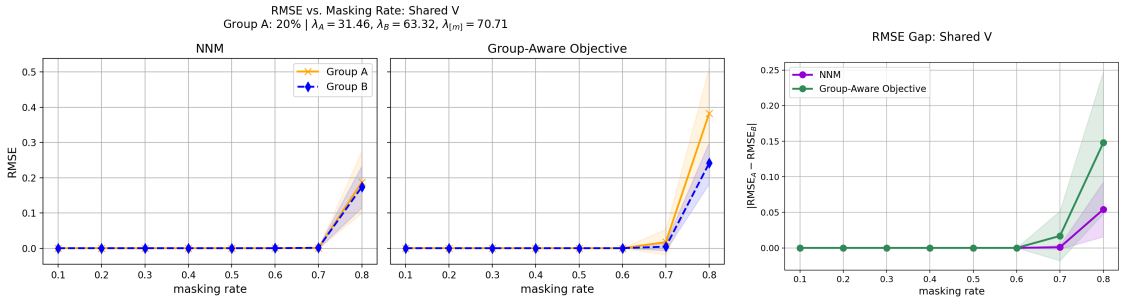
3.5.5 Group-Aware Matrix Completion Experiments

We evaluate the Group-Aware objective introduced in Eq. (3.4.4) across the synthetic and MovieLens datasets. We use the weighting scheme described in Section 3.2.4, with the weights for each experiment reported in the corresponding figure headings. As described in Section 3.5.1, synthetic experiments use 100×100 matrices, while MovieLens experiments use 300×300 matrices. Unless otherwise specified, the minority group comprises 20% of the rows.

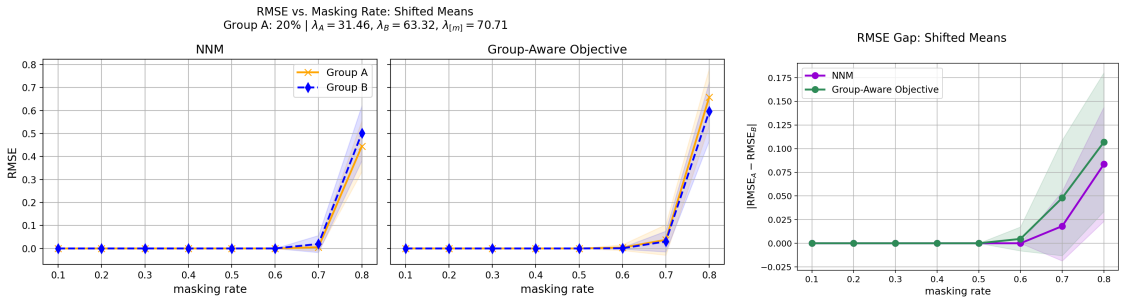
3.5.5.1 Synthetic Data Experiments

We first consider the Shifted Means and Shared V datasets, for which vanilla nuclear norm minimization exhibits little or no reconstruction-error disparity. As shown in Figure 3.16, the Group-Aware objective increases the reconstruction error of both groups in these settings and therefore does not provide an accuracy or fairness improvement over vanilla nuclear norm minimization. This is consistent with the fact that the two groups either share the same underlying latent structure or differ only through a mean shift, so separating the group-specific nuclear norm terms does not provide an apparent advantage.

We next consider the Perturbed V and Extended Minority V datasets, both of which exhibit reconstruction-error disparities under vanilla nuclear norm minimization. As shown in Figure 3.17, on Perturbed V, the Group-Aware objective reduces the reconstruction error of both groups, with a larger reduction for the majority group. However, the improvement in overall reconstruction accuracy does not translate into a reduction in the group-wise error gap. On Extended Minority V, the Group-Aware objective decreases the majority-group reconstruction error while slightly increasing the minority-group error, resulting in worsened



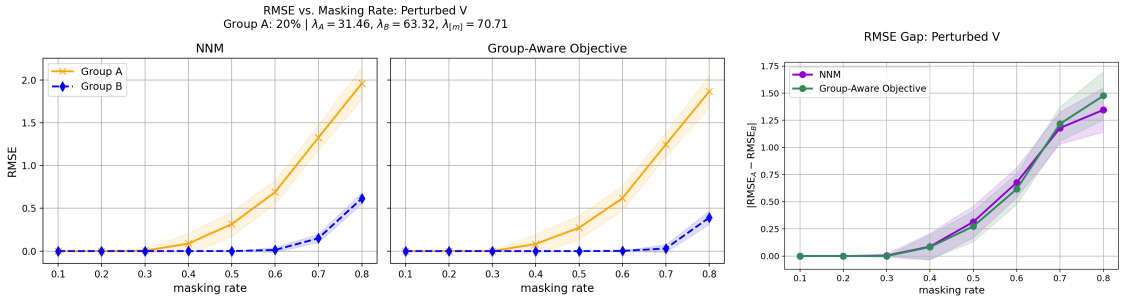
(a) Shared V.



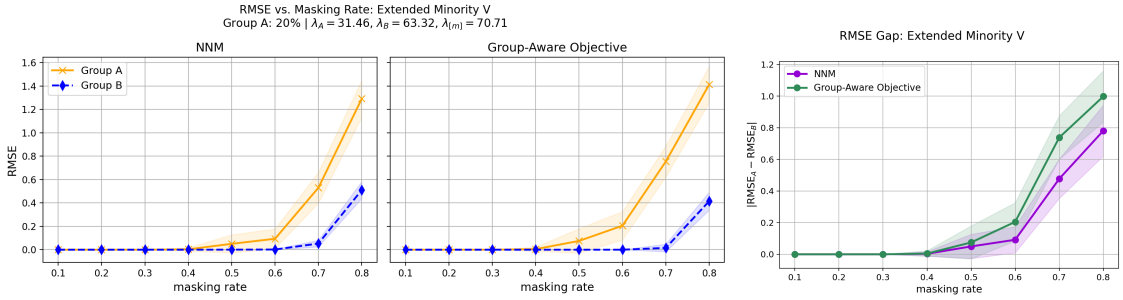
(b) Shifted Means.

Figure 3.16: Group-wise RMSE and the absolute difference in RMSE between groups as a function of masking rate for the Group-Aware objective on the Shared V and Shifted Means datasets, compared with vanilla nuclear norm minimization. The minority-group fraction is fixed at 20%.

error-based fairness. These results illustrate that allowing group-specific low-rank structure to be regularized separately can improve reconstruction accuracy without necessarily improving comparability of reconstruction error across groups.



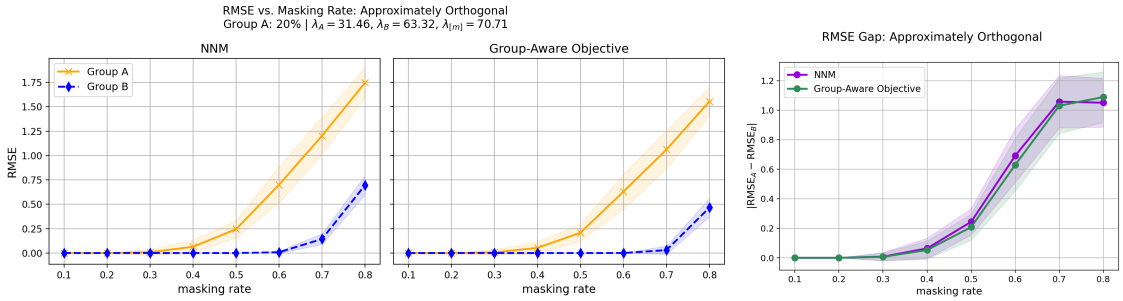
(a) Perturbed V.



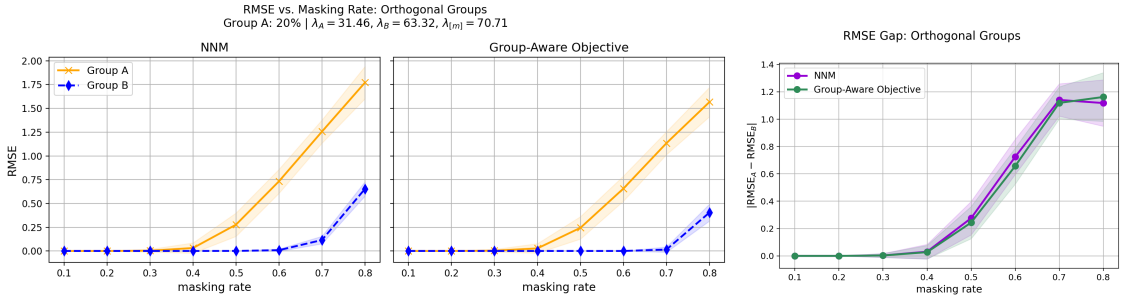
(b) Extended Minority V.

Figure 3.17: Group-wise RMSE and the absolute difference in RMSE between groups as a function of masking rate for the Group-Aware objective on the Perturbed V and Extended Minority V datasets, compared with vanilla nuclear norm minimization. The minority-group fraction is fixed at 20%.

We then consider the Approximately Orthogonal and Orthogonal Groups datasets, where the groups have relatively little or no overlap in their latent structures. As shown in Figure 3.18, on Approximately Orthogonal, the Group-Aware objective reduces reconstruction error for both groups, with the improvement benefiting the majority group more strongly. Despite this improvement in overall reconstruction accuracy, the group-wise error gap is not reduced. The method behaves similarly on the Orthogonal Groups dataset.



(a) Approximately Orthogonal.



(b) Orthogonal Groups.

Figure 3.18: Group-wise RMSE and the absolute difference in RMSE between groups as a function of masking rate for the Group-Aware objective on the Approximately Orthogonal and Orthogonal Groups datasets, compared with vanilla nuclear norm minimization. The minority-group fraction is fixed at 20%.

3.5.5.2 MovieLens Experiments

For the MovieLens experiments, the Group-Aware objective produces little meaningful change in either group's reconstruction error, beyond a slight increase in minority group error at high masking rates. As shown in Figure 3.19, the corresponding group-wise error gap slightly increases at high masking rates.

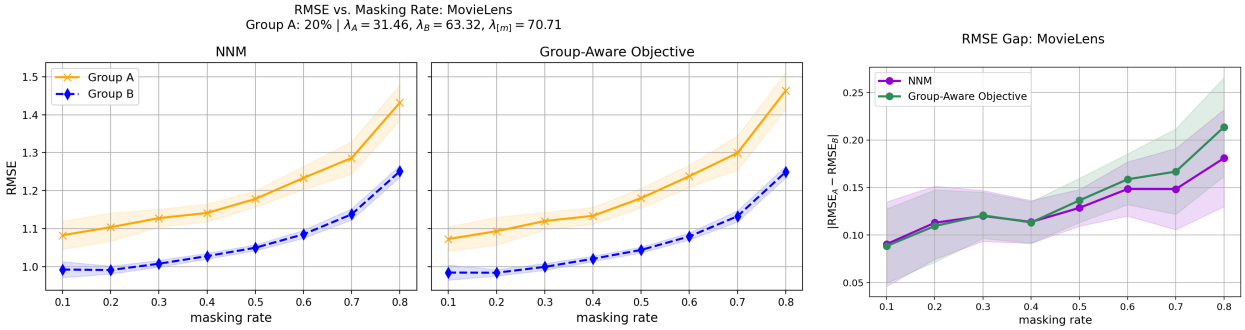


Figure 3.19: Group-wise RMSE and the absolute difference in RMSE between groups as a function of masking rate for the Group-Aware objective on the MovieLens 100K dataset, compared with vanilla nuclear norm minimization. The minority-group fraction is fixed at 20%.

The GAME method [GSA26] was originally developed to improve global reconstruction accuracy by accounting for group-specific low-rank structure. The authors of GAME report the greatest benefits when subgroups exhibit distinct low-rank structure, which is consistent with the improvements in overall reconstruction accuracy observed for the Perturbed V and Approximately Orthogonal datasets in our experiments. However, these improvements do not necessarily translate into improved error-based fairness. In particular, the Group-Aware objective can reduce reconstruction error more substantially for the majority group, leaving the gap in reconstruction accuracy unchanged or increased.

Our experiments use a weighting heuristic inspired by the GAME work, in which each group's weight is proportional to the square root of its number of observed entries. Under uniform masking, this gives the majority group a larger group-specific weight, placing more emphasis on its regularization. While this weighting may be appropriate for the original

goal of improving overall reconstruction accuracy, it does not explicitly target comparability of reconstruction error across groups. An alternative weighting scheme that places greater emphasis on the smaller or more difficult-to-reconstruct group may therefore have greater potential for improving error-based fairness. In our two-group formulation, this could be explored by assigning a larger weight to group A than to group B. Investigating how group-specific weights should be selected to balance global reconstruction accuracy and error-based fairness is a direction for future work.

3.5.6 Conclusion

Table 3.2 summarizes the empirical findings across the synthetic and MovieLens datasets. The experiments show that reconstruction accuracy can differ substantially across groups even when a single low-rank model provides a good overall fit. In the synthetic experiments, varying the relative size of the two groups and the relationship between their latent structures allows us to examine two sources of disparity: population imbalance and group-specific structure. Population imbalance is reflected in the dependence of the reconstruction-error gap on group size, while differences in group-specific structure can produce disparities that persist even when the groups are equally represented. The MovieLens experiments provide a corresponding real-data example in which female users have higher reconstruction error than male users even when female users constitute the larger group.

We then evaluated two approaches for incorporating group information or fairness criteria into matrix completion. The fairness-regularization experiments demonstrate that enforcing statistical parity or equal opportunity does not necessarily improve error-based fairness. On the synthetic datasets, the fairness penalties generally increase reconstruction error for the minority group and, in some cases, introduce a disparity where little or none existed under unregularized nuclear norm minimization. The Shifted Means dataset provides an expected example of this behavior because its two groups have different true means, making exact recovery incompatible with statistical parity. However, statistical parity also fails to improve

reconstruction-error disparities in the other synthetic settings where group means have similar values. On the MovieLens data, however, both statistical parity and equal opportunity reduce the reconstruction-error gap while producing only a small increase in majority-group error. Thus, the effectiveness of a fairness criterion depends on its relationship to the underlying data and to the particular notion of fairness being evaluated.

The Group-Aware objective improves reconstruction accuracy in some settings with distinct group-specific latent structure, however, these improvements do not consistently reduce disparities in reconstruction error across groups. In several settings, improvements in reconstruction accuracy are concentrated more strongly in the majority group, leaving the group-wise error gap largely unchanged or increasing it. The weighting scheme used in our experiments places greater weight on larger groups under uniform masking, which does not explicitly target comparable reconstruction accuracy across groups. Alternative choices of group-specific weights may therefore provide an avenue for balancing overall reconstruction accuracy with error-based fairness, and provide a direction for future work.

Table 3.2: Summary of empirical findings by dataset and method. For each dataset, we report whether vanilla nuclear norm minimization (nuclear norm minimization) exhibits a group-wise reconstruction error gap, the primary source of the disparity, and the change in group-wise reconstruction error under each fairness intervention relative to vanilla nuclear norm minimization. Here, $\uparrow$ denotes an increase in error, $\downarrow$ denotes a decrease in error, and $\approx$ denotes no meaningful change. Two arrows indicate a larger change for the indicated group relative to the other group.

Dataset	**Vanilla NNM error gap**	**Primary source of disparity**	**Group-aware objective**	**NNM + EO**	**NNM + SP**
Shifted Means	None	None	$\uparrow$ both	$\uparrow$ both	$\uparrow\uparrow$ minority, $\uparrow$ majority
Shared V	None	None	$\uparrow$ both	$\uparrow$ both	$\uparrow\uparrow$ minority, $\uparrow$ majority
Perturbed V	Yes	Both	$\approx$ minority, $\downarrow$ majority	$\uparrow$ minority, $\approx$ majority	$\uparrow$ minority, $\approx$ majority
Extended Minority V	Yes	Both	$\uparrow$ minority, $\downarrow$ majority	$\uparrow$ both	$\uparrow\uparrow$ minority, $\uparrow$ majority
Approx. Orthogonal	Yes	Population imbalance	$\downarrow$ both	$\uparrow$ both	$\approx$ both
Orthogonal Groups	Yes	Population imbalance	$\approx$ both	$\uparrow$ minority, $\approx$ majority	$\approx$ both
MovieLens 100K	Yes	Group-specific structure	$\approx$ both	$\downarrow$ minority, $\approx$ majority	$\downarrow$ minority, $\approx$ majority

Overall, the only setting in which we observed an improvement in error-based fairness was the MovieLens experiment, where both statistical parity and equal opportunity reduced the group-wise reconstruction-error gap at the cost of a marginal increase in majority-group

error. This result suggests potential for fairness regularization in real-data settings and motivates further investigation into the data characteristics and regularization strategies under which such improvements can be achieved.

CHAPTER 4

Conclusion

This dissertation developed and analyzed methods for accounting for group or source specific structure in matrix and tensor models. In the first part, we developed Stratified Non-negative Tensor Factorization that learns globally shared and stratum-level structures. In the second, we studied fairness in matrix completion via nuclear norm minimization, identifying sources of disparities in reconstruction accuracy across groups and evaluating approaches for addressing these disparities.

Chapter 2 developed Stratified Non-Negative Tensor Factorization, an extension of Stratified Non-Negative Matrix Factorization to tensor-valued data. Stratified Non-Negative Tensor Factorization jointly models globally shared topics and stratum-level features while preserving the geometric structure of tensor-valued data. This allows the method to recover both structure common to all data sources and structure specific to each source without flattening the data. We derived multiplicative update rules for the unregularized objective and for a total-variation-regularized variant, and showed that the explicit tensor representation makes mode-specific regularization possible. Applying total variation regularization to the spatial modes of the global topics allows the method to effectively denoise image content while preserving stratum-level watermarks. Empirically, the method produces interpretable stratum-level features on text data and outperforms Stratified Non-Negative Matrix Factorization on image data with substantially fewer learnable parameters, indicating that the tensor representation can provide greater representational efficiency for the settings considered.

Chapter 3 studied fairness in matrix completion through the lens of group-wise reconstruction accuracy. We showed that nuclear norm minimization can produce disparities in reconstruction accuracy when the data exhibit population imbalance or group-specific latent structure. In particular, disparities arising from group-specific structure can persist even when groups are equally represented. We formalized statistical parity and equal opportunity for the matrix completion setting, and investigated two approaches for incorporating fairness and group information into the reconstruction objective: fairness-metric regularization and a group-aware matrix completion formulation. Our experiments demonstrate that the effectiveness of these approaches depends on the underlying data structure and the fairness criterion being considered. Fairness-metric regularization did not consistently improve error-based fairness on the synthetic datasets, while it reduced disparities in reconstruction accuracy on the MovieLens data. The group-aware formulation improved reconstruction accuracy in some settings with distinct group-specific structure, but did not consistently reduce disparities between groups. These results demonstrate that group information can affect the behavior of matrix completion in meaningful ways, with different objectives offering different trade-offs between reconstruction accuracy and fairness metrics across groups.

As machine learning methods continue to be applied to increasingly large and complex datasets, large-scale models can provide valuable insight by revealing patterns at the population level. However, accurately representing individual- or group-level features can require attention to structure that may be obscured by global patterns. Accounting for meaningful local structure can therefore complement population-level trends and lead to more robust representations and predictions. This thesis demonstrates the value of accounting for group- or source-specific structure in machine learning. Doing so can lead to more flexible and interpretable methods or provide ways to evaluate fairness at the group level and, in some settings, improve it, providing a step toward fairer machine learning systems.

REFERENCES

[ABC17] Dylan Anderson, Aleksander Bapst, Joshua Coon, Aaron Pung, and Michael Kudenov. "Supervised non-negative tensor factorization for automatic hyperspectral feature extraction and target discrimination." In *Proc. SPIE*, volume 10198, pp. 263–275. SPIE, 2017.

[BCD19] Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Li Wei, Yi Wu, Lukasz Heldt, Zhe Zhao, Lichan Hong, Ed H Chi, et al. "Fairness in recommendation ranking through pairwise comparisons." In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2212–2220, 2019.

[Ber97] D P Bertsekas. "Nonlinear Programming." *Journal of the Operational Research Society*, **48**(3):334–334, 1997.

[BHN23] Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press, 2023.

[BSO18] Robin Burke, Nasim Sonboli, and Aldo Ordonez-Gauger. "Balanced neighborhoods for multi-sided fairness in recommendation." In *Conference on fairness, accountability and transparency*, pp. 202–214. PMLR, 2018.

[CC70] J Douglas Carroll and Jih-Jie Chang. "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition." *Psychometrika*, **35**(3):283–319, 1970.

[CCS10] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. "A singular value thresholding algorithm for matrix completion." *SIAM Journal on optimization*, **20**(4):1956–1982, 2010.

[CH24] Simon Caton and Christian Haas. "Fairness in machine learning: A survey." *ACM Computing Surveys*, **56**(7):1–38, 2024.

[CP10] Emmanuel J Candes and Yaniv Plan. "Matrix completion with noise." *Proceedings of the IEEE*, **98**(6):925–936, 2010.

[CR09] Emmanuel J Candès and Benjamin Recht. "Exact matrix completion via convex optimization." *Foundations of Computational Mathematics*, **9**(6):717–772, 2009.

[CYN23] James Chapman, Yotam Yaniv, and Deanna Needell. "Stratified-NMF for Heterogeneous Data." In *Asilomar Conf. Signals Syst. Comp.*, pp. 614–618. IEEE, 2023.

[DB16] Steven Diamond and Stephen Boyd. "CVXPY: A Python-embedded modeling language for convex optimization." *Journal of Machine Learning Research*, **17**(83):1–5, 2016.

[Dev08] Karthik Devarajan. "Nonnegative matrix factorization: An analytical and interpretive tool in computational biology." *PLoS Computational Biology*, **4**(7):e1000029, 2008.

[DHP11] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Rich Zemel. "Fairness Through Awareness.", 2011.

[EK04] Julian Eggert and Edgar Korner. "Sparse coding and NMF." In *Int. Jt. Conf. Neural Netw. (IEEE Cat. No. 04CH37541)*, volume 4, pp. 2529–2533. IEEE, 2004.

[Faz02] Maryam Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, March 2002.

[FNP21] Simon Foucart, Deanna Needell, Reese Pathak, Yaniv Plan, and Mary Wootters. "Weighted Matrix Completion From Non-Random, Non-Uniform Sampling Patterns." *IEEE Transactions on Information Theory*, **67**(2):1244–1263, 2021.

[FP20] James R. Foulds and Shimei Pan. "Are Parity-Based Notions of AI Fairness Desirable?" *IEEE Data Eng. Bull.*, **43**:51–73, 2020.

[Gil14] Nicolas Gillis. "The why and how of nonnegative matrix factorization." *Regularization, optimization, kernels, and support vector machines*, **12**(257):257–291, 2014.

[Gil20] Nicolas Gillis. *Nonnegative Matrix Factorization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2020.

[GSA26] Hamza Golubovic, Matthew Shen, Genevera I. Allen, and Tarek M. Zikry. "Group-Aware Matrix Estimation and Latent Subspace Recovery.", 2026.

[Har70] Richard A Harshman et al. "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis." *UCLA Work. Pap. Phonetics*, **16**(1):84, 1970.

[HHM53] Morris H Hansen, William N Hurwitz, and William G Madow. *Sample survey methods and theory. Vol. I. Methods and applications.* John Wiley, 1953.

[HK15] F. Maxwell Harper and Joseph A. Konstan. "The MovieLens Datasets: History and Context." *ACM Transactions on Interactive Intelligent Systems*, **5**(4):1–19, 2015.

[HPS16a] Moritz Hardt, Eric Price, and Nathan Srebro. "Equality of Opportunity in Supervised Learning.", 2016.

[HPS16b] Moritz Hardt, Eric Price, and Nati Srebro. "Equality of Opportunity in Supervised Learning." In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, pp. 3315–3323. Curran Associates, Inc., 2016.

[HPT20] Mark Herbster, Stephen Pasteris, and Lisa Tse. "Online Matrix Completion with Side Information.", 2020.

[KB09] Tamara G Kolda and Brett W Bader. "Tensor decompositions and applications." *SIAM Rev.*, **51**(3):455–500, 2009.

[KBV09] Yehuda Koren, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." *Computer*, **42**(8):30–37, 2009.

[KGN24] Lara Kassab, Erin George, Deanna Needell, Haowen Geng, Nika Jafar Nia, and Aoxi Li. "Towards a Fairer Non-negative Matrix Factorization.", 2024.

[KMR16] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. "Inherent Trade-Offs in the Fair Determination of Risk Scores.", 2016.

[KPA19] Jean Kossaifi, Yannis Panagakis, Anima Anandkumar, and Maja Pantic. "TensorLy: Tensor Learning in Python." *J. Mach. Learn. Res.*, **20**(26):1–6, 2019.

[LCB10] Yann LeCun, Corinna Cortes, and CJ Burges. "MNIST handwritten digit database." *ATT Labs. Available: http://yann.lecun.com/exdb/mnist*, **2**, 2010.

[LCX23] Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, Juntao Tan, Shuchang Liu, and Yongfeng Zhang. "Fairness in recommendation: Foundations, methods, and applications." *ACM Transactions on Intelligent Systems and Technology*, **14**(5):1–48, 2023.

[Lin07] Chih-Jen Lin. "On the convergence of multiplicative update algorithms for non-negative matrix factorization." *IEEE Trans. on Neural Netw.*, **18**(6):1589–1596, 2007.

[LS99] Daniel D Lee and H Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." *Nature*, **401**(6755):788–791, 1999.

[LS00] Daniel Lee and H Sebastian Seung. "Algorithms for non-negative matrix factorization." *Adv. Neural Inf. Process. Syst.*, **13**, 2000.

[MHT10] Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. "Spectral regularization algorithms for learning large incomplete matrices." *The Journal of Machine Learning Research*, **11**:2287–2322, 2010.

[MMS21] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. "A Survey on Bias and Fairness in Machine Learning." *ACM Computing Surveys*, **54**(6):1–35, 2021.

[MR10] Sébastien Marcel and Yann Rodriguez. "Torchvision the machine-vision package of torch." In *Proc. ACM 18 Int. Conf. Multimedia*, pp. 1485–1488, 2010.

[Nik04] Mila Nikolova. "A variational approach to remove outliers and impulse noise." *J. Math. Imaging Vision*, **20**(1):99–120, 2004.

[OCP16] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. "Conic optimization via operator splitting and homogeneous self-dual embedding." *Journal of Optimization Theory and Applications*, **169**(3):1042–1068, 2016.

[PRW17] Geoff Pleiss, Manish Raghavan, Felix Wu, Jon Kleinberg, and Kilian Q. Weinberger. "On Fairness and Calibration.", 2017.

[PVG11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: Machine Learning in Python." *J. Mach. Learn. Res.*, **12**:2825–2830, 2011.

[RL08] Jason Rennie and Ken Lang. "Home Page for 20 Newsgroups Dataset.", 2008.

[RO94] Leonid I Rudin and Stanley Osher. "Total variation based image restoration with free local constraints." In *IEEE Image Proc.*, volume 1, pp. 31–35. IEEE, 1994.

[Rod13] Paul Rodríguez. "Total variation regularization algorithms for images corrupted with different noise models: a review." *J. Electr. Comput. Eng.*, **2013**(1):217021, 2013.

[SG17] Yaser Esmaeili Salehani and Saeed Gazor. "Smooth and sparse regularization for NMF hyperspectral unmixing." *IEEE Trans. Geosci. Remote Sens.*, **10**(8):3677–3692, 2017.

[SH94] Ferdinando S Samaria and Andy C Harter. "Parameterisation of a stochastic model for human face identification." In *Proc. Appl. Comput. Vis.*, pp. 138–142. IEEE, 1994.

[SH05] Amnon Shashua and Tamir Hazan. "Non-negative tensor factorization with applications to statistics and computer vision." In *Proc. 22nd Int. Conf. Mach. Learn.*, pp. 792–799, 2005.

[STM18] Samira Samadi, Uthaipon Tantipongpipat, Jamie Morgenstern, Mohit Singh, and Santosh Vempala. "The Price of Fair PCA: One Extra Dimension.", 2018.

[SVC24] Alexander Sietsema, Zerrin Vural, James Chapman, Yotam Yaniv, and Deanna Needell. "Stratified Non-Negative Tensor Factorization." In *Proc. 58th Asilomar Conf. Signals, Syst., Comput.*, pp. 260–264, Pacific Grove, CA, USA, 2024.

[TA22] Riku Togashi and Kenshi Abe. "Fair matrix factorisation for large-scale recommender systems." *arXiv preprint arXiv:2209.04394*, 2022.

[TSS20] Uthaipon Tantipongpipat, Samira Samadi, Mohit Singh, Jamie Morgenstern, and Santosh Vempala. "Multi-Criteria Dimensionality Reduction with Applications to Fairness.", 2020.

[WMZ23] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. "A survey on the fairness of recommender systems." *ACM Transactions on Information Systems*, **41**(3):1–43, 2023.

[XQZ19] Fengchao Xiong, Yuntao Qian, Jun Zhou, and Yuan Yan Tang. "Hyperspectral Unmixing via Total Variation Regularized Nonnegative Tensor Factorization." *IEEE Trans. Geosci. Remote Sens.*, **57**(4):2341–2357, 2019.

[YH17] Sirui Yao and Bert Huang. "Beyond parity: Fairness objectives for collaborative filtering." *Advances in neural information processing systems*, **30**, 2017.

[ZVR17] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rogriguez, and Krishna P Gummadi. "Fairness constraints: Mechanisms for fair classification." In *Artificial intelligence and statistics*, pp. 962–970. PMLR, 2017.

[ZWS13] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. "Learning fair representations." In *International conference on machine learning*, pp. 325–333. PMLR, 2013.