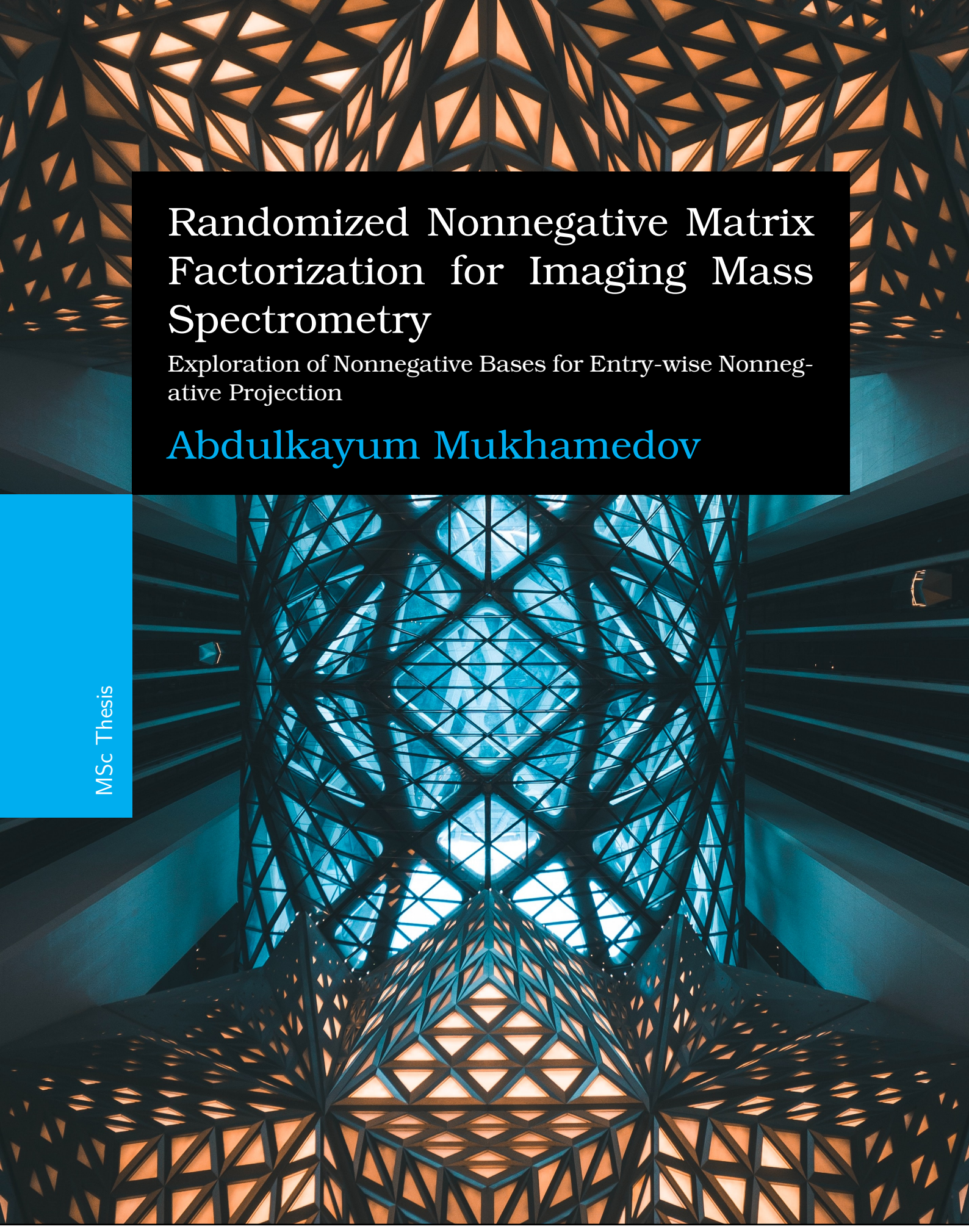


Randomized Nonnegative Matrix Factorization for Imaging Mass Spectrometry

Exploration of Nonnegative Bases for Entry-wise Nonnegative Projection

Abdulkayum Mukhamedov

MSc Thesis

TU Delft Delft University of Technology

Delft Center for Systems and Control

Randomized Nonnegative Matrix Factorization for Imaging Mass Spectrometry

Exploration of Nonnegative Bases for Entry-wise Nonnegative Projection

MSc Thesis

Abdulkayum Mukhamedov

April 16, 2025

Faculty of Mechanical Engineering (ME) · Delft University of Technology

Abstract

Imaging Mass Spectrometry (IMS) data calls for computational methods to perform factorizations efficiently that facilitate interpretable components. Despite the widespread use of Nonnegative Matrix Factorization (NMF) for IMS-related dimensionality reduction, conventional NMF algorithms are not up to par to scale for extremely large datasets. To remedy this problem, this thesis explores Randomized Nonnegative Matrix Factorization as an efficient alternative to compute IMS data on whole-body mouse pup tissue data.

In this thesis, the author demonstrates how Randomized NMF inherits its two-stage framework for computation from Randomized SVD. The main bottleneck between Randomized NMF replicating the success of its predecessor is the lower-dimensional basis used to compress the data matrix. Although the basis has been conventionally effective for Randomized SVD with provable guarantees, the lack of entry-wise nonnegativity constraint requires existing methods to compromise the projection quality and interpretability and rely on the relaxation of nonnegativity constraints within the update rules.

Therefore, this is the first work in the field of Randomized Nonnegative Matrix Factorization to motivate and explore the use of alternative bases that directly satisfy entry-wise nonnegativity. To achieve this, orthogonal variants of NMF, such as Projective NMF and Approximately Orthogonal NMF methods, are employed to extract nonnegative approximately orthogonal bases. To avert the downsides of such methods and their lack of compatibility within the pipeline, this thesis also proposes a method that approximates the subspace on the QR-based basis whilst satisfying nonnegativity constraints.

Contents

Preface

This final MSc Thesis report is a part of the graduation project of the Systems and Control Master's program at Delft University of Technology. The main purpose of the thesis is to present the Randomized Nonnegative Matrix Factorization as a powerful tool for dimensionality reduction for Imaging Mass Spectrometry data. The aim of the research is not only to justify the compatibility of the method with Imaging Mass Spectrometry, but simultaneously identify the possible research contribution towards improvement of the compressed NMF approach.

Higher spatial and spectral resolutions from continuously developing IMS technology contribute to higher amount of information stored, and thereby larger data sizes. A widely accepted approach to deal with such large data sizes is to employ linear dimensionality reduction techiques such as SVD or PCA. A lower-dimensional representation, or a lower-dimensional factorized version of the original dataset allows to represent the the data in terms of a combination (usually a product of) lower dimensional matrices. Given such lower-dimensional representations facilitating visual analysis, and the components carrying physical meaning facilitates interpretation for the involved analyst in research, diagnosis or other imaging-related tasks. Given this focus, this thesis thoroughly delves into an randomized Nonnegative Matrix Factorization to obtain a part-based representation of the data in an efficient manner.

Randomized factorization methods offer guarantees on the quality of approximation and allow to make informative claims on the amount of discarded information. The inherent guarantees on preservation of Euclidean geometry in lower-dimensional space informs on the appropriate methodology to take, where subsequent decisions aim to simultaneously catered to a parts-based interpretable representation by NMF whilst simultaneously enjoying the benefits of fast randomized NMF algorithms.

A physical limitation with regard to this project will concern the use of a personal laptop with 8GB of RAM Lenovo Yoga 720. This incentivizes to utilize memory-efficient approaches, and simultaneously serve as an example that the future of medicine and high-resolution equipment can become more accessible to institutions, individuals or entities with substantially lower available computational power. Dealing with methods concerned in our thesis will involve transferring a higher-dimensional problem into a lower-dimensional problem upon which fast algorithms could be leveraged.

Acknowledgements

I would like to acknowledge the efforts and guidance of my supervisors dr. Van de Plas, and Roger Moens throughout the research process. I would like to thank Roger for his expertise, regular meetings, and ability to convey feedback to make the most out of the research process. Dr. Van de Plas played an important role in the holistic guidance of the project, redirecting and and refining the goal throughout the thesis. I would also like to thank Paul-Louis Delacour and the rest of the research group for additional opportunities to learn new concepts within the IMS field, and get accustomed to presenting research in a compelling manner. The cohesion and attention to detail within the research group have overall contributed to a better standard of research.

The study environment and the MSc Systems and Control cohort has also played an important role in shaping me up as a student. The diverse, supportive, and hard-working student friend group made the study experience very motivating and contributed towards an improved vision for this project.

I would also like to thank Prof. N. Takahashi for sharing his perspective, as well as providing their implementation for the method featured in the 2021 Masuda et al. paper [1] [2]. It is a pleasure to see the authors provide an additional insight to their argumentation and share perspectives to researchers from across the globe.

Last but not least, I extend my gratitude, love and appreciation to my family for their unwavering support which invigorated my passion for learning. This journey has been very rewarding because of the understanding, support and trust from my family.

Chapter 1

Introduction

Modern medical imaging facilitates diagnosis, treatment selection, and a variety of other image-guided interventions [3]. Among modern medical imaging instrumentation is the rapidly advancing Imaging Mass Spectrometry (IMS) continuously developing in throughput, spatial, and mass resolutions [4], accompanied by large research interest reflected in an increasing number of IMS-related publications [5]. However, Imaging Mass Spectrometry datasets are very large and pose significant data processing challenges. Therefore, it is important to alleviate the computational strain by reducing the dimensionality of IMS datasets whilst retaining important information [6].

Linear matrix factorization is one of the most widely used methods to reduce the dimensionality of Imaging Mass Spectrometry data whilst preserving the majority of the information. Among such methods, Nonnegative Matrix Factorization (NMF) prevails in its ability to produce an interpretable parts-based representation and is a well-established method to reduce Imaging Mass Spectrometry.

However, as IMS data sizes grow, applying conventional NMF algorithms poses concerns over scalability and efficient computation. Therefore, the introductory chapter motivates the use of Randomized NMF and presents a key problem within its computational framework. Subsequently, the remainder of the thesis can steer towards the exploration of appropriate solutions.

1-1 Imaging Mass Spectrometry

Imaging Mass Spectrometry (IMS) is an advanced imaging instrumentation tool that visualizes the spatial distribution of bio-molecules and maps specific molecules to two-dimensional coordinates of the original sample [7] [8]. IMS also offers numerous advantages in direct molecular analysis of the whole body tissue section over traditional methods [9].

The imaging technique is label-free [10] and captures the distribution of a multitude of classes of proteins, metabolites, pharmaceuticals, and inorganics. The imaging technique is capable

of producing localization maps for biological samples varying in the degree of chemical and structural complexity in *in vivo*, *in vitro*, and *in situ* i.e. in living organisms, in a laboratory environment, and in natural environment respectively [11] [12]. The core principle behind IMS is ionization; a process characterized by the exposure of molecules to a beam of ions, also known as a microprobe, at an individual cell in a grid structure. The subsequent transfer of the generated ions to the mass analyzer results in the mass spectrum at each individual cell [11] [13] as shown in Figure 1-1 below.

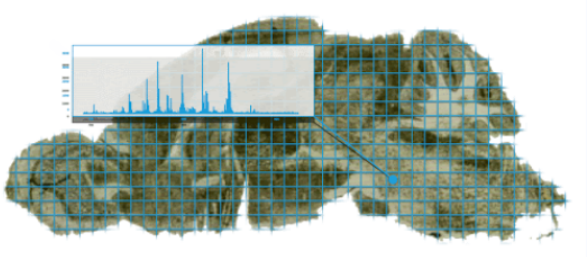


Figure 1-1: IMS Ionization Visual from Aspect Analytics [14]

The most widely used method of ionization is Matrix-assisted Laser Desorption/Ionization (MALDI) IMS [6], however depending on the applications other methods such as Secondary Ion Mass Spectrometry (SIMS), Laser Desorption Ionization (LDI) IMS, and Desorption Electrospray Ionization (DESI) IMS can be used [15]. Regardless of the ionization variant, the principle behind the growing data sizes remains the same, thus drawing focus toward a selection of an appropriate dimensionality reduction technique.

1-1-1 Growing Data Sizes IMS Data

The aforementioned description alludes to multiple factors related to growing data sizes. First, a finer grid structure with an increased number of pixels and an increased range of mass bins in the mass analyzer results in a larger data size.

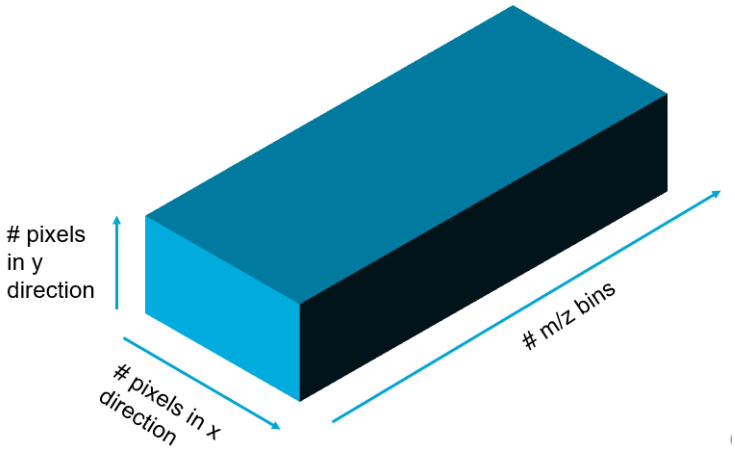


Figure 1-2: IMS Data Tensor Representation

Figure 1-2 above visualizes the three-way tensor data representation of IMS with pixels in x and y directions, as well as the m/z bins range. As imaging resolution increases, the pixel size decreases, and the number of pixels in x and y directions grows, contributing to a larger size of the data tensor. A similar contribution to the data size is evident in the third mode, m/z bins axis, representing the number of mass-to-charge ratio bins.

Given the general understanding of the IMS data tensor, one can further deduce the size of a common dataset with high resolution and mass bin range. A 512×512 image size, along with 100,000 mass bins stored in single precision, would amount to a total of 97.66GB of storage memory. Therefore, methods that involve memory-intensive operations exclude the use of computers with limited memory and processing capacity.

Moreover, the immense data sizes pose challenges for the generation of labels due to lack of availability or required labor time [16], which substantially impedes the use of supervised methods. Their unsupervised counterparts, such as clustering and dimensionality reduction, have recurrently proved their usefulness for facilitating the interpretation of IMS data [17]. In particular, complete simultaneous analysis of IMS data with minimal prior knowledge of the biologic sample is preferred during the identification of spatially relevant regions [18].

1-1-2 Matricized Dimensionality Reduction

The original three-way tensor can be considered in matricized form by unfolding along the third mode, the m/z bins mode, in turn concatenating the first two spatial modes [19] resulting in a two-dimensional matrix X.

$$\textit{Let } X \in \mathbb{R}^{M \times N} \textit{ be the Input Matrix}$$

In the context of IMS, the dimensions M and N are large, however a much smaller number of features K is assumed to represent the majority of the variance in the data. M thus represents the spatial dimension, whilst N represents the spectral dimension. The number of key features K is thus much smaller than the dimensions M and N as follows:

$$\textit{Let } K << M, N \textit{ be the Fixed Rank}$$

For dimensionality reduction via matrix factorization, the best possible approximation for rank-K factorization is known. The result for optimal rank-K factorization is famously described by the Eckart-Young-Mirsky theorem [20] [21]:

Theorem 1-1.1 (Eckart-Young-Mirsky)**.** *For either the 2-norm or the Frobenius norm*

$$\|X - X_K\| \leq \|X - M\|$$

In addition,

$$\|X - X_K\| = \begin{cases} \sigma_{K+1} & \textit{for the } \|\cdot\|_2 \textit{ norm} \\ \left(\sum_{i=K+1} \sigma_i^2\right)^{1/2} & \textit{for the } \|\cdot\|_F \textit{ norm} \end{cases}$$

Theorem 1-1.1 describes the relation between the minimal rank-K approximation error in terms of singular value(s) of the input matrix X. The error is equal to the largest singular value or the sum of singular values discredited due to truncation for the 2-norm and Frobenius norm, respectively. Any choice of an approximated matrix M will thus, at best, equate to this smallest error.

The optimal factorization X_K in the truncated version of Singular Value Decomposition (SVD):

$$\begin{gathered} X = U\Sigma V^T \\ U \in \mathbb{R}^{M\times M}, \Sigma \in \mathbb{R}^{M\times N}, V^T \in \mathbb{R}^{N\times N} \\ U^T U = I, V^T V = I \\ \Sigma = \text{diag}(\sigma_1, \cdots, \sigma_N) \end{gathered} \tag{1-1}$$

Computing X_K by selecting the index range up until the Kth index as such:

$$X_K := \sum_{k=1}^{K} \sigma_k u_k v_k^T \tag{1-2}$$

The computation of truncated SVD requires an SVD on a full matrix, thus for large data matrices this poses a significant computational bottleneck. Halko et al. have introduced a framework for Randomized Singular Value Decomposition with theoretical guarantees on low-rank approximation [22].

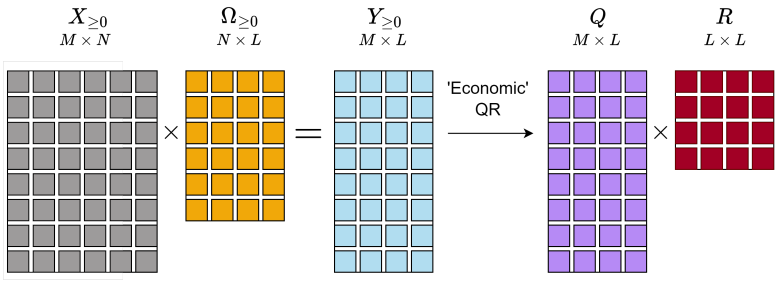


Figure 1-3: Sketching and Orthonormalization

The dimensionality of the dataset is reduced via multiplication of $X \in \mathbb{R}^{M\times N}$ by random matrix $\Omega \in \mathbb{R}^{N\times L}$ as shown in Figure 1-3. This step is called sketching, where L is the sketching dimension. Each entry of Ω is taken from a random distribution, and the subsequent matrix product results in a tall-skinny sample matrix $Y \in \mathbb{R}^{M\times L}$.

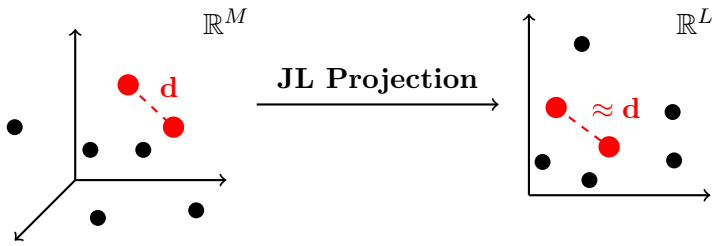


Figure 1-4: Johnson–Lindenstrauss Projection (Adapted from [23], [24])

The sketching step is motivated by the Johnson-Lindenstrauss lemma. Johnson-Lindenstrauss lemma states that for certain Lipschitz mapping functions, the distances are roughly preserved in the lower-dimensional space a tolerance factor [25] [26]. A common choice for Ω within Randomized Factorization includes Gaussian and Uniform distributions [22][27][2].

Similarly, $L << M, N$, where the sketching dimension L is slightly higher than fixed rank K. Formally, L is defined as $L = K + P$ where P is an oversampling parameter. Generally, the choice of P usually involves 2 to up K additional samples [22].

An orthonormal basis to the range space of Y is obtained via orthonormalization of tall-skinny matrix Y [22]. A common method for extracting such a basis is economic QR factorization.

Q from QR factorization is used to compress X and form a surrogate input matrix B:

$$\textit{Form Small Matrix } B \textit{ via } B = Q^T X \in \mathbb{R}^{L \times N}$$

$$\textit{where } Q \textit{ is an orthonormal basis to the range space of } Y$$

Instead of performing an SVD on a large input matrix, the dimensionality reduction strategy in Figure 1-3 allows to perform SVD on a reduced matrix:

$$\textit{Perform SVD on small matrix } B = \tilde{U}\Sigma V^T$$

where the U can be recovered via a decompression operation or a projection back to the higher-dimensional space:

$$\textit{Recover } U = Q\tilde{U}$$

.

The approximation of X via Q and B can be described as QB decomposition, and further related to SVD:

$$X = QQ^T X = QB = QQ^T U\Sigma V^T = Q\tilde{U}\Sigma V^T$$

where $QQ^T = \mathcal{P}$ is an orthonormal projection matrix. The orthonormal projection operator $\mathcal{P}$ onto the column space of X indicated by col(X) in Figure 1-5. Any vector $x \in \mathbb{R}^M$ in

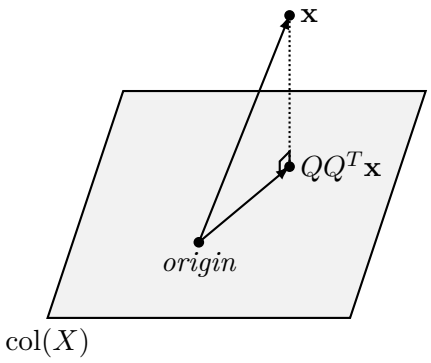


Figure 1-5: Figure A-5: Orthonormal Projection (Adapted from [24])

M-dimensional space is transformed to its orthogonal projection where the projected vector lies in the column space of X [24].

Conclusively, one may provide a simplified description of the expected error of QB decomposition with a multiplicative factor applied on $K+1$th largest singular value σ_K [22]:

$$\mathbb{E}\|X - QB\| \leq \left[1 + \sqrt{\frac{1}{P-1}} + \frac{e\sqrt{K+P}}{P} \cdot \sqrt{N-K}\right]^{\frac{1}{2Q+1}} \sigma_{K+1} \tag{1-3}$$

The power number Q is the power involved in power or subspace iterations used to enhance the amplify the effect of dominant singular vectors. The expect error can thus be controlled by increasing the oversampling parameter P, the rank K, and power Q. Consequently, the QB approximation of the original input matrix is only a small multiplicative factor away from the optimal Eckart-Young-Mirsky bound. Therefore, with these additional measures, it is possible to generate a Q such that QB factorization is near-optimal with regard to the Eckart-Young-Mirsky theorem.

The potential of Randomized Factorization will be explored on a whole-body mouse pup dataset using a 1 week old C57BL/6 control mouse. Mouse pup tissue was cryo-sectioned at 20 μm thickness using a CM3050 S cryostat and thaw-mounted onto conductive indium tin oxide coated glass slides [28]. The specific dataset used includes a spatial dimension M is 196514, whereas the spectral dimension is 500 where the m/z bins range is between 703.57 – 872.56.

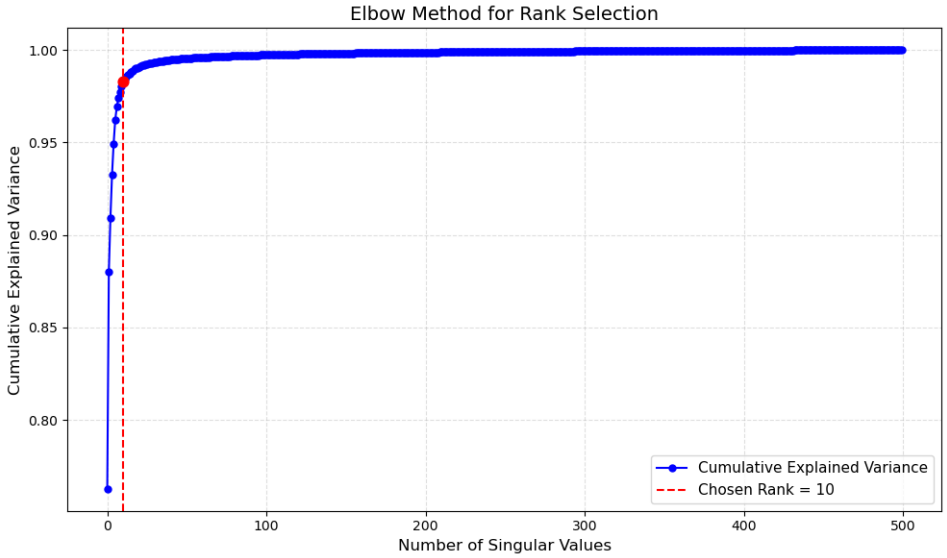


Figure 1-6: Rank Selection

The assumption of low rank is validated by Figure 1-6 above, where few components associated with the largest singular values account for most explained variance. As indicated by the red dotted line, rank 10 is chosen as the fixed rank K using the elbow method. 98% of variance in the data from the first 10 components.

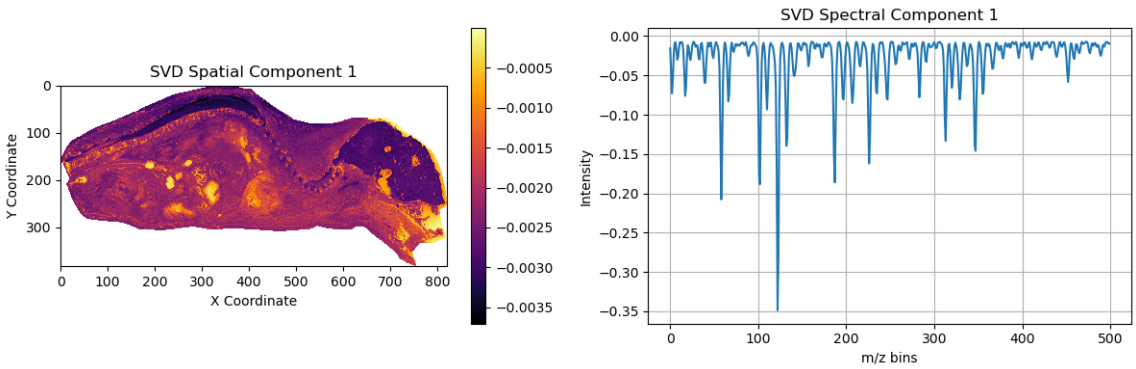


Figure 1-7: SVD on IMS dataset

Unfortunately, despite its optimal rank-K factorization, Singular Value Decomposition does not produce interpretable results. As evident in the spatial component associated with the largest singular value, the spatial abundances in Figure 1-7 are not interpretable. The spectral profile further produces negative peaks which do not carry physical meaning.

Therefore, an alternative factorization method that imposes entry-wise nonnegativity upon its factor matrices is explored. The factorization method is also known as Nonnegative Matrix Factorization (NMF).

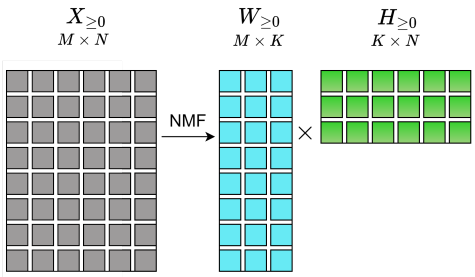


Figure 1-8: Standard NMF Visualization

$$\min_{W \geq 0, H \geq 0} \underbrace{\|X - WH^T\|_F^2}_{f(W,H)} \tag{1-4}$$

where $W \in \mathbb{R}^{M \times K}$ is the *Spatial Abundance Matrix*, and $H \in \mathbb{R}^{N \times K}$ is the *Spatial Signature* matrix. NMF is visualized in Figure 1-8, where an input data matrix X is factorized into a product of entry-wise nonnegative matrices W and H.

The optimization problem 1-4 is nonconvex [29]; however, minimizing over W, or H individually, is a convex sub-problem. In fact, solving simultaneously for both optimizers is an NP-hard problem [30]; hence update methods focus only on efficient ways of computing local optima [29]. As the sub-problems for computing two blocks, W and H, are convex, most algorithms approach the NMF problem via Block Coordinate Descent (BCD), which optimizes alternatively over the two variables [31]. Some common BCD methods include Multiplicative update rule [32], Hierarchical Alternating Least Squares, and Alternating Nonnegative Least Squares [33].

In the example below, one may observe the enhanced interpretability due to NMF factorization using HALS update rule. The dominant component highlights a clear anatomical structure in the spatial map, whilst the spectral profile exhibits a few select well-separated sharp peaks that are well contrasted in relation to the rest of the profile.

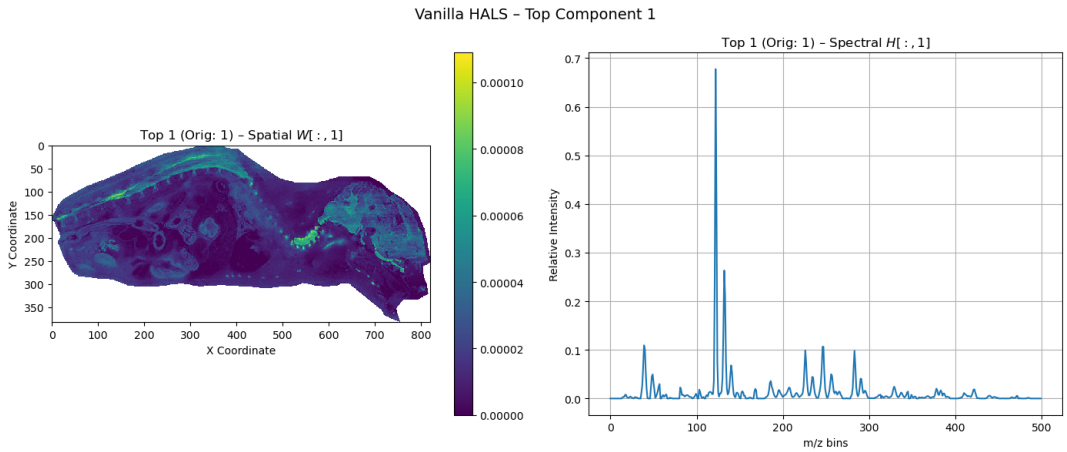


Figure 1-9: NMF on IMS dataset

Similarly, for extremely large data sizes common in Imaging Mass Spectrometry, conventional NMF update methods are not sufficiently scalable. Recent works on Randomized NMF have aimed to emulate Halko et al.'s strategy for Randomized SVD [27][2][34]. To facilitate this result, one can introduce how the pipeline from Randomized SVD translates into the randomized pipeline for NMF.

1-2 Introduction to Randomized NMF

Using the same sketching strategy as in Figure 1-3, Randomized NMF methods seek to employ the orthonormal basis Q from QR decomposition to perform NMF on B:

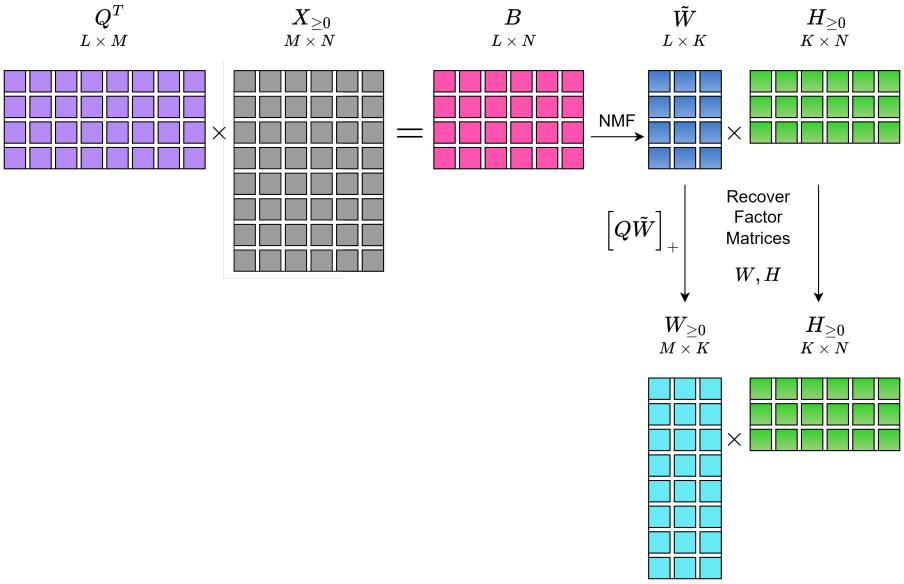


Figure 1-10: Solving Low-Dim NMF and Recovering Factors

A compressed version of the input matrix $B = Q^T X$ and a compressed version of the Spatial Abundance Matrix $\tilde{W} = Q^T X \in \mathbb{R}^{L \times L}$ are thus used within NMF update rules, resulting in the following lower-dimensional formulation:

$$\min_{Q\tilde{W} \geq 0, H \geq 0} \underbrace{\|Q^T X - Q^T W H^T\|_F^2}_{f(W,H)} \tag{1-5}$$

However, this thesis challenges this formulation due to lack of nonnegativity constraints on the basis Q. As a result, B and $\tilde{W}$ relax the nonnegativity constraint in the fashion of Semi-Nonnegative Matrix Factorization. In Figure 1-10, W is shown to be recovered in the form $[Q\tilde{W}]_+$ where $[\cdot]_+$ is an entry-wise projection to the nonnegative orthant. The proceeding chapter will thus further underpin the downside of the recovery in such form and explore other possible solutions.

1-3 Overview of the Thesis

Throughout this thesis, the main focus will concern the choice for the orthonormal basis Q.

Main Research Question: *In what ways can nonnegativity constraints be integrated into randomized SVD-inspired pipeline for NMF?*

1. Why do the existing approaches fail to accommodate nonnegativity constraints with a QR-based basis within the Randomized Framework for NMF?
2. Can alternative bases with a nonnegativity constraint be incorporated to alleviate the problems posed by the QR-based basis?

In order to answer the research question and its subquestions, this thesis outlines the following structure:

Chapter 2 revisits the role of an orthonormal basis Q within the framework and features the main computational properties leveraged by the choice of $Q = Q_{\text{qr}}$ from QR decomposition. Given that the QR-based basis is a unanimous choice for the compression matrix within existing Randomized NMF literature, this chapter demonstrates why its use is not sufficient within the framework, as well as the IMS context. In light of the first sub-question, multiple strategies to accommodate a nonnegativity constraint on $W = Q\tilde{W}$ will be discussed.

Chapter 3 presents the main exploration and research contribution within the thesis. It will involve an exploration of how alternative nonnegative bases can be integrated within the Randomized NMF pipeline to alleviate the problems introduced in Chapter 2. To the best knowledge of the author, this is the first work to apply a compression matrix other than a QR-based basis within the Randomized NMF pipeline.

Chapter 2

Approaches to Randomized Nonnegative Matrix Factorization

The problem of negative entries in projections to and from the lower-dimensional space will be further investigated and contextualized within the framework for Randomized Nonnegative Matrix Factorization. This chapter aims to highlight this problem as one of the key steps in the framework for state-of-the-art randomized algorithms and identify key properties of the QR-based basis. The said algorithms will be compared to baseline methods such as conventional and large-scale algorithms. On one hand, this will motivate the potential of randomized NMF algorithms and attribute its fast convergence to the lower-dimensional basis. On the other hand, this will facilitate the streamlining of the results during the exploration of solutions to the mixed-sign problem in the remainder part of the paper.

2-1 Comparison to Classical Approaches

Before delving into the main problem of entry-wise nonnegativity within the lower-dimensional formulation, it is important to baseline the state-of-art randomized algorithm. In order to accomplish this, vanilla versions of NMF update rules such as Multiplicative Update (MU) and Hierarchical Alternating Least Squares (HALS) are performed to factorize the IMS dataset. A Large Scale (LS) NMF algorithm further serves as a comparison of a Fast HALS learning rule via a block-wise strategy [35] [36].

The algorithms used for initial comparison are thus the following:

1. MU NMF from Sha et al. and Varasteh [37] [38]
2. HALS NMF from Cichocki [35]
3. Large Scale HALS from Cichocki et al. [35] [36]
4. Randomized HALS (RandHALS) from Erichson et al. [27]

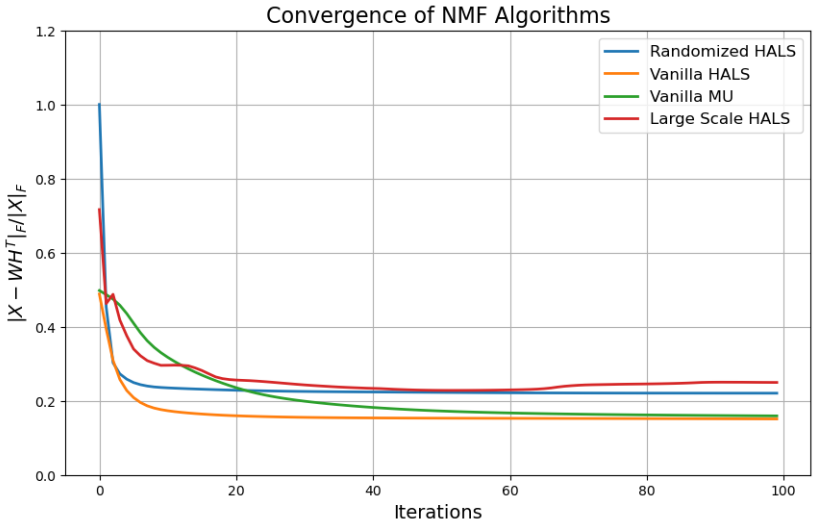


Figure 2-1: Factorization of Mouse Pup IMS data - Algorithm Comparison

Figure 2-1 presents the convergence of the four above-mentioned update rules over a hundred iterations. The vanilla algorithms do not employ any compression, whereas the Large Scale HALS applies a block-wise approach with sampling of rows and columns. The latter selects vectors or rows in the data matrix X based on a sampling method of choice [35]. As opposed to obliviously selecting columns, a more attuned sampling approach with leverage score-based sampling is applied. The update rule initializations for W and H are a modulus of the standard normal distribution $|\mathcal{N}(0, 1)|$.

Update Method	Sampling	Relative Frobenius Norm	Time (s)
Erichson RandHALS	Uniform	0.221	81.32
Cichocki LSHALS	Leverage Scores	0.250	248.30
Vanilla MU	N/A	0.153	292.52
Vanilla HALS	N/A	0.149	552.89

Table 2-1: Comparison of Update Methods against Approximation Time & Errors

Table 2-1 accompanies the plot in Figure 2-1 with the chosen sampling methods, the running time as well as the approximation errors. The relative Frobenius norm error is used as a metric for the factorization error computed via the equation below:

$$\frac{\|X - WH^T\|_F}{\|X\|_F} \tag{2-1}$$

The computation time includes the sampling time, as well as the time taken to execute the update rules. As a result, the Randomized HALS algorithms by Erichson et al. display roughly twice the speed of the conventional algorithms whilst exhibiting fast convergence behavior with low Frobenius errors. The above-mentioned relative errors from Table 2-1 may be further baselined against the optimal factorization result from Truncated SVD for rank 10, which is 0.13967.

For comparative purposes a leverage score sampling strategy was implemented in Large Scale HALS, which supersedes in approximation errors compared to random column subset selection sampling strategies. For Randomized NMF, however, one may use a data-oblivious sketching strategy with a JL-type of projection and still outperform Large Scale HALS in factorization accuracy and computational speed. For Randomized NMF

In the Erichson RandHALS algorithm, a uniformly distributed test matrix Ω Unif(0, 1) is used to sketch X to form a sampling matrix Y, which is then orthonormalized as previously mentioned in the introduction. The Large Scale HALS algorithm may yield a non-monotonous decrease as shown with the leverage score-based sampling example in Figure 2-1. Leverage scores attribute a probability to each row or column based on its importance towards spanning the range of an input matrix. The first step involves a computation of a Truncated SVD:

$$X \approx U_k \Sigma_k V_k^T \tag{2-2}$$

where U_k are top k left singular vectors, and V_k are top k right singular vectors. For row sampling $k = R$ number of rows, and for column sampling $k = C$ number of columns.

Leverage or "influence" score for each $i-$th row may be assigned via the following summation on top R left singular vectors are selected [39]:

$$\ell_i = \sum_{j=1}^{R} U_{ij}^2 \tag{2-3}$$

- U_{ij} is the (i, j)-th entry of U_k, the left singular vector.

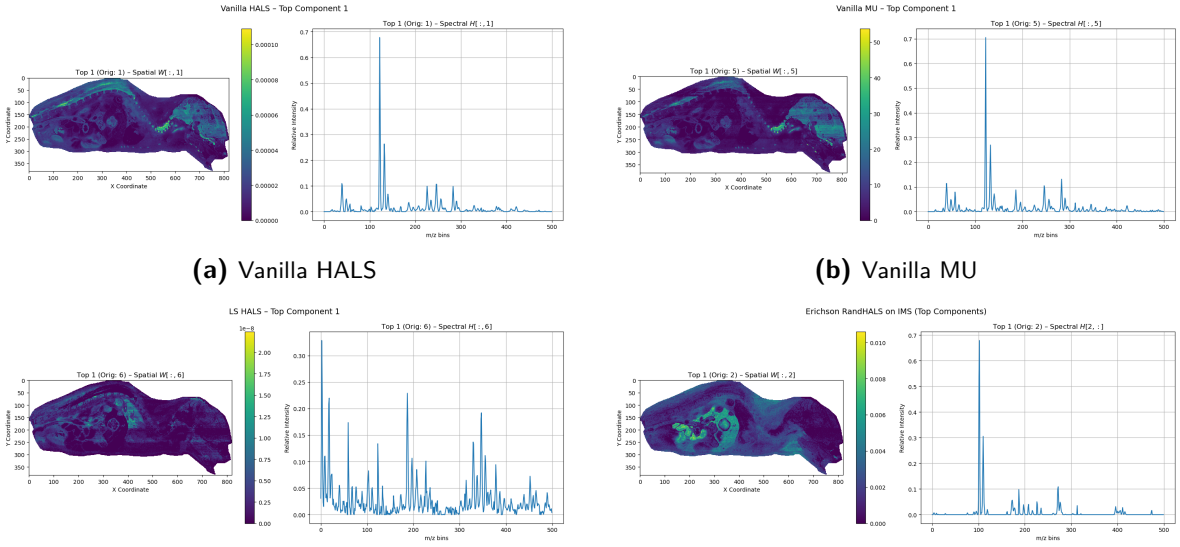
Thus the probability of selecting a row is as follows:

$$p_i = \frac{\ell_i}{\sum_{j=1}^{M} \ell_j} \tag{2-4}$$

Analogous steps can be performed for column sampling with top C right singular vectors of V. R selected rows and C columns are chosen as 30 according to the guideline below with respect to rank $K = 10$ factorization [35]:

$$K < (R = 30) \leq 4K, \quad K < (C = 30) \leq 4K \tag{2-5}$$

Error bounds on projections via column-subset selection have been shown in previous work [22]. Some of the existing approaches include leverage-based selection [39] and empirical interpolation approaches[40], however, optimal sub-matrix formation is likely to be performed with the former [22]. However, the column-subset selection requires a full SVD in $\Theta(MN^2)$ time for exact leverage score computation [41]. This is contrary to a mere sampling via data oblivious multiplication requiring only $LNT_{\text{rand}} + LT_{\text{mult}} + L^2M$ flops where T_{rand} is the cost of forming a random matrix, and T_{mult} is the multiplication cost [22].



(a) Vanilla HALS **(b)** Vanilla MU

(c) LS HALS **(d)** RandHALS

Figure 2-2: Component Visualizations based on the largest column norm of H.

Figure 2-2 highlights the top spatial and spectral components where the column index is selected by the index of the largest column norm of H. Due to variability in norms of columns of H, it is unnecessary for different update methods to highlight the same region in its most norm-dominant component and exhibit similar spectral profiles. However, for interpretability, beyond the nonnegativity of components, it is also important that spatial components are well-localized, and spectral profiles are well-separated with sharp, narrow peaks. A well-localized spatial component would display a concentration of positive intensities in a specific anatomical region, such as the brain or kidneys.

LS HALS, for example, exhibits a more noisy spectral profile in its dominant component with a dense spectrum and a high noise floor, unlike the other three methods. Similarly, the spatial map is faint, and the intensities appear diluted in different tissue parts. In contrast, the vanilla methods display narrow peaks in a consistent profile with few dominant peaks and a concentration of spatial abundances in the brain and spinal regions. However, some presence of noise is evident by the scattered patches in the brain region for vanilla MU and noise present near zero Y coordinate for vanilla HALS. RandHALS by Erichson et al. shows a well-contrasted spectral profile with biologically relevant localized parts such as the intestines region. The bottom part of the intenstines region seems to be diffused, as well as other parts close to the perimeter of the tissue.

Despite advantages and disadvantages in terms of interpretability for all of the above methods, the Erichson RandHALS update method suffices in its ability to map to anatomically relevant regions despite being partially diffused and noisy, which makes it suitable for initial dimensionality reduction. In the next sections, one may overview the framework as a concrete set of steps where the use of a QR-based basis will be further delved into, including its favorable and undesirable properties.

2-2 Framework for Randomized NMF

The state-of-the-art algorithm by Erichson et al. has been shown to demonstrate strong factorization performance to baseline algorithms. Whilst the update method exhibits fast convergence and low approximation errors, it is important to investigate the cause and impact of mixed-sign basis. In order to better understand the impact of the projection to lower-dimensional subspaces that are fully entry-wise nonnegative, it is important to understand how Randomized Nonnegative Matrix Factorization algorithms operate in retrospect. For visual purposes, the two-stage framework summarizes the main steps in state-of-the-art Randomized NMF algorithms in Figure 2-3 below:

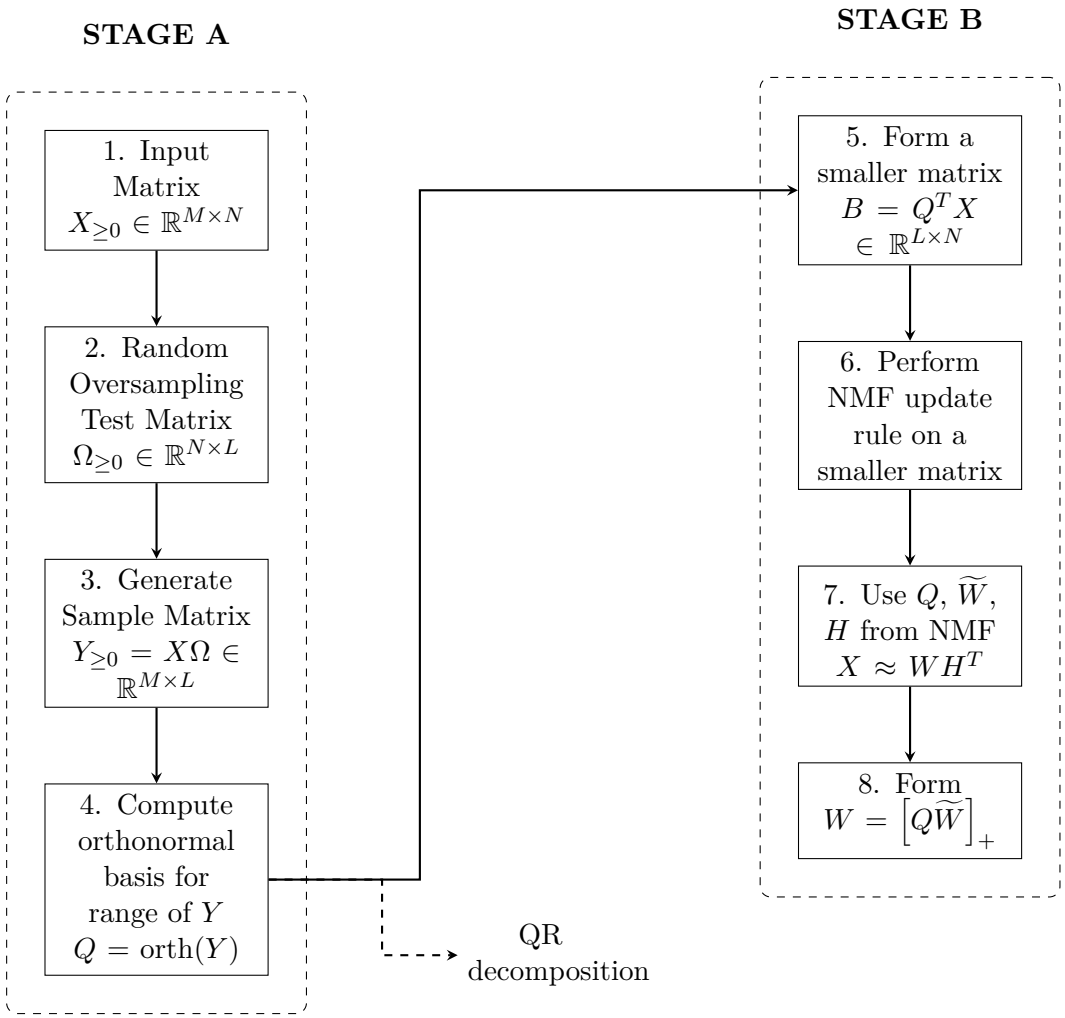


Figure 2-3: Randomized NMF Framework (Visualized based on [22][27])

The conventional higher-dimensional NMF problem in 1-4 can be solved as a lower-dimensional problem in 1-5 by following Stages A and B. The framework visualized in Figure 2-3 is based on the Randomized SVD two-stage approach from Halko et al. [22] and later adapted by Erichson et al. for Randomized NMF [27].

Stage A concerns the sketching process for the input matrix X. In steps 2 and 3, random projections are applied via multiplication of X by a randomly distributed test matrix Ω. The choice for the test matrix in line with the framework will be a Uniform distribution between 0 and 1, similarly used in the literature [27] [2].

The probability of capturing K-most dominant singular vectors of X increases via oversampling by an additional P columns. Therefore, for $K = 10$, the sampling dimension L is chosen to be 12 with an oversampling of $P = 2$ additional columns.

In step 4, the generated sample matrix Y is orthonormalized via QR decomposition to yield an orthonormal basis to the range space of Y. The orthonormal basis Q is then used to compress matrix X by sketching the columns of X via multiplication $Q^T X$ to generate a small surrogate input matrix X.

This step is a crucial step to impose a discriminative structure on X. The $Q^T X$ is essentially a lower-dimensional whitened version of matrix X. Alternatively, this may be referred to as a data-adapted sketch of X [42]. In general, the JL-based projections can also be referred to as data-oblivious sketches, which aim to form a sample matrix Y [42].

Subsequently, the small surrogate input matrix B is then used as an input to an NMF update rule of choice, such as Hierarchical Alternating Least Squares by Erichson et al. The obtained lower-dimensional factor matrix $\tilde{W}$ is projected back to the higher-dimensional space via the following relation $Q\tilde{W} \approx W$.

Consequently, the two-stage framework echoes most of the steps employed in the two-stage process for Randomized SVD. The key difference in existing algorithms lies in step 6, where the factorization method of choice, or NMF, is applied. However, state-of-the-art randomized algorithms fail to guarantee the nonnegativity of intermediate matrices B, $\tilde{W}$, and in some cases of the output of $W \approx Q\tilde{W}$. The Nonnegativity of H for one-sided column sketching is not a concern, however an analogous problem would be thus relevant if a row sketching is applied on the matrix X. Due to the Johnson-Lindenstrauss-based projections being one-way and non-reversible [43], orthonormalization is a crucial step to project to and from the lower-dimensional space.

$$\begin{aligned} B_{\not\geq 0} &= Q^T X \text{ surrogate input matrix} \\ \tilde{W}_{\not\geq 0} &= Q^T W \text{ compressed left factor matrix} \end{aligned} \tag{2-6}$$

QR decomposition can be further parallelized into level-3 BLAS operations, scales well and makes for an extremely rapid computational framework. However, the problem of negative entries may invoke a question on whether the approximation errors of the said method compare well.

State-of-the-art Randomized NMF algorithms compare well against other baselines. Randomized NMF has some similar assumptions to separability. For instance, Cichocki implements a block-wise approach assuming that not the whole matrix is used. Randomly chosen blocks can be used to approximate.

2-3 Importance of the Orthonormal Projection

2-3-1 Purpose and Role within the Framework

We have previously introduced the Q from QR decomposition. In this section, we would like to continue to reiterate how the properties of an orthonormal matrix are favorable within the framework.

Generating $Y \in \mathbb{R}^{M \times L}$ results in a tall-skinny matrix. In layman's term, Y can be described a data-oblivious sketch of X using a JL-type of projection. Unfortunately, the sample matrix Y will likely be ill-conditioned [22]. The orthonormalization step is an intermediary step that identifies the numerical rank of the sample matrix and discards excess samples [22].

$$\begin{aligned} \mathrm{qr}(Y) &= QR \\ \text{where } Y &\in \mathbb{R}^{M \times L} \\ Q \in \mathbb{R}^{M \times M}, R &\in \mathbb{R}^{M \times L} \end{aligned} \tag{2-7}$$

The dimension M is much larger than the sampling dimension L; hence the QR decomposition of Y can be partitioned in the following manner:

$$\begin{aligned} A &= \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \\ &\textit{for } M \geq N \\ &Q_1 \in \mathbb{R}^{M \times L}, Q_2 \in \mathbb{R}^{M \times (M-L)} \\ &R_1 \in \mathbb{R}^{L \times L}, R_2 \in \mathbb{R}^{L \times M-L} \end{aligned} \tag{2-8}$$

Therefore, the last $M - L$ rows of R do not provide additional information, leading to the following form:

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \quad = Q_1 R_1 \tag{2-9}$$

Hence, the thin or economic QR can efficiently retrieve $Q_1 \in \mathbb{R}^{M \times L}$. For this chapter, the reference to the orthonormal basis Q from QR decomposition refers to $M \times L$ matrix that is retrieved from economic QR decomposition of sample matrix Y.

The columns of Q form an orthonormal set with respect to the Hermitian inner product, and preserve the Euclidean geometry for any vector x such that [22]:

$$\|Qx\| = \|x\| \tag{2-10}$$

Traditionally, this basis may also be referred to as a rotation matrix, or more precisely a linear transformation matrix. The orthogonality property allows the circumventing of a computationally heavy inverse operation by a mere transpose. In the context of the randomized NMF framework, an orthonormal Q allows to rotate to and from the lower-dimensional space without inverting the matrix.

$$Q^T = Q^{-1} \tag{2-11}$$

An orthonormal matrix Q can be used to form an orthonormal projector $P_Y = QQ^T$ is an idempotent, Hermitian operator such that [44] [22]:

$$P^2 = P, P = P^{-\dagger} \tag{2-12}$$

where the $\dagger$ is the pseudo-inverse.

The projection to the lower-dimensional subspace has the favorable property of maintaining the energy norm of a matrix as follows:

$$\|QQ^T X\|_F^2 \approx \|X\|_F^2 \tag{2-13}$$

Where an orthogonal square projector QQ^T is a unitary real matrix under which the operator norm and Frobenius norms remain unitarily invariant. Given the suitable choice of fixed rank K and oversampling P, QR decomposition on Y yields near-optimal normal basis such that [27]:

$$X \approx Q \underbrace{Q^T X}_{B} \tag{2-14}$$

It is possible to further enhance the accuracy in stage A via the use of power/subspace iterations and two-sided sampling. However, due to the pronounced singular value spectrum shown in Figure 1-6, the power rule will overemphasize the influence of the component associated with the largest singular value. In practical circumstances, the size of the sketch dimension $L = 12$ will be sufficient in terms of factorization performance as well as the interpretability of factors. Due to the focus on the nonnegativity constraint for the basis for one-dimensional compression, two-sided compression is not discussed within the scope of this thesis.

The QR-based basis Q_{qr} is visualized below via heatmaps indicating negligible orthogonality residual and a mixed-sign dense basis. 54.34% of the entries for the given test matrix.

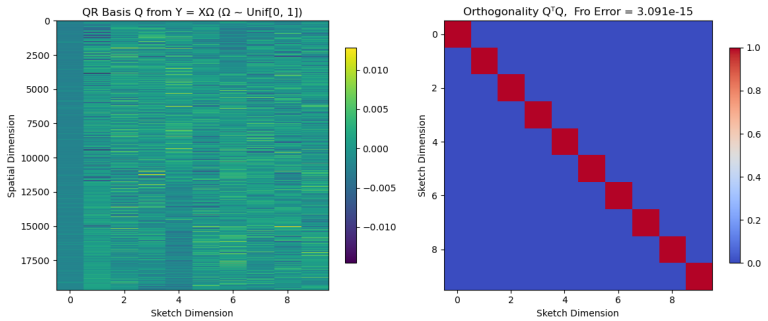


Figure 2-4: Heatmaps of Q_{qr} and its Orthogonality $Q^T Q$ given a test matrix $\Omega \sim \text{Unif}(0, 1)$.

2-4 Variants of RandHALS to accommodate Nonnegativity Constraints

Existing Randomized Algorithms in line with the randomized framework use the orthonormal basis from QR decomposition on the sample matrix as the main and only method to extract a basis. More recent works have explicitly recognized the importance of mitigating negative entries from the QR-based basis [2] [34] [42].

This section will thus review three update methods:

- Erichson Randomized HALS (RandHALS)
- Masuda RandHALS
- Chaudhry Randomized MU (RandMU)

Erichson et al. highlight that lower-dimensional factor $\tilde{W}$ is not necessarily nonnegative throughout iterations, however, a projection to nonnegative orthant can assure a nonnegative higher-dimensional factor W at each iteration as follows:

$$W = \left[Q\tilde{W}\right]_+ \tag{2-15}$$

where $[\cdot]_+ = max(\cdot, 0)$ is an entry-wise projection to nonnegative orthant

Simultaneously, at the end of each iteration, the basis is used to project W to lower-dimensional $\tilde{W}$:

$$\tilde{W} = Q^T \underbrace{\left[Q\tilde{W}\right]_+}_{W_{\geq 0}} \tag{2-16}$$

$\tilde{W}$ is not necessarily nonnegative through the iterations [27]. Masuda et al. have highlighted the the number of negative entries, as well as their large absolute values in $Q\tilde{W}$ term [2]. In order to alleviate this issue, an alternative objective function has been proposed where $L_{1,1}$ norm of H term is regularized, and the $Q\tilde{W}$ constraint is rewritten as follows:

$$-\epsilon 1_{M\times K} \leq Q\tilde{W} \leq U 1_{M\times K} \tag{2-17}$$

where 1 is a matrix of ones, ϵ is a small positive tolerance, and U is an upper bound. The approach towards the satisfaction of this constraint via an entry-wise bounding where each entry of $\tilde{w}$ is bounded between $\tilde{w}_{\min}$ and $\tilde{w}_{\max}$.

$$\tilde{w}_{k',k} \leftarrow P_{[\tilde{w}_{\min},\tilde{w}_{\max}]}\left(\frac{r_{k'k}^T h_k + \delta\tilde{w}_{k',k}}{\|h_k\|_2^2 + \delta}\right) \tag{2-18}$$

where $r_{k'k}$ is an entry of a lower-dimensional residual matrix, and δ is a small constant.

The feasible range for each element $\tilde{w}_{k'k}$ entry is computed as:

$$\tilde{w}_{\min} = \max\left(-\epsilon + \frac{\sum_{\ell \neq k'}^{K} q_{m\ell}\tilde{w}_{\ell k}}{q_{mk'}}, U - \frac{\sum_{\ell \neq k'}^{K} q_{m\ell}\tilde{w}_{\ell k}}{q_{mk'}}\right) \tag{2-19}$$

The maximum of the two entries is chosen as $\tilde{w}_{\min}$, and the first entry in equation 2-4 is used when $q_{mk'}$ is positive, and the second entry when $q_{mk'}$ is negative.

$$\tilde{w}_{\max} = \min\left(-\epsilon + \frac{\sum_{\ell \neq k'}^{K} q_{m\ell}\tilde{w}_{\ell k}}{q_{mk'}}, U - \frac{\sum_{\ell \neq k'}^{K} q_{m\ell}\tilde{w}_{\ell k}}{q_{mk'}}\right) \tag{2-20}$$

Similarly, for $\tilde{w}_{\max}$ in equation 2-20, the first entry is for negative $q_{mk'}$ whilst the second is for positive $q_{mk'}$.

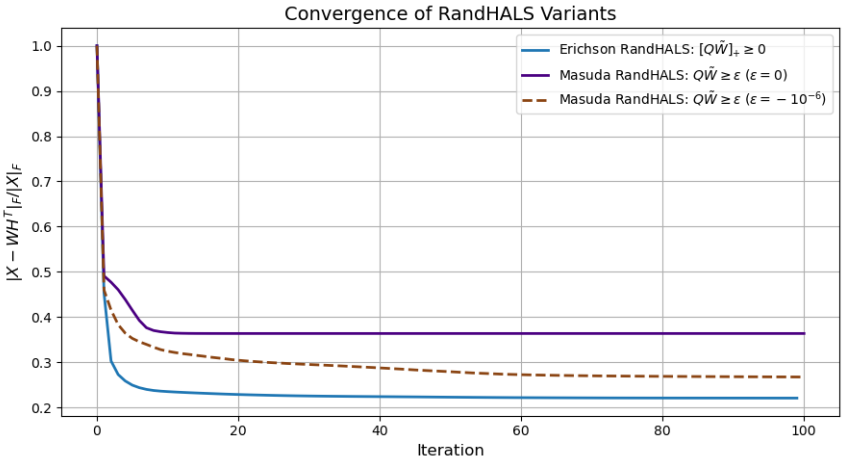


Figure 2-5: Rank 10 Factorization of Mouse Pup IMS data using Randomized HALS Algorithms

The bounding approach satisfies the relaxed entry-wise constraint on $Q\tilde{W}$ from equation 2-17. The tolerance is chosen as $\epsilon = -10^{-6}$ and the upper bound is $U = 10^9$. Figure 2-5 displays the relative Frobenius norm error approximations for rank 10 according to equation 2-1. The bounding approach yields a slightly higher approximation error; however, it is subject to a slow entry-by-entry update rule.

Whilst this approach reduces the magnitude of negative entries to a small tolerance of ϵ, it is interesting to highlight that allowing a small negative tolerance may have poor implications on the interpretability of IMS data.

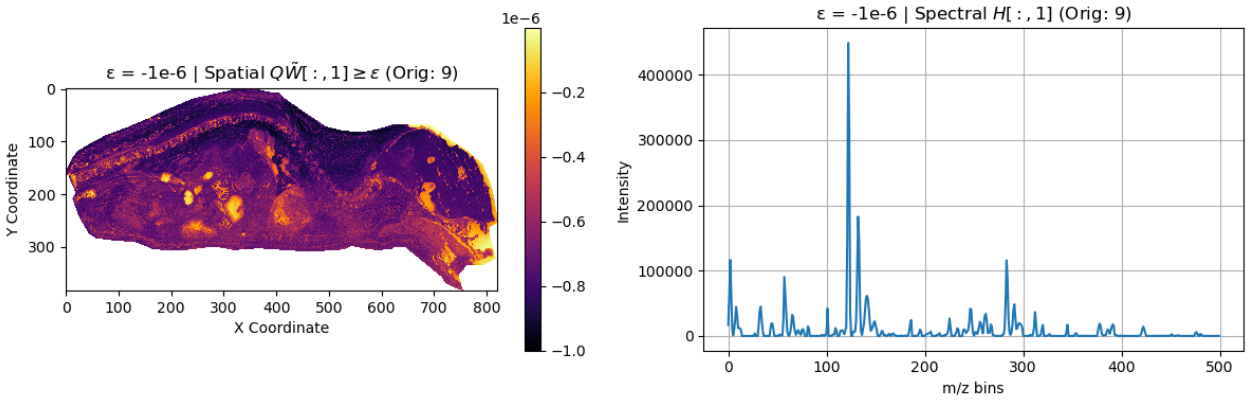
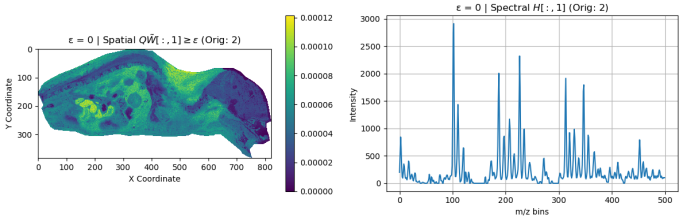
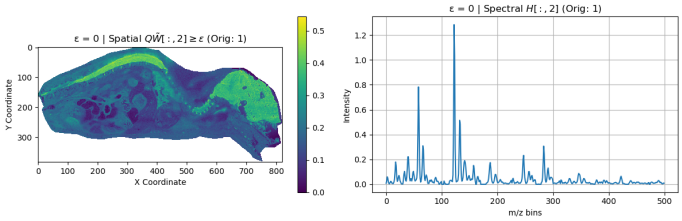


Figure 2-6: Component Visualization (Relaxed Nonnegativity Constraint) - Masuda RandHALS

Figure 2-6 demonstrates that on the largest column associated with the spectral component norm $\|H_k\|$, the corresponding spatial map $W \approx Q\tilde{W}$ includes small negative values. Whilst for some applications this may be useful, for interpretability purposes in Imaging Mass Spectrometry a relaxed nonnegativity constraint is therefore not preferred. Modifying the tolerance ϵ to a fully entry-wise nonnegative bounded constraint displays favorable results in the most dominant spectral component.



(a) Top Component 1



(b) Top Component 2

Figure 2-7: Component Visualizations - Masuda RandHALS ($\epsilon = 0$)

The brain and spinal parts of the mouse pup tissue are well-contrasted in comparison, thus vastly improving interpretability. However, the other 8 components from the rank 10 factorization are fully entry-wise zero not providing additional information. The relaxed nonnegativity constraint counterpart does not face this issue, but the first three most dominant components include negative spatial values.

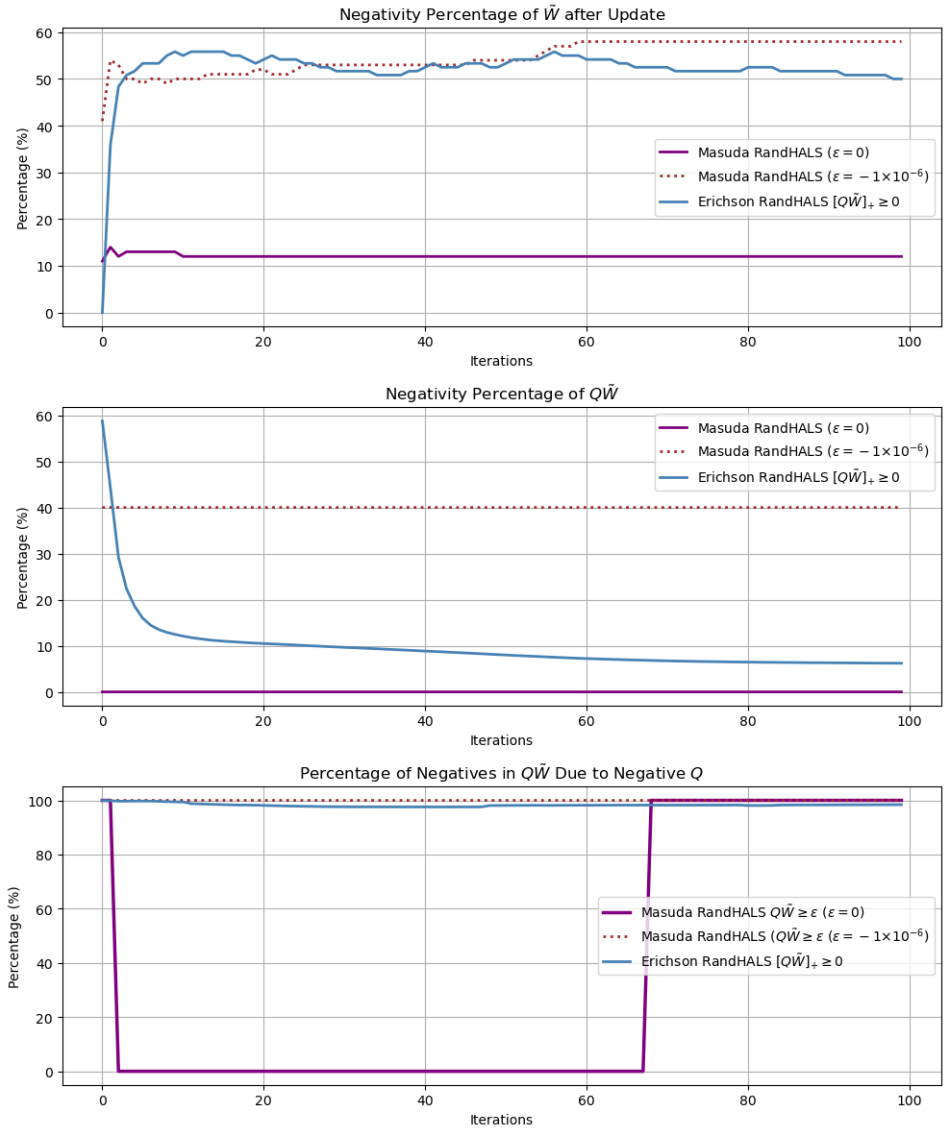


Figure 2-8: Percentage of Nonnegative Entries in $\tilde{W}$, $Q^T\tilde{W}$ during Iterations - RandHALS Methods

Furthermore, one may obtain further insight into how the QR-based basis affects the nonnegativity of the factors $W = Q\tilde{W}$ and $\tilde{W}$ within iterations in Figure 2-8. For the case of zero tolerance, Masuda et al. successfully retain full entry-wise nonnegativity of $Q\tilde{W}$ throughout all iterations as seen in the middle plot 2-8. The Erichson et al. algorithm manages to continuously increase the percentage of entry-wise nonnegative entries to around 92% due to continuous projections to the nonnegative orthant, which is partially attributed to increased sparsity in W as a result of entry-wise clipping.

The percentage of negative entries due to Q being negative is calculated via the formula below:

$$\frac{\text{Total Number of Entries}(Q_{-}\tilde{W}_{\geq 0})}{\text{Total Number of Entries}(Q\tilde{W})} \tag{2-21}$$

where $Q_{-}\tilde{W}_{\geq 0}$ indicates the number of entries $Q_{m,l}\tilde{W}_{l,l} < 0$ entries such that $Q_{m,l} < 0$ and $\tilde{W}_{l,l} \geq 0$.

Almost a hundred percent of entry-wise negativity in $Q\tilde{W}$ is caused by the basis matrix Q, which is a concerning result. This is consistent with the result in Randomized HALS implementation by Masuda et al. where bounding only slightly improves this result. A projection designed explicitly to preserve Euclidean geometry in lower-dimensional space is no longer fully maintained each each iteration step, with the largest projection loss taking place during the first few iterations.

It is important to emphasize that Erichson et al. assure an entry-wise nonnegative spatial factor W by projecting $Q\tilde{W}$ to the nonnegative orthant as such:

$$\left[Q\tilde{W}\right]_{+} \approx W \tag{2-22}$$

Thus, the plots demonstrate the nonnegativity percentage right before this step. However, given the above result, it is no longer possible to guarantee that the projection to Nonnegative orthant step does not eliminate significant information during the projection to higher-dimensional space. Additionally, the iterative update rules are applied in Semi-NMF fashion where the surrogate input matrix is no longer entry-wise nonnegative, and the nonnegativity constraint on lower-dimensional factor $\tilde{W}$ is relaxed.

The update rule by Erichson et al., beyond its speed and low factorization errors, is more indicative of Low-rank Nonnegative Matrix Approximation. That is, a computational framework that admits negative entries in intermediate low-rank matrices, however, converges to a low-rank nonnegative matrix [45]. Alternating projections between a low-rank manifold and nonnegative orthant have been shown to be effective in the context of Low-rank Nonnegative Matrix Approximation, where a better low-rank approximation on the input matrix can facilitate a lower approximation error [46].

Additionally, one may be alarmed that this may violate the geometric property of NMF, which mainly involves finding a cone involving all the data points in the positive orthant [47]. Thus, the negative entries caused by the mixed-sign basis may cause the data points to no longer lie in the orthant. Different work by Yahaya et al. refers to matrices with unconstrained basis matrix as instances of Randomized Semi-NMF algorithms in their Randomized NeNMF implementation [48].

It is important to note that the use of the QR-based basis is standard within the current state-of-the-art Randomized (semi-)NMF algorithms [27] [2] [2] [49][48]. For example, Yahaya et al. NMF directly refers to an analogous approach as semi-Nonnegative Matrix Factorization.

2-5 Correctivity and Regularization on the Projection Matrix

A recent paper by Chaudhry et al. somewhat deviates from the SVD-inspired pipeline and instead focuses on shifting the projection matrix to become nonnegative, directly dealing with an input matrix $QQ^TX + correction$ [42]:

$$(QQ^T + \sigma_{correction} 1_M 1_M^T)(X - WH^T) \tag{2-23}$$

where $\sigma_{correction} = abs(min(QQ^T)))$ is the absolute value of the most negative entry in the projector $QQ^T \in \mathbb{R}^{M \times M}$ and 1_M is a vector of ones [42].

Unfortunately, given the dimension M being 196514 for the Mouse Pup dataset, computing the minimum entry is a significant computational bottleneck for a computer with limited power. Therefore, based on a sample of 10,000 vector products from the QQ^T matrix, the penalty term has been slightly overestimated$\sigma_{correction} \geq \sigma_{min}$. The term is thus used for one-sided data-adapted compressed multiplicative update rule to utilize QQ^TX instead of an original input X. A regularization parameter $\lambda = 0.1$ is applied on the complementary space of X to mitigate the influence of sub-optimal shift on the initially orthonormal basis [42]:

$$W \leftarrow W \circ \frac{QQ^TXH + \sigma \mathbf{1}\mathbf{1}^TXH}{(1-\lambda)QQ^TWH^TH + \sigma \mathbf{1}\mathbf{1}^TWH^TH + \lambda WH^TH} \tag{2-24}$$

$$H \leftarrow H \circ \frac{X^TQQ^TW + \sigma X^T\mathbf{1}\mathbf{1}^TW}{(1-\lambda)HW^TQQ^TW + \sigma HW^T\mathbf{1}\mathbf{1}^TW + \lambda HW^TW}. \tag{2-25}$$

The update rule shown in equations 2-24 and 2-25 recovers a good solution given that the original NMF error and the norm of the complementary space is small [42]. Additionally, the σ correction term successfully assures a fully entry-wise nonnegative solution across all iterations. Given the growing data sizes of Imaging Mass Spectrometry datasets, the update rule does not require storing the entire data matrix and accessing only once or twice [42]. However, the key difference with the algorithm is that it does not use an update rule directly on small surrogate input matrix $B = Q^TX$ however, instead opting to use the full QB factorization of X in the form QQ^TX. Aside from this deviation from the randomized framework, another difference is the much slower convergence behavior of the shifted projection approach:

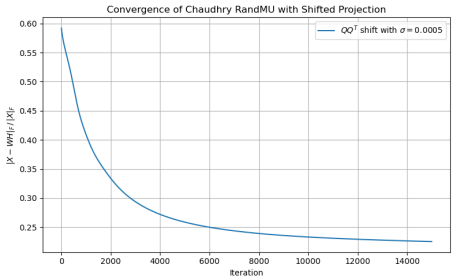
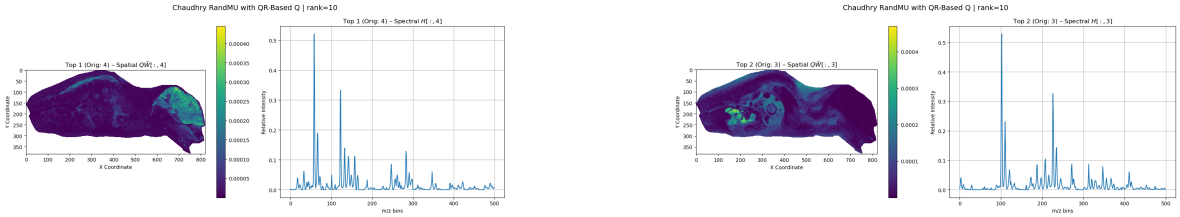


Figure 2-9: Convergence of Randomized MU Algorithm with Shifted Projection



(a) Dominant Component 1 **(b)** Dominant Component 2

Figure 2-10: Component visualizations from Chaudhry RandMU with QR-based projection and 15,000 iterations. Highlights dominant structures in IMS with spatial and spectral clarity.

The convergence behavior is extremely slow and converges to a relative error slightly above Erichson RandHALS. As shown in Table 2-2, Masuda variants of RandHALS are also much slower due to the entry-by-entry update rule. The shifted nonnegative projector QQ^T initial promising signs with the desired extraction in spatial and spectral components. However, the use of compression also conflicts within the framework beyond its use of QQ^TX. The Chaudhry RandMU method no longer benefits from fast convergence properties that are displayed in other update rules with the use of Q_{qr}.

Figure 2-9 visualizes the convergence of Randomized MU by Chaudhry et al. where the algorithm takes over 15 000 iterations to converge, unlike only 100 iterations for the state-of-the-art randomized NMF algorithms. The shift of 0.0005 to the orthonormal basis in this case is sufficiently influential to no longer maintain the fast convergence properties and likely disturb the orthogonality structure of Q. The algorithm is still a strong example of an update rule that learns NMF from compressed data. The performance of the algorithms which explicitly employ a method to post-process an orthonormal basis from QR decomposition are thus summarized below:

Update Method	**Iterations**	**Relative Frobenius Norm**	**Time (s)**
Erichson RandHALS	100	0.221	81.32
Masuda RandHALS, $\epsilon = 0$	100	0.365	8866.72
Masuda RandHALS, $\epsilon = -10^{-6}$	100	0.267	8884.76
Chaudhry RandMU	15 000	0.225	16465.68

Table 2-2: Comparison of methods mitigating negative entries in Q

The Table 2-2 displays the factorization results with the relaxed nonnegativity version of the Masuda RandHALS algorithm. For strictly nonnegative constraint on $Q\tilde{W}$, the algorithm's relative Frobenius norm error increases to 0.365. As a consequence in their original formulations, the algorithms can reach low approximation errors, however they fail to simultaneously address nonnegativity and orthonormality of the basis or the projector.

Chapter 3

Exploration of Nonnegative Bases

In contrast to the previous chapter, this chapter explores alternative bases for the QR-based basis. Namely, this part aims to demonstrate new results within update methods upon imposing an additional nonnegativity constraint on the orthonormal basis Q. For this reason, the traditionally-used QR-based basis Q_{qr} in two update methods: *Erichson RandHALS* and *Chaudhry RandMU* will be replaced by alternative bases obtained from Projective NMF (PNMF), Approximately Orthogonal NMF (AONMF) and a Proposed Method.

The discussed alternatives to QR-based basis include bases Q_{PNMF}, Q_{AONMF}, and Q_{prop}, which will be retrieved with varying degree of orthogonality, initialization, and even inputs. For example, PNMF and AONMF bases will be extracted from sample matrix Y and input data matrix X. The proposed method is designed to approximate the subspace on a QR-based basis. Extraction of the nonnegative approximately orthonormal bases from sample matrix Y will attempt to directly replace the orthonormalization step in step 4 of the Randomized Framework. This chapter will demonstrate which cases are insufficient and provide baseline results against basis extraction directly from X. Therefore, for PNMF and AONMF, this chapter presents the results of two experiments: Experiment A and Experiment B.

This part of the thesis highlights the successes and challenges associated with nonnegative basis extraction from orthogonal variants of NMF, such as PNMF and AONMF. As a logical progression to the first two experiments, a third experiment with a proposed methodology to compute a basis $Q_{\text{prop}} \geq 0$. The proposed method is mainly motivated by the incompatibility of the first two methods within the Erichson RandHALS update method. The need for the method stems from its ability to control the tradeoff between orthogonality and fit on the data.

The summary of experiments A, B and C is shown in Table 3-1 below:

Experiment	Input	Extracted Basis	Update Rules	Description
A	Y	$Q_{\text{PNMF-Y}}$, $Q_{\text{AONMF-Y}}$	Erichson RandHALS, Chaudhry RandMU	Nonnegative bases are extracted from the sample matrix $Y = X\Omega$, where $\Omega \sim \mathcal{U}(0,1)$. A given nonnegative basis $Q_{\geq 0}$ is applied within the update rules instead of Q_{qr}
B	X	$Q_{\text{PNMF-X}}$, $Q_{\text{AONMF-X}}$	Erichson RandHALS, Chaudhry RandMU	Nonnegative bases are extracted from the input data matrix X. A given nonnegative basis $Q_{\geq 0}$ is applied within the update rules instead of Q_{qr}
C	Q_{qr}	Q_{prop}	Erichson RandHALS, Chaudhry RandMU	Proposed Basis approximates the subspace spanned by Q_{qr} via optimization of orthogonality, alignment, and angle similarity terms. Similarly, the basis is applied within update rules.

Table 3-1: Summary of Experiments

The table emphasizes that *Experiment A* seeks to extract a nonnegative approximately orthonormal basis from orthogonal variants of NMF. Experiment A is designed to unveil whether Step 4 in Randomized Framework in Figure 2-3 can be directly replaced by an alternative orthonormalization method of Y.

The design of *Experiment B* circumvents the entire Stage A pipeline from Figure 2-3 and directly computes the basis from X. Throughout the next sections, the inefficiency of this approach will be demonstrated, however this experiment will facilitate and baseline the results for Experiment A and C.

As the last experiment, *Experiment C* will propose an additional step to the Randomized Framework, namely an intermediary step between steps 4 and 5. Unlike the other experiments, instead of relying on X or Y, this method will directly compute a nonnegative approximation of the QR-based basis. This experiment will demonstrate the flexibility of the method with various sets of hyperparameters for μ (orthogonality), λ alignment, and β principal angles.

Table 3-2: Hyperparameter Configurations - Proposed Method

Set	μ	λ	β	**Description**
1	0.5	1.5	2.0	alignment and angle emphasis
2	1.0	0.1	0.1	orthogonality dominant
3	0.1	2.0	0.1	alignment dominant
4	0.1	0.1	5.0	principal angle dominant
5	0.0	2.0	0.0	alignment only
6	2.0	0.0	0.0	orthogonality only
7	0.0	0.0	3.0	principal angle only
8	1.0	1.0	1.0	balanced configuration

Table 3-2 outlines eight different sets of hyperparameters for Experiment C with a mix of balanced and less balanced configurations as specified in the description. The AONMF method also allows the control of the orthogonality penalty μ; thus, it will also be varied between 0.1, 0.5, 1, 2, and 5. Subsequently, for all experiments, different initializations of $Q_{\geq 0}$ will be utilized including:

- Modulus of Standard Normal Distribution - $|\mathcal{N}(0,1)|$
- Uniform Distribution between 0 and 1 - $|Unif(0,1)|$
- Uniform Distribution between 0 and 0.04 - $|Unif(0,0.04)|$
- Uniform Distribution between 1 and 2 - $|Unif(1,2)|$
- Modulus of Standard Normal Distribution with offset - $|\mathcal{N}(0,1)|+0.1)$

Since all the involved objective functions are nonconvex, it is important to employ different initializations for $Q_{\geq 0}$. Q_{PNMF}, Q_{AONMF} and Q_{prop} will thus adopt the above initializations. In the context of this thesis, the combination of orthogonality and nonnegativity constraints renders the explored $Q_{\geq 0}$ bases to become sparse [50]. The initializations are thus fully entry-wise nonnegative, and some initializations assure that Q is not initialized close to its orthogonal, and namely not on the Stiefel manifold. For instance, for PNMF, the multiplicative update rule would work poorly on a matrix that is initialized to be orthogonal, as the update rule cannot recover entries that have been rendered zero [51].

3-1 Nonnegative Orthogonal Basis Optimization

As an introduction to methods that seek to extract a nonnegative (approximately) orthonormal basis, such as PNMF and AONMF, this section underpins the theoretical implications of imposing an additional constraint on entry-wise nonnegativity of Q.

$Q_{\geq 0} \in \mathbb{R}^{M\times L}$ is now a nonnegative (approximately) orthonormal basis of rank less than or equal to L such that $Q^TQ \approx I$. Given this formulation as a starting point, the chapter will focus on other possible formulations to facilitate the computational of a nonnegative orthonormal basis $\tilde{Q}$.

Optimization with orthogonality and nonnegativity constrained is commonly known as optimization on (plus-sign) sign-constrained Stiefel manifold. First, the Stiefel manifold without sign constraints is defined as a set of orthogonal matrices J such that $J^TJ = I$ or:

$$\mathbb{S}^{N,L} := \{J \in \mathbb{R}^{N\times L} : J^TJ = I \in \mathbb{R}^{L\times L}\} \tag{3-1}$$

The sign-constrained, or namely nonnegatively constrained Stiefel manifold is thus defined as [52]:

$$\mathbb{S}_+^{N,L} := \{J \in \mathbb{S}^{N,L} : J \geq 0\} \tag{3-2}$$

The distance of any matrix $Z \in \mathbb{R}^{M\times L}$ to a nonnegative Stiefel manifold can be bounded based on the norm of entries that violate the sign constraints and/or the norm of the orthogonality residual with positive constant factors ν and q as follows [52]:

$$\begin{aligned} \text{dist}(Z) &\leq \nu\left(\|Z_-\|_F^q\right) \\ \text{dist}(Z,\mathbb{S}_+^{M,L}) &\leq \nu\left(\|Z^TZ - I\|_F^q\right) \\ \text{dist}(Z,\mathbb{S}_+^{M,L}) &\leq \nu\left(\|Z_-\|_F^q + \|Z^TZ - I\|_F^q\right) \end{aligned} \tag{3-3}$$

Chen et al. investigate and derive tight error bounds with specific constants for the above bounds [52]. The third expression in equation 3-3 suggests that extracting a basis from a matrix with violating entries will increase the Euclidean distance.

Within the context of experiments, since X and Y are nonnegative, the distance to the nonnegative Stiefel manifold $\mathbb{S}_+^{N,L}$ is bounded by the orthogonality residual term. However, approximating a nonnegative basis $Q_{\geq 0}$ by approximating the subspace of Q_{qr} will admit an additional term based on the negative entries in Q_{qr}.

3-1-1 Projective NMF

Given the theoretical understanding, the first method in discussion to compute a nonnegative basis within Experiments A, and B is Projective Nonnegative Matrix Factorization. The corresponding basis $Q_{\text{PNMF}} \in \mathbb{R}^{N \times L}$ will be used to compress X within the aforementioned update rules.

As the next step, one may formulate the relevant PNMF objective function for Experiment A, which seeks to extract a nonnegative basis from sample matrix Y. In that case, the PNMF objective function takes the following form:

$$\min_{Q_{\geq 0}, Q^T Q = I} \|Y - QQ^T Y\|_F \tag{3-4}$$

Upon the computation of $Q_{\geq 0}$, the PNMF basis extracted from Y is thus referred to $Q_{\text{PNMF-Y}} \leftarrow Q$ corresponding to the objective function in 3-4.

Similarly, input matrix $X \in \mathbb{R}^{M \times N}$ can serve as an input to the PNMF objective function below:

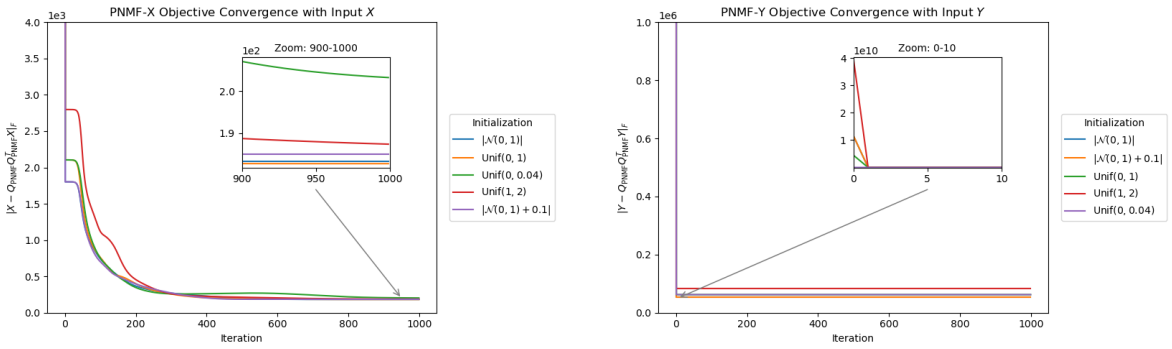
$$\min_{Q_{\geq 0}, Q^T Q = I} \|X - QQ^T X\|_F \tag{3-5}$$

The PNMF basis retrieved from X is thus assigned $Q_{\text{PNMF-X}} \leftarrow Q$ as the minimizer of the objective function in 3-5. This will thus serve as a PNMF basis for Experiment B. The projective formulation in optimization problems 3-4 and 3-5 incentivizes the term QQ^T to become an orthonormal projection matrix, which is the desired effect from this projective formulation [51].

Yang and Oja have proposed the following iterative update rule to monotonically decrease the objective function in 3-4 [51]:

$$Q \leftarrow Q \cdot \frac{XX^T Q}{QQ^T XX^T Q} \tag{3-6}$$

The update rule can thus analogously be applied on Y by replacing X. Furthermore, one may thus perform the update rule to obtain the PNMF bases extracted from X and Y, respectively. The corresponding PNMF objective function convergence is shown in Figure 3-1 below:



(a) Convergence of PNMF-X **(b)** Convergence of PNMF-Y

Figure 3-1: Convergence of PNMF Objective Function on X and Y.

Across various initializations and a maximum of 1000 iterations for sketch dimensions $L = 10$ and $L = 12$, respectively. PNMF-X extracts the basis directly from X; hence, oversampling is no longer performed, and the sketch dimension matches fixed-rank $L = K = 10$. Whilst PNMF-X is not so sensitive to initialization, the results for PNMF-Y show otherwise. The uniformly distributed initializations of Q_{prop} converge within the first 5 iterations, whereas the normally distributed initializations exhibit smoother convergence behavior.

Furthermore, one may observe the Frobenius norms of projection and orthogonality errors as displayed in Tables 3-3 and 3-4. The PNMF bases extracted from Y display poor orthogonality results with objective value results in the range 10^5; however, they reduce peak memory usage by more than three times. A much lower orthogonality residual is observed for PNMF-X bases as well as substantially lower projection errors. Unfortunately, the improvement in results for PNMF-X comes at the cost of slow computation and higher memory peaks.

Table 3-3: Projection and Orthogonality Metrics for PNMF-Y with Different Initialization (L = 12).

Initialization	$\|QQ^TX - X\|_F$	$\|Q^TQ - I\|_F$	**Time (s)**	**Mem (MB)**
$\|\mathcal{N}(0,1) + 0.1\|$	2195.01	7.1230	1134.18	746.78
$\|\mathcal{N}(0,1)\|$	2198.80	7.1352	1123.44	765.41
$\text{Unif}(0, 0.004)$	2576.75	8.3691	1.57	724.20
$\text{Unif}(0,1)$	2575.69	8.3656	1.75	737.93
$\text{Unif}(1,2)$	3406.09	11.0745	1.19	710.44

Table 3-4: Projection and Orthogonality Metrics for PNMF-X with Different Initialization (L = 10).

Initialization	$\|QQ^TX - X\|_F$	$\|Q^TQ - I\|_F$	**Time (s)**	**Mem (MB)**
$\|\mathcal{N}(0,1) + 0.1\|$	185.0456	0.7648	2149.99	2512.68
$\|\mathcal{N}(0,1)\|$	183.2189	0.7411	2517.07	2500.70
$\text{Unif}(0, 0.04)$	203.1779	0.8174	3541.82	2435.70
$\text{Unif}(0,1)$	182.7363	0.7624	3297.11	2512.68
$\text{Unif}(1,2)$	187.3563	0.7733	3542.58	2420.59

In line with expectations, an example in Figure 3-2 with a modulus of standard normal distribution initialization highlights a sparse nonnegative matrix with entries. Similar results

appear for other initializations of Q_{PNMF}.

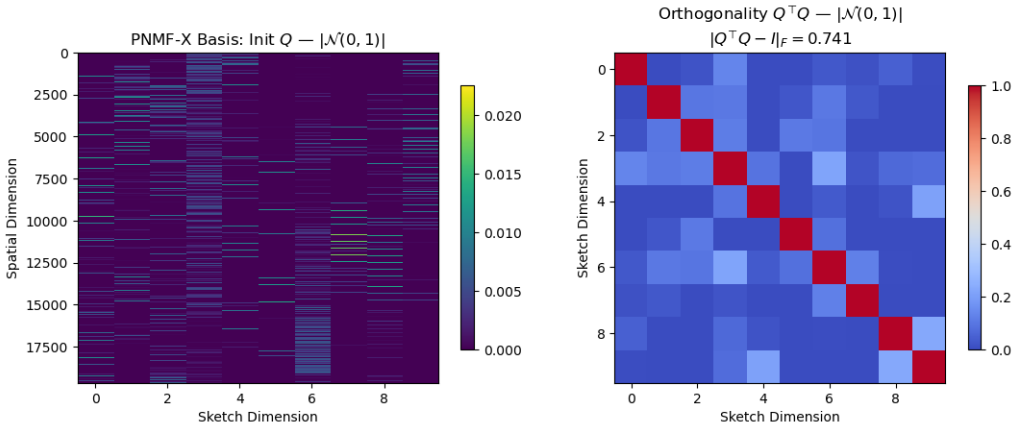


Figure 3-2: Heatmap of a PNMF-X Basis

The heatmap visualizes the magnitude of nonnegative entries based on the specified color maps. The orthogonality heatmap on the right-hand side of Figure 3-2 demonstrates that off-diagonal terms are still substantially greater than zero.

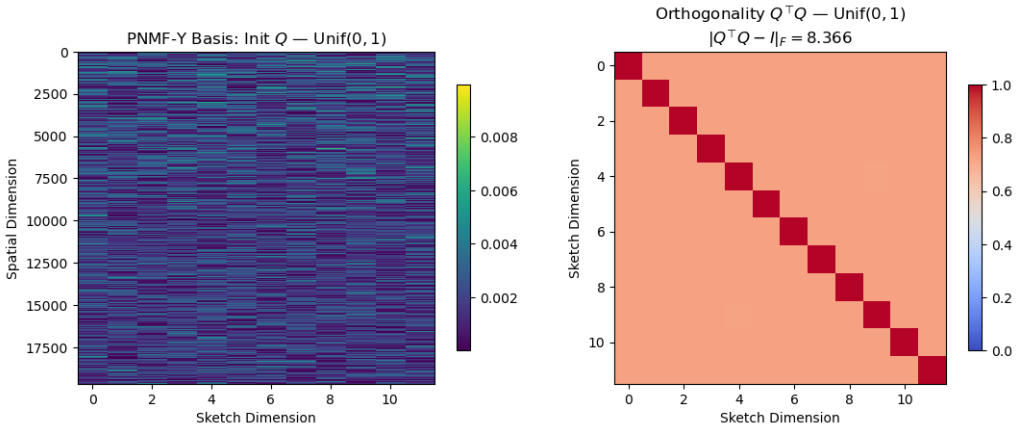


Figure 3-3: Heatmap of a PNMF-Y Basis

Figure 3-3 affirms the orthogonality results from the PNMF-Y table, where the off-diagonal terms in the Q^TQ matrix are close to 1. The corresponding orthonormal basis matrix is more dense, unlike its previous sparse counterpart. Therefore, upon forming $B = Q^TX$, one may expect the structure of B to be more discriminative for sparse basis Q_{PNMF}.

In PNMF literature, there exist more memory-efficient alternatives to the term XX^T within the PNMF update rule present in stochastic and QR-based variants [53] [54]. For the latter variant, the inclusion of QR factorization on X is a concern because of the aforementioned emphasis on entry-wise nonnegativity. Despite the availability of approximated PNMF variants that are more memory-efficient, extraction via conventional PNMF is prioritized to best maintain the properties of the basis as in the objective function 3-5.

3-1-2 Approximately Orthogonal NMF

In light of experiments A and B, this section aims to demonstrate the second method for extracting a nonnegative basis. Namely, Approximately Orthogonal NMF (AONMF) will be performed to extract basis from Y and X.

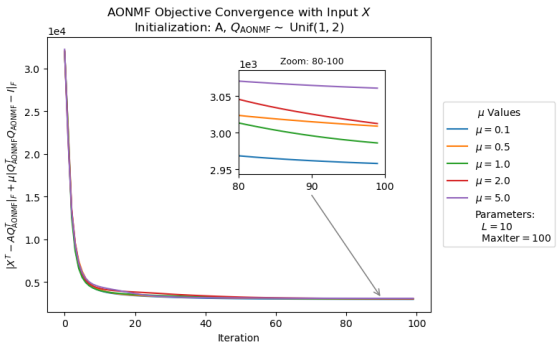
AONMF objective function takes the following form:

$$\min_{Q\geq 0, A_{\geq 0}, Q^TQ=I} \quad \| X - QA^T\|_F + \mu\|A^TA - I\| \tag{3-7}$$

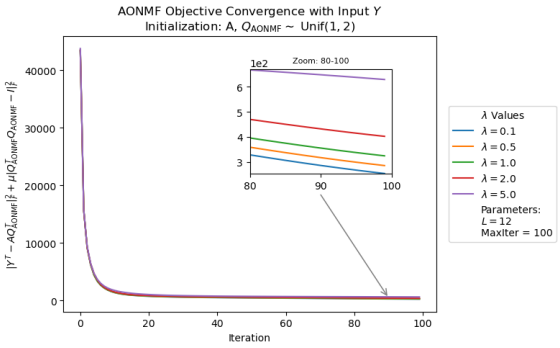
An update method via HALS for the Approximately Orthogonal NMF formulation by Li, Zhou, and Cichocki is chosen [50]. The HALS-AONMF is originally formulated for the right factor matrix to be orthogonal, however by transposing the input matrix, one may impose the orthogonality constraint on the left factor factor instead via a transpose:

$$X^T \approx AQ^T \tag{3-8}$$

What separates this method from orthogonal and projective formulations is the ability to control the degree of orthogonality necessary to achieve the needed balance between fitting error and orthogonality. Orthogonal NMF with a rigid orthogonality constraint is equivalent to hard k-means clustering, whereas soft orthogonality makes NMF synonymous with fuzzy C-means cluster [50].



(a) Convergence of AONMF-X (Example with Unif(1, 2) Initialization



(b) Convergence of AONMF-Y (Valid only with Unif(1, 2) Initialization)

Figure 3-4: Convergence of AONMF Objective Function on X and Y.

Similarly, AONMF has been performed across a maximum of 100 iterations to extract a basis Q_{AONMF} from X and Y respectively as shown in convergence plots in Figure 3-4. The convergence plots demonstrate similar rate of convergence, namely settling around 40 iterations, however due to marginal improvements a maximum of 100 iterations have been chosen. Due to similar results from alternative initializations, other AONMF-X plots are omitted.

For AONMF-Y, the updates with input Y have been extremely numerically sensitive, where only the results for uniform distribution between 1 and 2 have been able to produce a basis. Whilst further research into AONMF with tall-skinny matrix Y would be favorable, from this point, AONMF-Y-related exploration will involve only the above-mentioned initialization.

Table 3-5: Projection, Orthogonality, and Objective Value Metrics for AONMF-X with Unif$(1, 2)$ Initialization (L $= 10$).

μ	$\|QQ^TX - X\|_F$	$\|Q^TQ - I\|_F$	AONMF Obj Value	Time (s)	Mem (MB)
0.1000	1291.9441	4.4102	2958.2	1054.53	1966.46
0.5000	1285.3598	4.4118	3009.2	1057.21	1970.07
1.0000	1263.8609	4.3187	2986.2	1059.85	1969.21
2.0000	1248.1861	4.2867	3012.7	1054.86	1968.85
5.0000	1203.0548	4.1413	3061.1	1075.70	1967.53

Similar to PNMF-X, the sketch dimension is selected to be 10. The soft orthogonality in the objective function is reflected within the results, where the orthogonality residuals are higher than the PNMF counterpart. Depending on various degrees of orthogonality μ, the degree of orthogonality improves but does not change drastically. The overall bases for AONMF-X are sparse, however they tend to have higher magnitude of entries such as entries in the range 0.016 as opposed PNMF-X with 0.008. It will be interesting to observe whether that could produce more well-contrasted components.

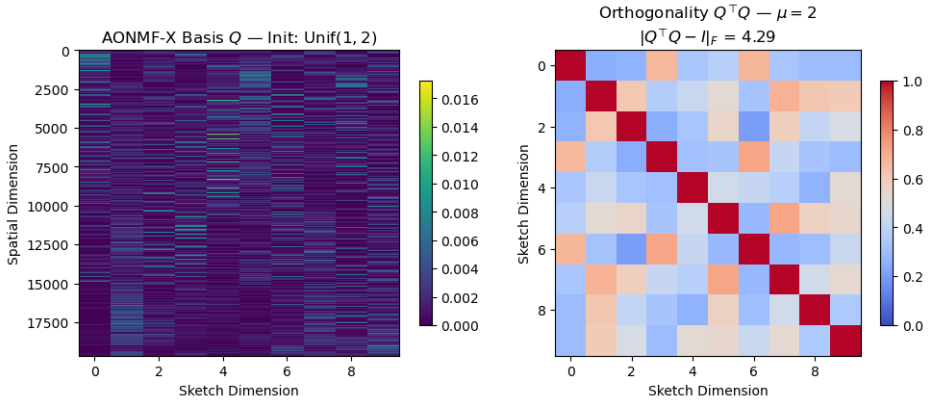


Figure 3-5: Heatmap of a AONMF-X Basis

Table 3-6: Projection, Orthogonality, and Objective Value Metrics for AONMF-Y with Unif$(1, 2)$ Initialization (L $= 12$).

μ	$\|QQ^TX - X\|_F$	$\|Q^TQ - I\|_F$	AONMF Obj Value	Time (s)	Mem (MB)
0.1000	2604.7836	8.7032	253.44	56.70	924.13
0.5000	2595.4341	8.6735	285.26	55.87	921.59
1.0000	2584.2677	8.6387	324.20	55.31	921.99
2.0000	2553.0668	8.5462	402.20	55.90	923.54
5.0000	2318.3433	7.8236	629.58	56.17	923.84

For AONMF-Y, however, the degree of orthogonality is the highest, and as previously mentioned, only one initialization was feasible. With much smaller peak memory, the computation

of the basis was within around 50 seconds. All bases incur high orthogonality errors, and generally, the bases with higher orthogonality errors would be also more dense as opposed to sparse.

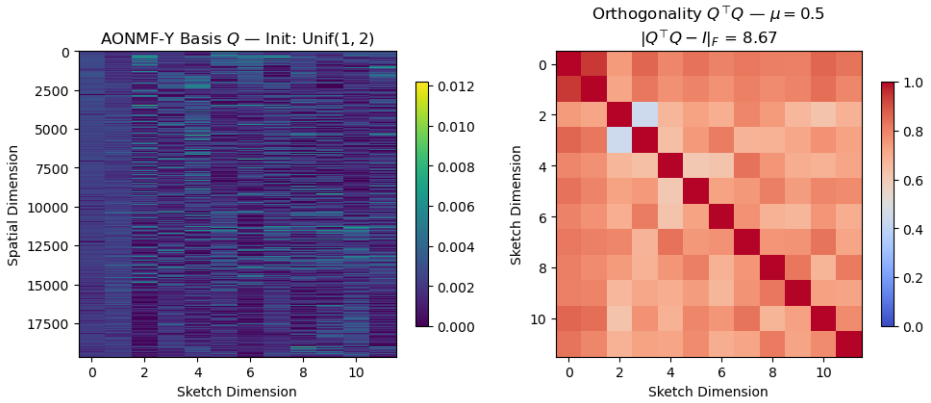


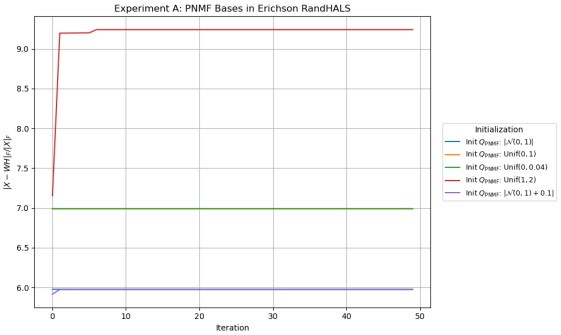
Figure 3-6: Heatmap of an AONMF-Y Basis

Overall, extracting a basis from X has shown to be much more computationally expensive, however, they will serve as a baseline for bases extracted from Y and the proposed method. Given the overall sensitivity of optimization with input Y, this suggests that further exploring methods that are better equipped to work with tall-skinny matrices may be favorable.

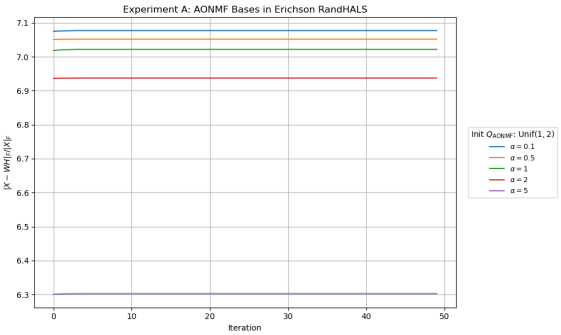
3-2 Experiments A and B

Given the overview of PNMF and AONMF, one may proceed with experiments A, and B, and namely apply the bases extracted from Y and X within the update rules:

- Erichson RandHALS
- Chaudhry RandMU

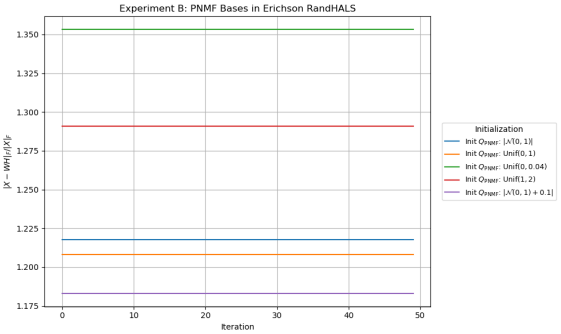


(a) Experiment A — Erichson RandHALS on PNMF-Y

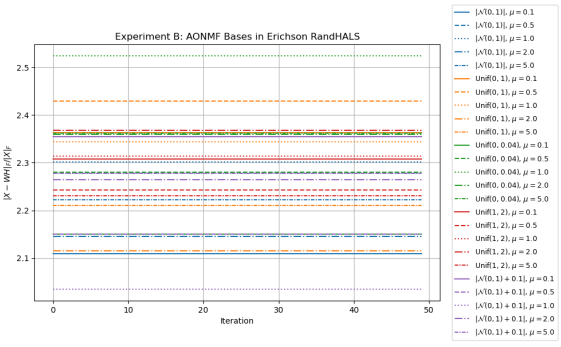


(b) Experiment A — Erichson RandHALS on AONMF-Y

Figure 3-7: Experiment A - Erichson RandHALS - PNMF-Y and AONMF-Y Bases



(a) Experiment B — Erichson RandHALS on PNMF-X

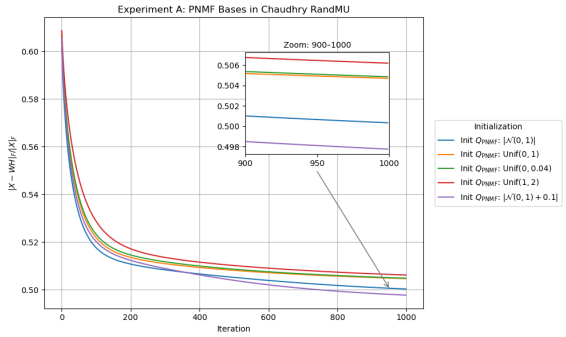


(b) Experiment B — Erichson RandHALS on AONMF-X

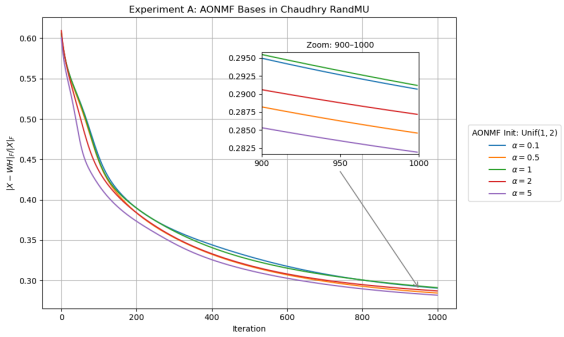
Figure 3-8: Experiment B - Erichson RandHALS - PNMF-X and AONMF-X Bases

The results from Experiments A and B with all PNMF and AONMF bases fail to converge within the update rule Erichson RandHALS. An interesting indication of why this could be a reason is provided in Appendix A in Table A-1. Some of the metrics indicated that the first dominant component may be responsible for just under 100% of explained variance, indicating imbalance. The condition number is substantially higher for the bases extracted from Y, also possibly suggesting numerical issues in using such a basis within the Erichson RandHALS update rule. Similarly, the bases with smaller spectral gaps tend to have slightly

improved initialization. Although more thorough analysis would be required on the desired properties of $B = Q^T X$, Appendix A serves as an initial starting point. Overall, due to the lack of convergence within the RandHALS update method, it is not sufficient to consider the interpretability of the resulting factor matrices W and H.



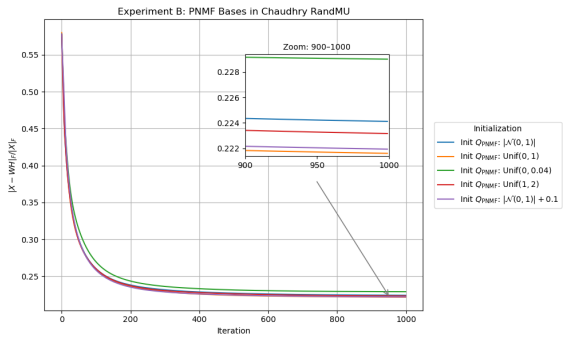
(a) Experiment A — Chaudhry RandMU on PNMF-Y



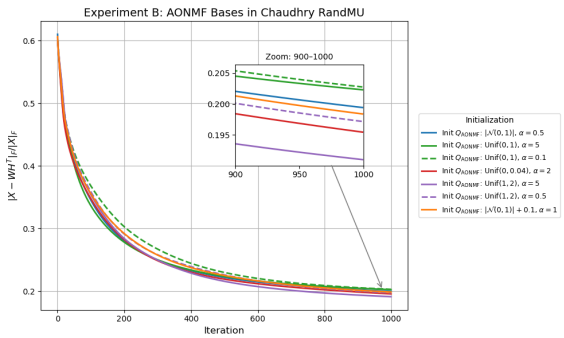
(b) Experiment A — Chaudhry RandMU on AONMF-Y

Figure 3-9: Experiment A - Chaudhry RandMU - PNMF-Y and AONMF-Y Bases

Chaudhry RandMU update rule, with its regularization on the complementary space, is able to successfully converge for all initializations for both PNMF-Y and AONMF-Y bases in Experiment A. Gaussian initialization with offset achieved lowest factorization from the PNMF-Y bases, whereas the term with highest orthogonality penalty μ for AONMF-Y similarly achieves lower factorization error. In general, the PNMF-Y bases in Experiment A tend to be more sensitive to initialization than in fact all other Chaudhry RandMU results for Experiments A and B.



(a) Experiment B — Chaudhry RandMU on PNMF-X

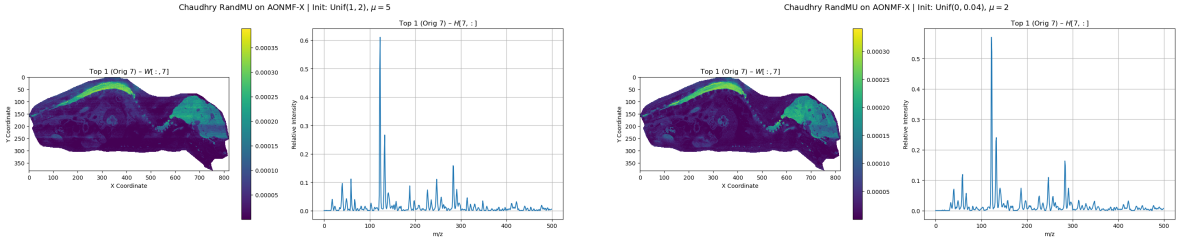


(b) Experiment A — Chaudhry RandMU on AONMF-X

Figure 3-10: Experiment B - Chaudhry RandMU - PNMF-Y and AONMF-X Bases

Experiment B with Chaudhry RandMU update rule demonstrates even lower approximation errors achieving approximation errors comparable to the QR-based and vanilla methods in Chapter 2. Given the plethora of initializations available to extract AONMF from X, only a handful of convergence behaviors are shown as reference in Figure 3-10b; however, this experiment is not significantly sensitive to initialization as can be seen.

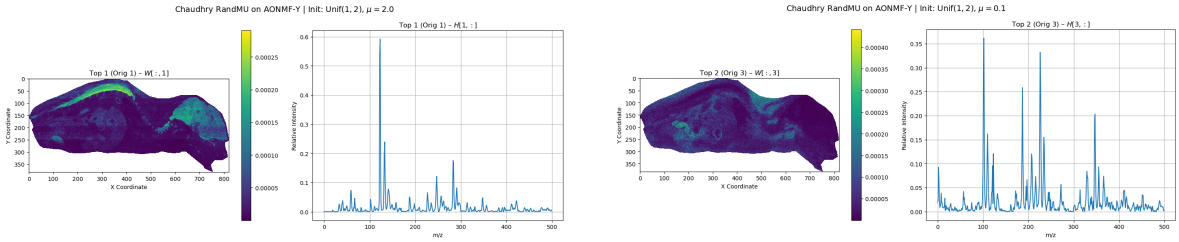
The factorized dominant IMS components from AONMF-X demonstrate the best interpretable result so far. Generally, the Chaudhry RandMU results for $\mu = 1$ and $\mu = 2$ display well-contrasted anatomical regions with little presence of noise and scattering. The spectral profiles for $\mu = 0.5$ may include na oisy spectral floor. Higher $\mu = 5$ may occasionally reduce the contrast in spatial maps; however, overall, it offers interesting insight.



(a) Init: Unif$(1, 2)$, $\mu = 5$ **(b)** Init: Unif$(0, 0.04)$, $\mu = 2$

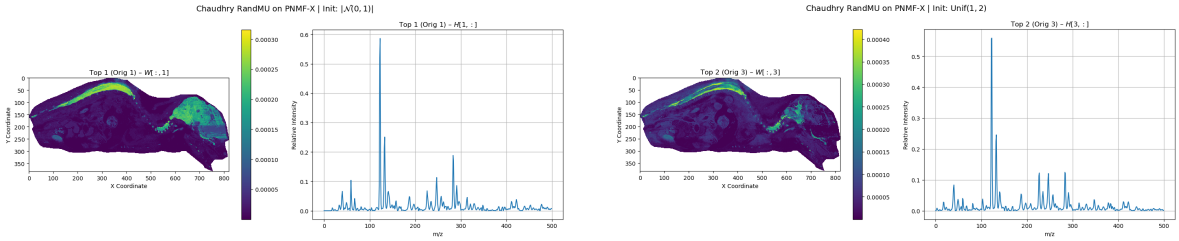
Figure 3-11: Experiment B - Chaudhry RandMU - IMS components AONMF-X

The spatial and spectral component results for orthogonality within AONMF-Y display better results for $\mu = 2$. The first dominant components offers interpretable components, however the remaining components do not offer significant value. For other values of μ, the relevant anatomical regions are no longer as well-contrasted and generally include more faint parts and scattered noise. The spectral profiles are analogous to AONMF-X.



(a) Init: Unif$(1, 2)$, $\mu = 2$ **(b)** Init:Unif$(1, 2)$, $\mu = 0.1$

Figure 3-12: Experiment A - Chaudhry RandMU - IMS components AONMF-Y



(a) Init: $|\mathcal{N}(0, 1)|$ **(b)** Init: Unif$(1, 2)$

Figure 3-13: Experiment B - Chaudhry RandMU - IMS components PNMF-X

As for PNMF-X results, the interplay between rigid orthogonality and nonnegativity offers interesting comparative insight with regard to AONMF-X. In this case, applying a more

selective basis with stronger orthogonality may occasionally be too discriminative and result in some incomplete selections of spatial regions unlike the AONMF-X counterpart. Previously highlighted higher magnitude in the basis entries, combined with an overall sparse structure with a soft orthogonality constraint may offer the best interpretation value. Unfortunately, using such bases comes at a cost of long computation time to extract the bases.

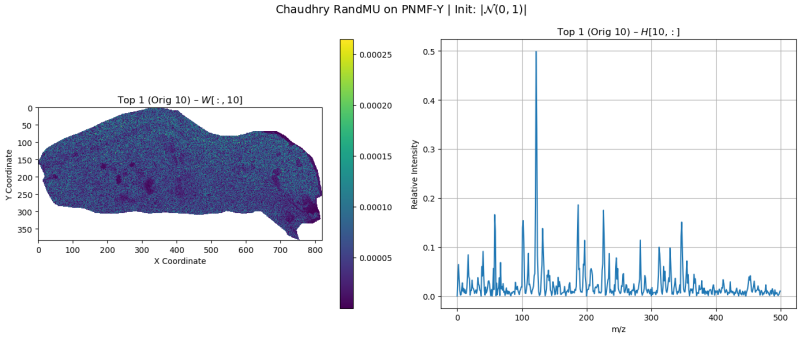


Figure 3-14: Experiment A - Chaudhry RandMU - IMS components PNMF-Y

Lastly, the bases extracted from PNMF-Y display the salt-and-pepper noise across the spatial components. Overall, PNMF-Y is insufficient for further use for investigating relevant parts of the tissue and peaks in the spectral profile.

3-3 Proposed Method

Orthogonal variants of Nonnegative Matrix Factorization such as PNMF, and AONMF have been able to compute nonnegative bases with varying degree of orthogonality, however they failed to converge within the Erichson RandHALS update method.

This chapter aims to propose an approximate basis Q_{prop} such that its objective function decreases the distance to the subspace spanned by QR-based basis Q. The objective function is shown in equation 3-9 below:

$$\min_{Q \geq 0} \mu \|Q^T Q - I\|_F^2 - \lambda \operatorname{Tr}(Q^T Q) + \beta \sum_i \theta_i^2 \tag{3-9}$$

The trace term seeks to capture the direction of maximum variance of near-optimal basis Q, whilst the second term seeks to control the orthogonality of the input matrix. The last term aims to minimize the principal angles θ_l from $l = 1$ to L for each of the L columns of Q_{qr} and Q_{prop}.

The objective function designed explicitly to approximate Q_{qr} allows only add an additional step within the pipeline and approximate a basis that is able to converge within the Erichson RandHALS update method. Simultaneously, the input matrix to the algorithm is a tall-skinny $M \times L$ Q_{qr} basis as opposed to $M \times N$ original input matrix.

Theorem 3-3.1. *[55] Let the columns of matrices $Q_{qr} \in \mathbb{R}^{M \times L}$ and $Q_{prop} \in \mathbb{R}^{M \times L}$ form orthonormal bases for the subspaces $\mathcal{Q}$ and $\tilde{\mathcal{Q}}$ respectively. Let the corresponding Singular Value Decomposition $Q_{qr}^T Q_{prop}$ be*

$$Q_{qr}^T Q_{prop} = U \Sigma V^T,$$

U and V are unitary matrices, and Σ is a $L \times L$ diagonal matrix with real diagonal elements $s_1(Q_{qr}^T Q_{prop}), \ldots, s_L(Q_{qr}^T Q_{prop})$ in nonincreasing order, with $L = K + P$. Then,

$$\cos \Theta^{\uparrow}(\mathcal{Q}, \tilde{\mathcal{Q}}) = S(Q_{qr}^T Q_{prop}) = \left[s_1(Q_{qr}^T Q_{prop}), \ldots, s_L(Q_{qr}^T Q_{prop}) \right],$$

where $\Theta^{\uparrow}(\mathcal{Q}, \tilde{\mathcal{Q}})$ denotes the vector of principal angles between $\mathcal{Q}$ and $\tilde{\mathcal{Q}}$ arranged in nondecreasing order, and $S(A)$ denotes the vector of singular values of A. Moreover, the principal vectors associated with this pair of subspaces are given by the first m columns of XU and YV, respectively.

According to this theorem, the term Q_{prop} should be an orthonormal matrix during each of the iterations the iterations. On one hand, projecting matrix Q_{prop} to the nearest orthonormal matrix is also used to ensure the orthogonality of a factor in ONMF [35]. At the same time, two orthonormal matrices at each iteration can thus be compared such that the distances between subspaces could be tracked via principal angles. Therefore, the method will make use of the projection to the nearest orthonormal matrix at each update step. A closed-form solution to the projection to the nearest orthonormal matrix may bypass the matrix inversion operation.

$$Q_{\text{orth}} = Q(Q^T Q)^{-1/2} \tag{3-10}$$

where $Q_{\text{prop}} \leftarrow Q_{\text{orth}}$

The matrix inverse operation can be circumvented via Eigenvalue Decomposition of $\tilde{Q}^T\tilde{Q}$ as follows:

$$Q_{\text{prop}}^T Q_{\text{prop}} = U\Lambda U^T \tag{3-11}$$

Thus the corresponding eigenvalues in diagonal Λ may be inverted:

$$(Q_{\text{prop}}^T Q_{\text{prop}})^{-1/2} = U\Lambda^{-1/2}U^T \tag{3-12}$$

$$\Lambda^{-1/2} = \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \ldots, \frac{1}{\sqrt{\lambda_L}}\right) \tag{3-13}$$

$$Q_{\text{prop}} \leftarrow Q_{\text{orth}} = Q_{\text{prop}}U\Lambda^{-1/2}U^T \tag{3-14}$$

The overall function is nonconvex, however each of the terms are differentiable hence a simple gradient descent based algorithm is proposed. Subsequently, the regularization parameter λ, μ, and β allows to attribute the weight to the degree of alignment between the subspaces against the level of orthogonality. Lastly, orthormalization via a closed-form equation and a projection to nonnegative orthant ensures the constraints are satisfied.

The algorithm can thus be summarized in the Algorithm 1 below:

Algorithm 1 Proposed Method to Approximate a Nonnegative Approximately Orthonormal Matrix Q_{prop} from Q_{qr}

Require: $Q_{\text{qr}} \in \mathbb{R}^{M \times L}$, Sketch Dimension L, Step Size η, Regularization Parameters μ, λ, β, Max Iterations `max-iter`, Tolerance *tol*

Ensure: Approximately Orthonormal matrix $Q_{\text{prop}} \in \mathbb{R}^{M \times L}$

Initialize: Q_{prop}

for iteration $= 1$ to `max-iter` **do**

Compute objective:

$$f(Q_{\text{prop}}) = \mu \|Q_{\text{prop}}^T Q_{\text{prop}} - I\|_F^2 - \lambda \operatorname{Tr}(Q_{\text{prop}}^T Q_{\text{qr}}) + \beta \sum_i \theta_i^2$$

Compute gradients for the first two terms:

$$\nabla_{Q_{\text{prop}}} f_1 = 4\mu Q_{\text{prop}}(Q_{\text{prop}}^T Q_{\text{prop}} - I), \quad \nabla_{Q_{\text{prop}}} f_2 = -\lambda Q_{\text{qr}}$$

Compute principal angles via SVD: $Q_{\text{qr}}^T Q_{\text{prop}} = U_k \Sigma_k V_k^T$

Compute third gradient term:

$$\nabla_{Q_{\text{prop}}} f_3 = -2\beta Q_{\text{qr}} U_k \operatorname{diag}(\theta_i) V_k^T$$

Gradient descent update:

$$Q_{\text{prop}} \leftarrow Q_{\text{prop}} - \eta(\nabla_{Q_{\text{prop}}} f_1 + \nabla_{Q_{\text{prop}}} f_2 + \nabla_{Q_{\text{prop}}} f_3)$$

Project to the nearest orthonormal matrix and clip negative entries:

$$Q_{\text{prop}} \leftarrow Q_{\text{prop}}(Q_{\text{prop}}^T Q_{\text{prop}})^{-1/2}, \quad Q_{\text{prop}} \leftarrow \max(Q_{\text{prop}}, 0)$$

if $\|\nabla_{Q_{\text{prop}}} f\|_F < tol$ **then**

break

end if

end for

return $Q_{\text{prop}} \in \mathbb{R}^{M \times L}$

3-4 Experiment C

For *Experiment C*, the Q_{prop} basis directly approximated the Q_{qr} basis. The experiments are run analogously for different initializations, as well as eight sets of hyperparameters demonstrated in Table 3-2. The step size in the gradient descent-based method is set to $\eta = 0.01$, and the basis is extracted over a maximum of 1000 iterations. As can be seen in Table A-4, the proposed basis is able to approximate within around 150 seconds, almost 10 times faster than retrieving AONMF-X bases and almost 20 times faster for 1000 iterations than PNMF-X bases.

Most importantly, unlike the other parts, this method succeeds in converging within almost all hyperparameter sets and initializations except for set 6, which emphasizes only the orthogonality term.

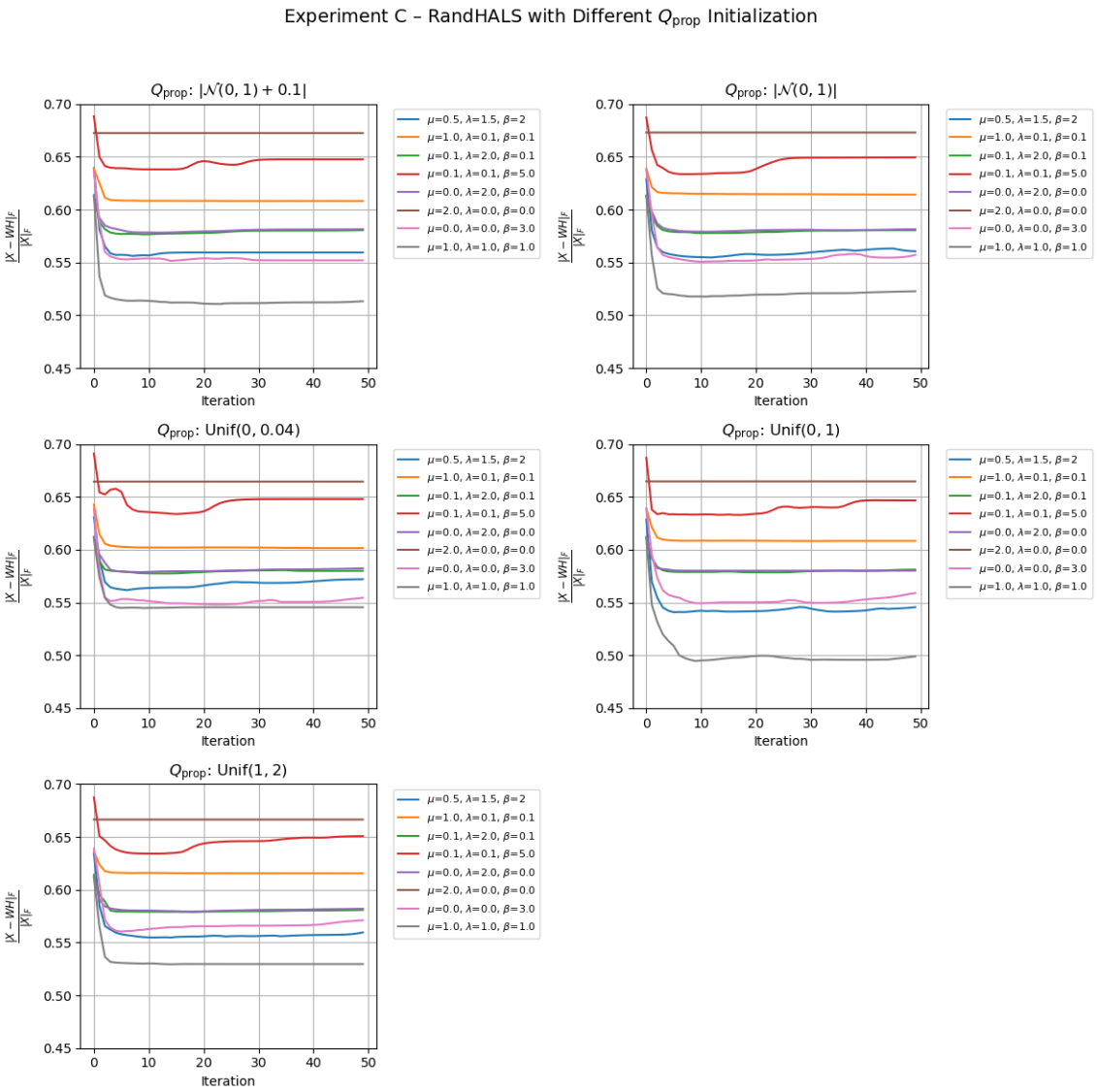
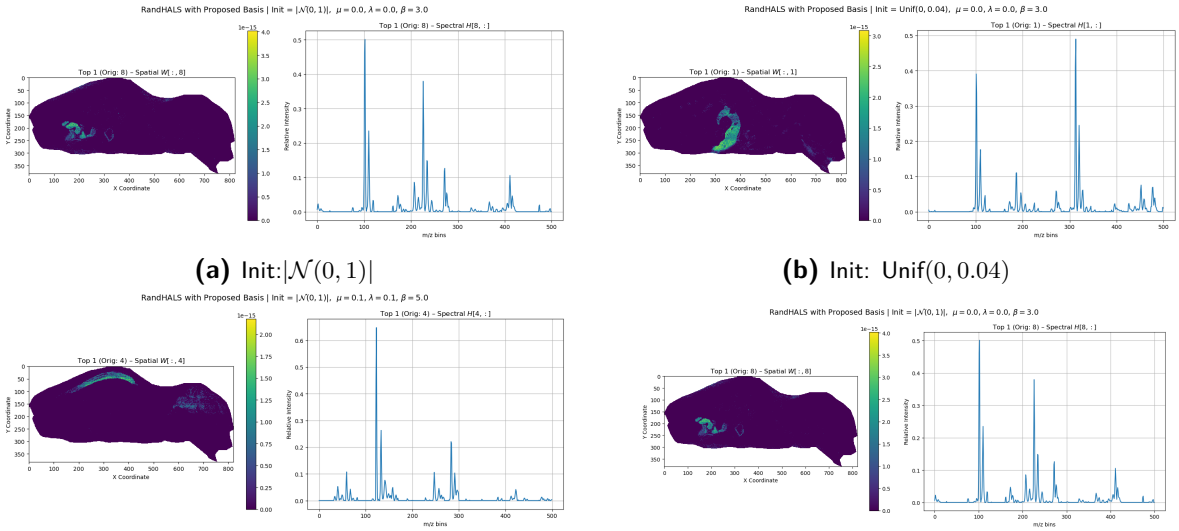


Figure 3-15: Experiment C - Erichson RandHALS



(a) Init:$|\mathcal{N}(0,1)|$ **(b)** Init: Unif$(0, 0.04)$

(c) Init: $|\mathcal{N}(0,1)|$ **(d)** Init: $|\mathcal{N}(0,1)+0.1|$

Figure 3-16: Experiment C — Erichson RandHALS with Proposed Bases - IMS Components

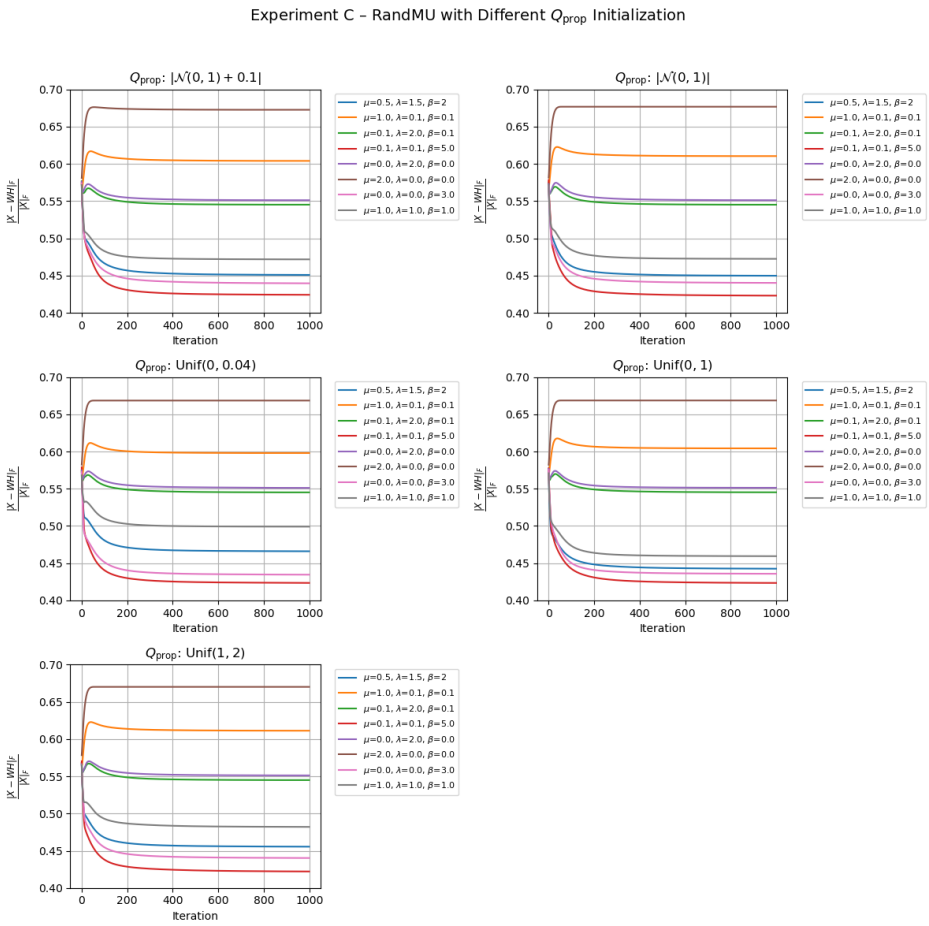
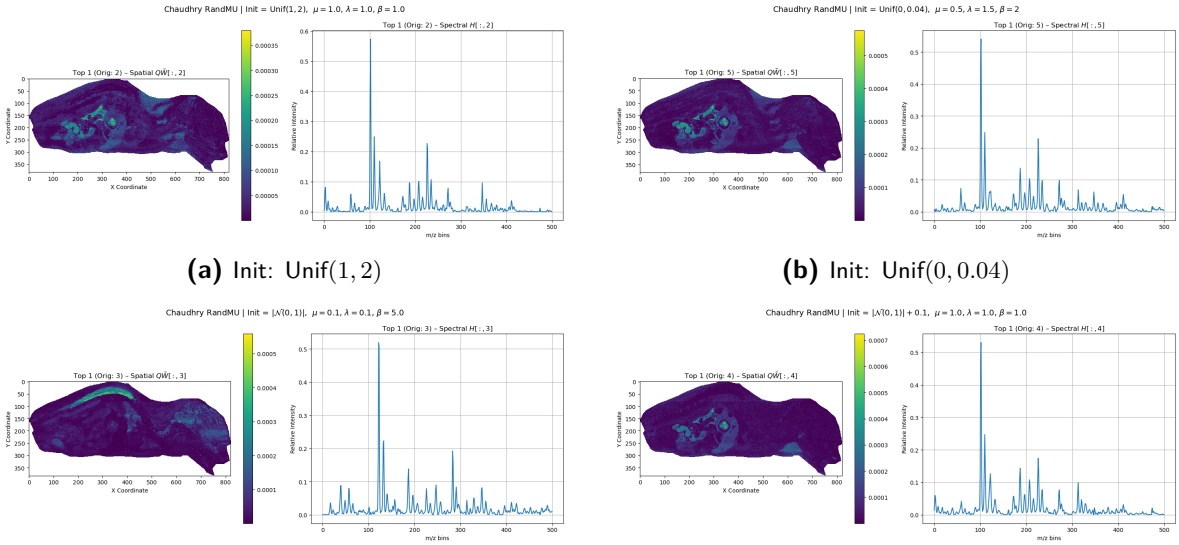


Figure 3-17: Experiment C - Chaudhry RandMU



(a) Init: Unif$(1, 2)$ **(b)** Init: Unif$(0, 0.04)$

(c) Init: $|\mathcal{N}(0, 1)|$ **(d)** Init: $|\mathcal{N}(0, 1) + 0.1|$

Figure 3-18: Experiment C — Chaudhry RandMU with Proposed Bases - IMS Components

Lastly, the Chaudhry RandMU offers improved interpretability where the regularization of the complementary space negates some of the overselectiveness of the basis. Similarly, the basis with the strongest interpretability had stronger regularization on the principal angles. The summary of the principal angles for all hyperparameter sets and other remaining tables is available in Appendix A.

3-5 Summary

Given the numerous update rules and numerous choices for bases, a summary table in Table 3-7 below:

Table 3-7: High-level Comparison of various conventional and randomized NMF update methods for specific choices of (nonnegative) bases. Green indicates good interpretability, yellow indicates moderate interpretability, orange indicates poor interpretability and red indicates very poor interpretability

Method	Component Interpretability	Spatial Localization	Spectral Clarity	Init Robustness	Notes
Baseline Methods					
Vanilla HALS	Moderate	Scattered but interpretable	Recognizable peaks	Stable	Some presence of noise but interpretable. Lacks scalability
Vanilla MU	Moderate	Scattered regions	Low peak resolution	Stable	Analogous to HALS
LS HALS	Mildly interpretable	Patchy	Modest spectral shape	Fairly stable	Block-wise efficient, but lacks clear anatomical structure in W. Needs SVD for column selection
Update Methods with QR-based Basis					
RandMU	High	Anatomically plausible	Well-separated peaks	Stable	Extremely slow convergence; occasionally faint spatial maps. Requires $\sigma > 0$ computation for Q_{qr}
Erichson RandHALS	Moderate-High	Main subregions recovered	Occasional instability	Mildly sensitive	Faster than RandMU, iteratively loses projection quality.
Masuda RandHALS ($\varepsilon = 0$)	High	Well-separated maps	Only two dominant spectra	N/A	Slow update rule, and zero components besides 2 dominant components
Masuda RandHALS ($\varepsilon = -10^{-6}$)	Low	Unclear features	Negative/noisy	N/A	Uninterpretable due to negative values
Update Methods with Nonnegative Bases					
Erichson RandHALS (PNMF-Y)	Poor	No convergence	No convergence	Unstable	Cannot converge
Erichson RandHALS (AONMF-Y)	Poor	Unstable maps	No convergence	Inconsistent	Cannot converge
RandMU (PNMF-Y)	Low	Diffuse	Noisy	Slight improvement	Salt and pepper noise
RandMU (AONMF-Y)	Moderate	Partial cortex	Mixed clarity	Variable	Sensitive to initialization, fast
RandMU (AONMF-X)	Very High	Highly consistent regions	Sharp peaks	Slow basis learning	Slow in computation, but excellent spatial and spectral interpretability.
RandMU (PNMF-X)	High	Coherent maps	Sharp peaks	Slow basis learning	Similar to AONMF-X, but more selective
Erichson RandHALS (Proposed)	Selective	Pinpoints regions	Sharp, low-noise peaks	Fast but unstable for insufficiently orthogonal Q_{prop}	Low orthogonality errors in basis, and fastly converging within updates. Significant loss due to selectivity, well-separated components for high β
Chaudhry RandMU (Proposed)	High	Consistent maps	Some mild noise	Stable	Similar to Erichson RandHALS, but less selective. Improves interpretability. Still converges fast within update method

Chapter 4

Discussion

Upon the completion of the exploration of alternative nonnegative bases within update methods such as Erichson RandHALS and Chaudhry RandMU, there are numerous key insights that are worth noting with respect to the framework. Experiment A, in an attempt to directly substitute the orthonormalization step, encounters difficulties in extracting a basis with a high degree of orthonormality. Whilst Experiment B, to a large extent, produces more orthogonal bases, they still do not suffice for use within Erichson RandHALS.

A proposed approximately orthogonal basis Q_{prop} managed to consistently converge within the update rules, and offer a diverse range of hyperparameter tunings. Whilst facilitating a more compatible nonnegative basis to the update rule, the proposed method has also demonstrated the importance of the principal angle alignment term. Experiment C suggests that a slightly relaxed orthogonality penalty combined with high principal angle penalties may be further worth investigating. Additionally, given the very selective extraction of spatial and spectral features, the experiments have shown that a high degree of orthogonality may not always be favorable.

For example, bases from AONMF's ability to control soft orthogonality has been shown to be very useful for interpretable factorizations of IMS data. Unfortunately, this result comes with an inefficient approach of extracting the AONMF basis directly from X. Based on the results from this paper, a more efficient alternative to AONMF-X would be an extremely useful area of research. The contrast with results from PNMF bases allowed for a balanced comparison between hard and soft orthogonality methods of NMF. With this in mind, further research in possibly combining said basis extraction methods with clustering algorithms may also offer value for interpretability.

Additionally, the ability of the RandMU update method to accommodate a diverse range of bases is attributed to the regularization of the complementary space of the projector. The use of nonnegative approximately orthogonal bases eliminates the use of shifted correctivity on the projection term QQ^T and vastly accelerates convergence. Given the success of the update method, it would be interesting to further explore the use of double-sided compression [42], and develop the relevant theoretical guarantees in relation use of nonnegative bases.

Chapter 5

Conclusion

In conclusion, this thesis, one may revisit the research question: *In what ways can nonnegativity constraints be integrated into randomized SVD-inspired pipeline for NMF?*

First, Chapter 2 described three main methods for directly mitigating the influence of negative entries in Q_{qr}, which included entry-wise clipping, bounding approach, and shifted projections. The Chapter highlighted that the lack of sign constraints on compressed matrices further forces it to iteratively compromise on projection quality without the ability to precisely qualify whether essential features are lost. In the bounding approach, the relaxation of the nonnegativity constraint on W has been shown to render components uninterpretable for IMS data. A strict orthogonality $\epsilon = 0$ has worked to extract the two most dominant components, however rendered other components to be zero. Lastly, the shifted projection vastly slows down convergence behavior; however, it offered a new perspective concerning the application of regularization of complementary space for nonnegative bases.

In retrospect, the strategies to maintain the use of the QR-based basis fully eliminate the influence of its negative entries. This problem has been solved from its root by direct extraction of fully entry-wise nonnegative bases. In contrast, the exploration demonstrated that nonnegative approximately orthogonal bases extracted from X demonstrate high factorization performance. Unfortunately, this comes at the cost of higher computation, which conflicts with the design of the Randomized NMF Framework. Therefore, it is important to highlight that the Proposed Method has succeeded to be compatible with two randomized update methods, however the combination of the orthogonality and nonnegativity constraints, make the basis extraction problem non-trivial. The approach via AONMF has demonstrated valuable insights toward leveraging soft orthogonality to enhance the interpretability of factor matrices.

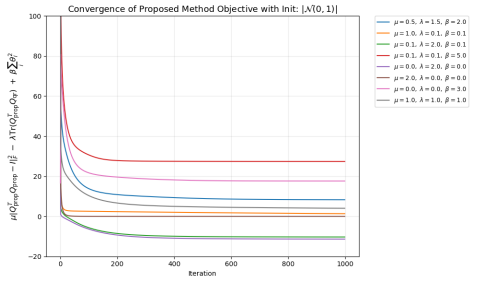
Therefore, this thesis has been able to propose multiple strategies to eliminate the problem of entry-wise negativity on a QR-based basis and offer multiple strategies to integrate nonnegative solutions within the data pipeline. Not only did this work challenge the status-quo choice for the basis, but also offered numerous improvements in terms of interpretability of resulting factor matrices for IMS data and also contributed a solution to maintain nonnegativity even in lower-dimensional subspaces.

Appendix A

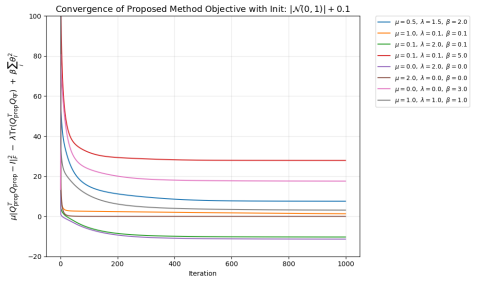
Supplementary Material

Table A-1: B-matrix statistics across PNMF-X, PNMF-Y, AONMF-Y, and AONMF-X bases (Frobenius norm, nuclear norm, condition number, rank, spectral gap, and explained variance)

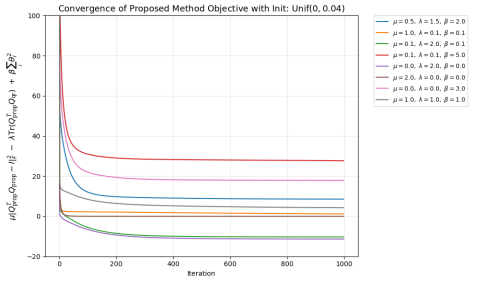
Method	**Init / μ**	$\|B\|_F$	$\|B\|_*$	$\mathrm{cond}(B)$	rank	Gap	EV_1	$\mathrm{EV}_{90\%}$	$\mathrm{EV}_{95\%}$	$\mathrm{EV}_{99\%}$
PNMF-X	$\lvert\mathcal{N}(0,1)+0.1\rvert$	431.43	775.73	46.13	10	252.21	0.8369	2	3	5
PNMF-X	$\lvert\mathcal{N}(0,1)\rvert$	430.33	772.51	42.91	10	251.34	0.8373	2	3	5
PNMF-X	$\mathrm{Unif}(0, 0.04)$	437.96	775.03	49.85	10	259.94	0.8447	2	2	5
PNMF-X	$\mathrm{Unif}(0, 1)$	430.48	773.76	47.33	10	251.76	0.8366	2	3	5
PNMF-X	$\mathrm{Unif}(1, 2)$	432.25	775.22	43.51	10	253.35	0.8382	2	3	5
PNMF-Y	$\lvert\mathcal{N}(0,1)\rvert$	897.83	914.87	14498.66	12	882.32	0.9997	1	1	1
PNMF-Y	$\lvert\mathcal{N}(0,1)+0.1\rvert$	897.25	918.91	14686.36	12	877.29	0.9995	1	1	1
PNMF-Y	$\mathrm{Unif}(0, 1)$	962.85	964.71	16273.35	12	962.30	1.0000	1	1	1
PNMF-Y	$\mathrm{Unif}(1, 2)$	1092.68	1093.38	62782.91	12	1092.48	1.0000	1	1	1
PNMF-Y	$\mathrm{Unif}(0, 0.04)$	963.03	964.70	18931.84	12	962.53	1.0000	1	1	1
AONMF-Y	$\mu = 0.1$	962.65	1192.07	699.81	12	855.04	0.9854	1	1	2
AONMF-Y	$\mu = 0.5$	961.13	1191.54	689.04	12	853.01	0.9852	1	1	2
AONMF-Y	$\mu = 1.0$	959.30	1190.89	681.76	12	850.43	0.9850	1	1	2
AONMF-Y	$\mu = 2.0$	954.02	1188.97	686.96	12	842.47	0.9842	1	1	2
AONMF-Y	$\mu = 5.0$	916.10	1180.02	541.72	12	778.57	0.9761	1	1	2
AONMF-X	$\mu = 0.1$	739.42	1081.26	92.37	10	563.11	0.9429	1	2	3
AONMF-X	$\mu = 0.5$	739.33	1081.46	97.74	10	554.96	0.9397	1	2	3
AONMF-X	$\mu = 1.0$	733.69	1078.21	87.03	10	556.62	0.9413	1	2	3
AONMF-X	$\mu = 2.0$	729.94	1074.28	93.94	10	550.24	0.9398	1	2	3
AONMF-X	$\mu = 5.0$	720.62	1070.32	86.35	10	537.49	0.9364	1	2	3



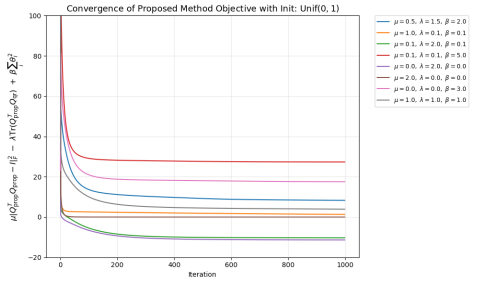
(a) Init: $|\mathcal{N}(0,1)|$



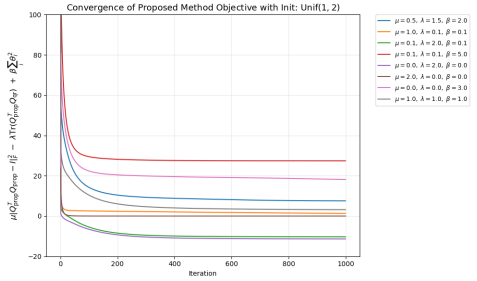
(b) Init: $|\mathcal{N}(0,1) + 0.1|$



(c) Init: $\text{Unif}(0, 0.04)$



(d) Init: $\text{Unif}(0,1)$



(e) Init: $\text{Unif}(1,2)$

Figure A-1: Convergence - Proposed Method Objective with different initialization and hyper-parameter sets.

Table A-2: Erichson RandHALS - Proposed Basis Metrics (Including Explained Variance thresholds EV's)

Initialization	Set	Rel. Error	Time (s)	Mem (MB)	Rank	EV_1	EV90%	EV95%
$\lvert\mathcal{N}(0,1)+0.1\rvert$	1	0.5594	25.40	3476.0	12	0.8724	2	2
	2	0.6081	57.75	2711.9	12	0.9374	1	2
	3	0.5806	57.80	2711.8	12	0.9355	1	2
	4	0.6476	45.12	2711.9	12	0.8726	2	2
	5	0.5815	58.26	2712.1	12	0.9377	1	2
	6	0.6725	6.02	2712.1	12	1.0000	1	1
	7	0.5520	42.03	2712.2	12	0.8630	2	3
	8	0.5132	58.32	2712.3	12	0.8634	2	3
$\lvert\mathcal{N}(0,1)\rvert$	1	0.5606	57.15	2712.4	12	0.8779	2	2
	2	0.6143	57.75	2712.4	12	0.9490	1	2
	3	0.5805	57.08	2712.5	12	0.9355	1	2
	4	0.6495	59.69	2712.9	12	0.8712	2	3
	5	0.5814	57.46	2712.9	12	0.9376	1	2
	6	0.6730	6.05	2712.9	12	1.0000	1	1
	7	0.5571	57.32	2713.0	12	0.8625	2	3
	8	0.5228	58.15	2713.1	12	0.8814	2	2
$\mathrm{Unif}(0, 0.04)$	1	0.5720	57.91	2713.1	12	0.8833	2	3
	2	0.6015	58.34	3480.8	12	0.9205	1	2
	3	0.5800	57.61	2711.7	12	0.9354	1	2
	4	0.6479	57.63	2711.9	12	0.8715	2	2
	5	0.5824	57.66	2712.0	12	0.9376	1	2
	6	0.6645	5.88	2712.1	12	1.0000	1	1
	7	0.5545	57.88	2712.2	12	0.8613	2	3
	8	0.5455	24.39	2712.3	12	0.8938	2	2
$\texttt{Unif}(0, 1)$	1	0.5456	58.56	3479.4	12	0.8671	2	3
	2	0.6083	55.65	2711.9	12	0.9376	1	2
	3	0.5811	58.53	2712.0	12	0.9355	1	2
	4	0.6467	57.84	2712.4	12	0.8713	2	3
	5	0.5801	13.93	2711.6	12	0.9377	1	2
	6	0.6647	6.08	3479.4	12	1.0000	1	1
	7	0.5589	58.40	3479.8	12	0.8610	2	3
	8	0.4989	57.70	2712.3	12	0.8678	2	3
$\texttt{Unif}(1, 2)$	1	0.5596	57.22	2712.3	12	0.8713	2	3
	2	0.6154	57.68	2712.4	12	0.9544	1	1
	3	0.5808	57.62	2712.4	12	0.9355	1	2
	4	0.6507	64.78	3480.1	12	0.8714	2	3
	5	0.5819	63.41	2708.1	12	0.9377	1	2
	6	0.6664	7.00	2711.7	12	1.0000	1	1
	7	0.5711	61.28	3193.8	12	0.8687	2	3
	8	0.5296	36.19	2711.8	12	0.8793	2	2

Table A-3: Chaudhry RandMU - Proposed Basis Metrics

Initialization	Set	Rel. Error	Time (s)	Mem (MB)
$\lvert\mathcal{N}(0,1)+0.1\rvert$	1	0.4509	1156.67	1928.40
	2	0.6040	937.94	2739.03
	3	0.5452	966.74	2739.18
	4	0.4242	1111.50	2738.80
	5	0.5513	924.71	2738.88
	6	0.6727	734.24	2739.12
	7	0.4396	1112.62	2738.81
	8	0.4718	1047.05	2738.85
$\lvert\mathcal{N}(0,1)\rvert$	1	0.4498	1109.12	2738.80
	2	0.6105	863.07	2738.87
	3	0.5452	985.20	2738.87
	4	0.4231	1121.93	2738.88
	5	0.5512	1039.30	2739.07
	6	0.6768	65.57	2739.66
	7	0.4402	1112.29	2739.76
	8	0.4725	1046.03	2739.96
$\mathrm{Unif}(0,0.04)$	1	0.4660	1112.52	2738.87
	2	0.5979	864.48	2738.84
	3	0.5451	1014.83	2738.80
	4	0.4235	1115.88	2739.06
	5	0.5511	1104.58	2738.84
	6	0.6684	63.15	2739.07
	7	0.4346	1117.18	2739.17
	8	0.4991	1051.18	2738.79
$\mathrm{Unif}(0,1)$	1	0.4426	1196.84	2738.80
	2	0.6042	901.46	2738.78
	3	0.5452	1044.60	2738.82
	4	0.4234	1118.05	2739.05
	5	0.5513	931.64	2738.88
	6	0.6685	63.99	2739.06
	7	0.4358	1117.12	2738.49
	8	0.4594	1122.46	2738.79
$\mathrm{Unif}(1,2)$	1	0.4555	1184.49	1918.09
	2	0.6113	977.41	2750.27
	3	0.5449	996.02	2749.91
	4	0.4222	1131.51	2751.50
	5	0.5512	1114.61	2749.18
	6	0.6702	71.18	2689.67
	7	0.4403	1223.51	2745.49
	8	0.4821	1220.66	2749.12

Table A-4: Proposed Method – Metrics Across Hyperparameters and Initializations Sweep

Init	**Hyperparams**	$\|Q_{\text{prop}}Q_{\text{prop}}^T X - X\|_2$	$\|\cdot\|_F$	$\|Q_{\text{prop}}Q_{\text{prop}}^T X - Q_{\text{qr}}Q_{\text{qr}}^T X\|_2$	$\|\cdot\|_F$	$\|Q_{\text{prop}}^T Q_{\text{prop}} - I\|_2$	$\|\cdot\|_F$	Time (s)	Mem (MB)
$\|\mathcal{N}(0,1)+0.1\|$	$\mu=0.5, \lambda=1.5, \beta=2.0$	147.98	190.59	148.13	179.80	0.3164	0.3366	145.31	3141.0
	$\mu=1.0, \lambda=0.1, \beta=0.1$	152.67	222.94	152.22	209.70	0.0494	0.0518	142.94	3117.9
	$\mu=0.1, \lambda=2.0, \beta=0.1$	140.89	207.96	139.46	193.91	0.1843	0.1947	144.64	3188.7
	$\mu=0.1, \lambda=0.1, \beta=5.0$	178.56	210.10	178.66	200.05	0.4700	0.5031	150.09	3090.3
	$\mu=0.0, \lambda=2.0, \beta=0.0$	141.64	209.16	140.23	195.05	0.1715	0.1810	143.05	3083.5
	$\mu=2.0, \lambda=0.0, \beta=0.0$	172.51	247.88	171.45	234.82	0.0109	0.0114	82.43	2938.6
	$\mu=0.0, \lambda=0.0, \beta=3.0$	150.32	188.47	150.37	177.17	0.3267	0.3481	143.30	3011.4
	$\mu=1.0, \lambda=1.0, \beta=1.0$	134.67	182.60	134.71	170.23	0.1930	0.2043	146.19	2969.0
$\|\mathcal{N}(0,1)\|$	$\mu=0.5, \lambda=1.5, \beta=2.0$	146.19	191.17	146.32	179.30	0.3187	0.3390	158.32	3358.8
	$\mu=1.0, \lambda=0.1, \beta=0.1$	153.52	225.32	152.97	212.29	0.0495	0.0519	145.15	3207.6
	$\mu=0.1, \lambda=2.0, \beta=0.1$	140.91	207.98	139.47	193.93	0.1844	0.1948	144.26	3176.7
	$\mu=0.1, \lambda=0.1, \beta=5.0$	177.97	209.20	178.10	199.03	0.4668	0.5000	140.68	3121.4
	$\mu=0.0, \lambda=2.0, \beta=0.0$	141.61	209.13	140.20	195.02	0.1714	0.1810	145.36	3008.1
	$\mu=2.0, \lambda=0.0, \beta=0.0$	172.74	248.05	171.66	234.98	0.0109	0.0114	87.02	3085.6
	$\mu=0.0, \lambda=0.0, \beta=3.0$	151.08	188.84	151.13	177.62	0.3283	0.3495	142.89	3032.8
	$\mu=1.0, \lambda=1.0, \beta=1.0$	130.49	184.77	130.35	172.00	0.2091	0.2211	142.74	3114.1
$\text{Unif}(0, 0.04)$	$\mu=0.5, \lambda=1.5, \beta=2.0$	147.45	194.57	147.30	182.80	0.3112	0.3298	160.53	2671.3
	$\mu=1.0, \lambda=0.1, \beta=0.1$	154.68	220.60	154.26	207.52	0.0461	0.0484	167.06	2729.4
	$\mu=0.1, \lambda=2.0, \beta=0.1$	140.89	207.90	139.45	193.86	0.1842	0.1946	153.42	2751.6
	$\mu=0.1, \lambda=0.1, \beta=5.0$	178.41	209.78	178.52	199.43	0.4691	0.5031	162.33	2791.8
	$\mu=0.0, \lambda=2.0, \beta=0.0$	141.62	209.14	140.21	195.04	0.1716	0.1812	172.65	2750.6
	$\mu=2.0, \lambda=0.0, \beta=0.0$	168.40	244.91	167.27	231.68	0.0109	0.0114	96.53	2783.2
	$\mu=0.0, \lambda=0.0, \beta=3.0$	150.53	187.72	150.62	176.70	0.3318	0.3541	172.81	2793.6
	$\mu=1.0, \lambda=1.0, \beta=1.0$	134.89	192.98	134.40	180.05	0.2021	0.2133	158.28	2650.4
$\text{Unif}(0, 1)$	$\mu=0.5, \lambda=1.5, \beta=2.0$	145.10	186.52	145.30	176.01	0.3069	0.3269	145.63	2925.5
	$\mu=1.0, \lambda=0.1, \beta=0.1$	152.63	222.96	152.16	209.70	0.0490	0.0514	143.43	2973.4
	$\mu=0.1, \lambda=2.0, \beta=0.1$	140.93	207.96	139.49	193.91	0.1843	0.1947	145.93	2814.7
	$\mu=0.1, \lambda=0.1, \beta=5.0$	177.98	209.26	178.11	199.12	0.4667	0.4998	159.07	2798.0
	$\mu=0.0, \lambda=2.0, \beta=0.0$	141.63	209.18	140.22	195.08	0.1716	0.1811	152.07	2844.5
	$\mu=2.0, \lambda=0.0, \beta=0.0$	168.61	244.98	167.47	231.76	0.0109	0.0114	90.39	2803.2
	$\mu=0.0, \lambda=0.0, \beta=3.0$	150.65	187.72	150.67	176.64	0.3285	0.3502	153.15	2895.9
	$\mu=1.0, \lambda=1.0, \beta=1.0$	129.85	179.53	129.66	168.06	0.2048	0.2174	154.99	2819.7
$\text{Unif}(1, 2)$	$\mu=0.5, \lambda=1.5, \beta=2.0$	148.48	191.06	148.36	178.90	0.3074	0.3267	156.69	2910.7
	$\mu=1.0, \lambda=0.1, \beta=0.1$	152.24	225.64	151.59	212.26	0.0505	0.0529	155.17	2927.7
	$\mu=0.1, \lambda=2.0, \beta=0.1$	140.89	207.94	139.45	193.89	0.1843	0.1947	152.71	2921.3
	$\mu=0.1, \lambda=0.1, \beta=5.0$	177.87	209.21	178.00	199.04	0.4663	0.4995	170.95	2916.0
	$\mu=0.0, \lambda=2.0, \beta=0.0$	141.63	209.20	140.23	195.09	0.1716	0.1811	161.36	2820.8
	$\mu=2.0, \lambda=0.0, \beta=0.0$	169.45	245.64	168.34	232.44	0.0109	0.0114	97.25	2565.5
	$\mu=0.0, \lambda=0.0, \beta=3.0$	151.97	191.48	151.99	179.15	0.3402	0.3633	163.35	2687.1
	$\mu=1.0, \lambda=1.0, \beta=1.0$	136.05	188.33	135.73	174.43	0.2122	0.2242	153.88	2709.0

Table A-5: B-matrix statistics across Proposed bases (Frobenius norm, nuclear norm, condition number, spectral gap, and explained variance)

Init.	Hyperparams	$\|B\|_F$	$\|B\|_*$	cond(B)	rank(B)	gap	EV_1	EV90%	EV95%	EV99%
$\|\mathcal{N}(0,1)+0.1\|$	μ=0.5, λ=1.5, β=2.0	369.36	668.18	55.59	12	240.54	0.8724	2	2	6
	μ=1.0, λ=0.1, β=0.1	300.49	467.14	119.30	12	229.87	0.9374	1	2	4
	μ=0.1, λ=2.0, β=0.1	333.83	533.41	180.27	12	259.67	0.9355	1	2	5
	μ=0.1, λ=0.1, β=5.0	392.38	709.42	57.11	12	256.94	0.8726	2	2	6
	μ=0.0, λ=2.0, β=0.0	330.63	523.60	609.05	12	258.47	0.9377	1	2	5
	μ=2.0, λ=0.0, β=0.0	274.29	278.28	3080.60	12	272.98	1.0000	1	1	1
	μ=0.0, λ=0.0, β=3.0	372.57	684.39	52.97	12	239.15	0.8630	2	3	6
	μ=1.0, λ=1.0, β=1.0	350.12	639.32	54.09	12	222.96	0.8634	2	3	6
$\|\mathcal{N}(0,1)\|$	μ=0.5, λ=1.5, β=2.0	370.24	660.22	66.44	12	245.47	0.8779	2	2	6
	μ=1.0, λ=0.1, β=0.1	298.75	453.35	110.21	12	239.14	0.9490	1	2	4
	μ=0.1, λ=2.0, β=0.1	333.83	533.38	185.41	12	259.70	0.9355	1	2	5
	μ=0.1, λ=0.1, β=5.0	392.19	710.28	56.68	12	256.18	0.8712	2	3	6
	μ=0.0, λ=2.0, β=0.0	330.64	523.74	553.81	12	258.41	0.9376	1	2	5
	μ=2.0, λ=0.0, β=0.0	274.14	277.07	3382.38	12	273.33	1.0000	1	1	1
	μ=0.0, λ=0.0, β=3.0	372.57	685.15	54.06	12	238.73	0.8625	2	3	6
	μ=1.0, λ=1.0, β=1.0	352.45	627.06	66.82	12	236.56	0.8814	2	2	6
uniform_0_004	μ=0.5, λ=1.5, β=2.0	365.91	661.58	50.00	12	250.09	0.8833	2	3	6
	μ=1.0, λ=0.1, β=0.1	301.69	490.22	80.90	12	220.77	0.9205	1	2	5
	μ=0.1, λ=2.0, β=0.1	333.84	533.36	208.43	12	259.57	0.9354	1	2	5
	μ=0.1, λ=0.1, β=5.0	392.36	707.80	54.72	12	255.77	0.8715	2	2	6
	μ=0.0, λ=2.0, β=0.0	330.65	523.83	499.49	12	258.43	0.9376	1	2	5
	μ=2.0, λ=0.0, β=0.0	276.98	280.45	2774.44	12	276.09	1.0000	1	1	1
	μ=0.0, λ=0.0, β=3.0	374.19	687.09	47.97	12	237.67	0.8613	2	3	6
	μ=1.0, λ=1.0, β=1.0	345.98	618.32	48.72	12	244.44	0.8938	2	2	6
uniform_0_1	μ=0.5, λ=1.5, β=2.0	369.58	673.77	63.00	12	238.86	0.8671	2	3	6
	μ=1.0, λ=0.1, β=0.1	300.43	467.87	88.92	12	231.13	0.9376	1	2	4
	μ=0.1, λ=2.0, β=0.1	333.84	533.58	166.63	12	259.68	0.9355	1	2	5
	μ=0.1, λ=0.1, β=5.0	392.11	709.88	56.47	12	256.20	0.8713	2	3	6
	μ=0.0, λ=2.0, β=0.0	330.63	523.76	491.48	12	258.50	0.9377	1	2	5
	μ=2.0, λ=0.0, β=0.0	276.91	280.20	3202.09	12	276.13	1.0000	1	1	1
	μ=0.0, λ=0.0, β=3.0	373.60	687.75	54.06	12	237.47	0.8610	2	3	6
	μ=1.0, λ=1.0, β=1.0	354.14	642.47	70.15	12	229.00	0.8678	2	3	6
uniform_1_2	μ=0.5, λ=1.5, β=2.0	367.15	667.29	57.01	12	240.22	0.8713	2	3	6
	μ=1.0, λ=0.1, β=0.1	298.68	450.88	96.98	12	244.38	0.9544	1	1	4
	μ=0.1, λ=2.0, β=0.1	333.84	533.53	177.12	12	259.64	0.9355	1	2	5
	μ=0.1, λ=0.1, β=5.0	392.06	709.77	56.30	12	256.23	0.8714	2	3	6
	μ=0.0, λ=2.0, β=0.0	330.62	523.62	493.58	12	258.48	0.9377	1	2	5
	μ=2.0, λ=0.0, β=0.0	276.32	279.83	3271.46	12	275.29	1.0000	1	1	1
	μ=0.0, λ=0.0, β=3.0	374.31	680.73	54.82	12	242.97	0.8687	2	3	6
	μ=1.0, λ=1.0, β=1.0	350.65	630.16	63.16	12	234.04	0.8793	2	2	6

Table A-6: Principal angles (in degrees) between Q_{prop} and Q_{qr} across different hyperparameters and initializations.

Init.	Hyperparams	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}
$\lvert\mathcal{N}(0,1)+0.1\rvert$	μ=0.5, λ=1.5, β=2.0	0.00	32.79	38.65	39.24	40.36	41.98	44.84	46.24	47.55	49.07	51.95	53.13
	μ=1.0, λ=0.1, β=0.1	24.61	58.27	59.23	61.45	63.32	65.25	66.39	66.63	69.73	71.42	72.67	74.86
	μ=0.1, λ=2.0, β=0.1	0.00	40.28	45.08	47.66	50.04	50.70	52.50	53.75	55.59	59.08	67.75	87.60
	μ=0.1, λ=0.1, β=5.0	0.00	28.67	33.73	36.06	37.98	39.06	41.60	41.98	43.98	45.10	45.59	49.93
	μ=0.0, λ=2.0, β=0.0	0.00	41.81	45.40	48.73	50.61	50.99	53.10	54.64	56.45	60.20	70.10	89.87
	μ=2.0, λ=0.0, β=0.0	30.77	89.26	89.34	89.41	89.48	89.51	89.62	89.75	89.82	89.84	89.91	89.97
	μ=0.0, λ=0.0, β=3.0	0.00	29.75	34.72	35.86	39.12	39.31	41.86	43.96	44.50	47.51	48.37	50.30
	μ=1.0, λ=1.0, β=1.0	0.00	32.85	37.80	38.50	41.71	43.07	44.87	47.61	49.24	51.93	54.57	56.49
$\lvert\mathcal{N}(0,1)\rvert$	μ=0.5, λ=1.5, β=2.0	0.00	31.32	36.85	39.66	42.07	43.98	44.61	46.63	48.54	49.19	51.93	59.71
	μ=1.0, λ=0.1, β=0.1	24.60	57.90	60.66	62.43	62.59	63.43	67.40	68.78	69.82	71.93	77.14	79.11
	μ=0.1, λ=2.0, β=0.1	0.00	40.23	45.08	47.63	50.02	50.72	52.49	53.75	55.60	59.07	67.74	87.67
	μ=0.1, λ=0.1, β=5.0	0.00	27.85	34.26	35.10	37.91	38.65	40.90	41.86	43.91	45.09	46.41	47.81
	μ=0.0, λ=2.0, β=0.0	0.00	41.80	45.40	48.76	50.62	50.96	53.10	54.65	56.46	60.20	70.09	89.94
	μ=2.0, λ=0.0, β=0.0	30.83	89.38	89.46	89.50	89.61	89.62	89.70	89.74	89.78	89.86	89.95	89.99
	μ=0.0, λ=0.0, β=3.0	0.00	30.14	34.54	35.85	39.07	39.20	41.79	43.33	44.91	47.26	47.32	52.27
	μ=1.0, λ=1.0, β=1.0	0.00	35.42	40.42	41.87	44.66	45.31	45.88	48.92	51.28	52.79	54.03	66.37
Unif(0,0.04)	μ=0.5, λ=1.5, β=2.0	0.00	34.10	36.32	39.07	40.30	41.75	45.17	45.42	48.40	50.94	52.21	57.73
	μ=1.0, λ=0.1, β=0.1	25.38	45.84	56.68	59.73	60.88	62.01	62.92	64.84	67.91	68.53	74.26	79.03
	μ=0.1, λ=2.0, β=0.1	0.00	40.22	45.04	47.60	50.06	50.72	52.45	53.75	55.58	59.09	67.71	87.94
	μ=0.1, λ=0.1, β=5.0	0.00	26.73	34.85	35.28	38.63	38.96	40.66	42.59	44.32	45.22	46.44	48.15
	μ=0.0, λ=2.0, β=0.0	0.00	41.82	45.40	48.76	50.68	51.00	53.08	54.65	56.43	60.21	70.08	89.94
	μ=2.0, λ=0.0, β=0.0	29.81	89.11	89.29	89.47	89.50	89.62	89.68	89.73	89.84	89.85	89.90	89.99
	μ=0.0, λ=0.0, β=3.0	0.00	27.84	35.32	36.08	39.82	40.40	41.53	43.02	46.41	47.25	47.98	52.40
	μ=1.0, λ=1.0, β=1.0	0.00	34.22	39.72	41.52	42.80	46.23	48.05	49.26	52.46	53.11	57.42	62.50
Unif(0,1)	μ=0.5, λ=1.5, β=2.0	0.00	30.79	34.91	40.03	40.33	41.73	44.19	46.10	48.66	50.27	53.92	55.33
	μ=1.0, λ=0.1, β=0.1	24.66	55.54	60.36	61.37	64.41	65.20	66.98	68.32	71.05	72.20	72.44	74.10
	μ=0.1, λ=2.0, β=0.1	0.00	40.26	45.06	47.64	50.04	50.73	52.52	53.74	55.62	59.09	67.72	87.42
	μ=0.1, λ=0.1, β=5.0	0.00	27.79	34.31	35.04	37.85	38.73	40.90	41.86	43.95	45.25	46.40	47.68
	μ=0.0, λ=2.0, β=0.0	0.00	41.83	45.44	48.78	50.63	50.96	53.09	54.64	56.46	60.19	70.10	89.98
	μ=2.0, λ=0.0, β=0.0	29.82	89.21	89.34	89.44	89.46	89.57	89.68	89.73	89.83	89.88	89.91	89.96
	μ=0.0, λ=0.0, β=3.0	0.00	29.68	35.33	36.40	38.16	40.09	41.51	43.19	46.29	46.72	47.64	49.83
	μ=1.0, λ=1.0, β=1.0	0.00	33.73	37.56	40.77	42.33	45.50	47.59	49.98	51.49	54.81	57.96	60.56
Unif(1,2)	μ=0.5, λ=1.5, β=2.0	0.00	32.47	36.22	37.18	40.06	41.50	43.28	44.97	47.47	49.99	51.51	54.26
	μ=1.0, λ=0.1, β=0.1	24.26	57.14	59.05	62.88	64.53	66.20	68.88	69.78	70.11	71.92	73.79	75.45
	μ=0.1, λ=2.0, β=0.1	0.00	40.23	45.08	47.66	50.02	50.72	52.47	53.76	55.60	59.06	67.72	87.55
	μ=0.1, λ=0.1, β=5.0	0.00	27.84	34.28	35.08	37.88	38.69	40.97	41.94	43.97	45.12	46.40	47.66
	μ=0.0, λ=2.0, β=0.0	0.00	41.81	45.44	48.81	50.61	51.00	53.14	54.64	56.42	60.16	70.13	89.90
	μ=2.0, λ=0.0, β=0.0	30.03	89.27	89.34	89.44	89.57	89.63	89.67	89.73	89.81	89.84	89.93	89.99
	μ=0.0, λ=0.0, β=3.0	0.00	30.48	35.55	36.02	39.35	40.86	41.63	44.03	46.26	47.77	49.56	50.94
	μ=1.0, λ=1.0, β=1.0	0.00	37.32	38.88	40.00	40.47	45.36	46.42	47.88	50.39	52.74	58.29	59.86

Bibliography

[1] N. Takahashi, "Code implementation of "an algorithm for randomized nonnegative matrix factorization and its global convergence" paper." Email Communication with the Authors, 2024. Code received from one of the authors.

[2] T. Masuda, T. Migita, and N. Takahashi, "An algorithm for randomized nonnegative matrix factorization and its global convergence," *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, Dec 2021.

[3] E. Bercovich and M. C. Javitt, "Medical imaging: From roentgen to the digital revolution, and beyond," *Rambam Maimonides Medical Journal*, vol. 9, Oct 2018.

[4] K. Krijnen, P. Blenkinsopp, R. M. Heeren, and I. G. Anthony, "Processing next-generation mass spectrometry imaging data: Principal component analysis at scale," *Journal of the American Society for Mass Spectrometry*, vol. 35, p. 3063–3069, Oct 2024.

[5] D. Pietkiewicz, S. Plewa, M. Zaborowski, T. Garrett, E. Matuszewska-Mach, Z. Kokot, and J. Matysiak, "Mass spectrometry imaging in gynecological cancers: the best is yet to come," *Cancer Cell International*, vol. 22, 12 2022.

[6] A. R. Buchberger, K. DeLaney, J. Johnson, and L. Li, "Mass spectrometry imaging: A review of emerging advancements and future insights," *Analytical Chemistry*, vol. 90, p. 240–265, Dec 2017.

[7] R. M. Caprioli, T. B. Farmer, and J. Gile, "Molecular imaging of biological samples: localization of peptides and proteins using maldi-tof ms," *Analytical Chemistry*, vol. 69, p. 4751–4760, Dec 1997.

[8] M. Nijs, T. Smets, E. Waelkens, and B. De Moor, "A mathematical comparison of non-negative matrix factorization related methods with practical implications for the analysis of mass spectrometry imaging data," *Rapid Communications in Mass Spectrometry*, vol. 35, Sep 2021.

[9] M. L. Reyzer, P. Chaurand, P. M. Angel, and R. M. Caprioli, "Direct molecular analysis of whole-body animal tissue sections by maldi imaging mass spectrometry," *Methods in Molecular Biology*, p. 285–301, 2010.

[10] C. Murayama, Y. Kimura, and M. Setou, "Imaging mass spectrometry: Principle and application," *Biophysical Reviews*, vol. 1, p. 131–139, Sep 2009.

[11] S. S. Rubakhin and J. V. Sweedler, "A mass spectrometry primer for mass spectrometry imaging," *Methods in Molecular Biology*, vol. 656, p. 21–49, 2010.

[12] L. A. McDonnell, A. van Remoortere, N. de Velde, R. J. van Zeijl, and A. M. Deeldera, "Imaging mass spectrometry data reduction: Automated feature identification and extraction," *Journal of the American Society for Mass Spectrometry*, vol. 21, p. 1969–1978, Dec 2010.

[13] S. S. Rubakhin, J. C. Jurchen, E. B. Monroe, and J. V. Sweedler, "Imaging mass spectrometry: Fundamentals and applications to drug discovery," *Drug Discovery Today*, vol. 10, p. 823–837, Jun 2005.

[14] A. Analytics, "Msi experiment visual (gif)," April 2024. Accessed on February 2025.

[15] E. K. Neumann, K. V. Djambazova, R. M. Caprioli, and J. M. Spraggins, "Multimodal imaging mass spectrometry: Next generation molecular mapping in biology and medicine," *Journal of the American Society for Mass Spectrometry*, vol. 31, p. 2401–2415, Jul 2020.

[16] D. Peralta and Y. Saeys, "Robust unsupervised dimensionality reduction based on feature clustering for single-cell imaging data," *Applied Soft Computing*, vol. 93, p. 106421, Aug 2020.

[17] P. W. Siy, R. A. Moffitt, R. M. Parry, Y. Chen, Y. Liu, M. C. Sullards, A. H. Merrill, and M. D. Wang, "Matrix factorization techniques for analysis of imaging mass spectrometry data," *2008 8th IEEE International Conference on BioInformatics and BioEngineering*, Oct 2008.

[18] M. Prasad, G. Postma, P. Franceschi, L. M. Buydens, and J. J. Jansen, "Evaluation and comparison of unsupervised methods for the extraction of spatial patterns from mass spectrometry imaging data (msi)," *Scientific Reports*, vol. 12, Sep 2022.

[19] R. Van de Plas, F. Ojeda, M. Dewil, L. Van Den Bosch, B. De Moor, and E. Waelkens, "Prospective exploration of biochemical tissue composition via imaging mass spectrometry guided by principal component analysis," *Biocomputing 2007*, Dec 2006.

[20] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, p. 211–218, Sep 1936.

[21] L. Mirsky, "Symmetric gauge functions and unitarily invariant norms," *The Quarterly Journal of Mathematics*, vol. 11, no. 1, p. 50–59, 1960.

[22] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Review*, vol. 53, p. 217–288, Jan 2011.

[23] M. Banf and T. Hartwig, "The reasonable effectiveness of randomness in scalable and integrative gene regulatory network inference and beyond," *Computation*, vol. 9, p. 146, Dec 2021.

[24] N. B. Erichson, S. Voronin, S. L. Brunton, and J. N. Kutz, "Randomized matrix decompositions using r," *Journal of Statistical Software*, vol. 89, no. 11, 2019.

[25] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," *Conference on Modern Analysis and Probability*, p. 189–206, 1984.

[26] N. Meinshausen, "Random projections and Johnson-Lindenstrauss lemma," Mar 2019. Lecture notes for Multivariate Statistics at ETH Zurich.

[27] N. B. Erichson, A. Mendible, S. Wihlborn, and J. N. Kutz, "Randomized nonnegative matrix factorization," *Pattern Recognition Letters*, vol. 104, p. 1–7, Mar. 2018.

[28] J. M. Spraggins, K. V. Djambazova, E. S. Rivera, L. G. Migas, E. K. Neumann, A. Fuetterer, J. Suetering, N. Goedecke, A. Ly, R. V. de Plas, and R. M. Caprioli, "High-performance molecular imaging with maldi trapped ion-mobility time-of-flight (timstof) mass spectrometry," *Analytical Chemistry*, vol. 91, no. 22, pp. 14552–14560, 2019.

[29] J. Kim and H. Park, "Fast nonnegative matrix factorization: An active-set-like method and comparisons," *SIAM Journal on Scientific Computing*, vol. 33, p. 3261–3281, Jan 2011.

[30] S. A. Vavasis, "On the complexity of nonnegative matrix factorization," *CoRR*, vol. abs/0708.4149, 2007.

[31] N. Gillis, "The why and how of nonnegative matrix factorization," *Regularization, Optimization, Kernels, and Support Vector Machines*, vol. 12, 01 2014.

[32] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, vol. 401, 1999.

[33] J. Kim, Y. He, and H. Park, "Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework," *Journal of Global Optimization*, vol. 58, p. 285–319, Mar 2013.

[34] T. Masuda, K. Yamada, T. Migita, and N. Takahashi, "A novel hals-based iterative algorithm for randomized nonnegative matrix factorization," *Proceedings of the 2023 International Symposium on Nonlinear Theory and Its Applications*, pp. 704–707, 9 2023.

[35] A. Cichocki, *Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation.* John Wiley, 2009.

[36] A. Cichocki and A.-H. Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE Transactions*, vol. 92-A, pp. 708–721, 03 2009.

[37] A. Varasteh, "Non-negative-matrix-factorization—implemented-in-python." https://github.com/ahmadvh/Non-Negative-Matrix-factorization---Implemented-in-python, 202.

[38] F. Sha, L. Saul, and D. Lee, "Multiplicative updates for nonnegative quadratic programming in support vector machines," *Advances in Neural Information Processing Systems*, vol. 15, 02 2003.

[39] J. Yang, O. Rübel, Prabhat, M. W. Mahoney, and B. P. Bowen, "Identifying important ions and positions in mass spectrometry imaging data using CUR matrix decompositions," *Analytical Chemistry*, vol. 87, p. 4658–4666, Mar 2015.

[40] Z. Drmač and S. Gugercin, "A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions," *SIAM Journal on Scientific Computing*, vol. 38, p. A631–A648, Jan. 2016.

[41] C. Boutsidis, M. W. Mahoney, and P. Drineas, "An improved approximation algorithm for the column subset selection problem," *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, Jan 2009.

[42] A. Chaudhry and E. Rebrova, "Learning nonnegative matrix factorizations from compressed data," 2024.

[43] Y. Zhang, L. Jin, F. Guo, X. Ni, Y. Zhao, Y. Cheng, and H. Wang, "Matrix factorization-based dimensionality reduction algorithms - a comparative study on spectroscopic profiling data," *Analytical Chemistry*, vol. 94, no. 39, pp. 13385–13395, 2022. PMID: 36130041.

[44] S. Hotta, *Ch. 14 Hermitian Operators and Unitary Operators*, p. 571–605. Springer, 2023.

[45] A. Sultonov, S. Matveev, and S. Budzinskiy, "Low-rank nonnegative tensor approximation via alternating projections and sketching," *Computational and Applied Mathematics*, vol. 42, p. 68, 2023.

[46] G.-J. Song and M. K. Ng, "Nonnegative low rank matrix approximation for nonnegative matrices," *Applied Mathematics Letters*, vol. 105, p. 106300, 2020.

[47] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, p. 1336–1353, Jun 2013.

[48] F. Yahaya, M. Puigt, G. Delmaire, and G. Roussel, "Faster-than-fast nmf using random projections and nesterov iterations," in *Proceedings of iTWIST: International Traveling Workshop on Interactions between Low-Complexity Data Models and Sensing Techniques*, (Marseille, France), November 21-23 2018.

[49] M. Tepper and G. Sapiro, "Compressed nonnegative matrix factorization is fast and accurate," *IEEE Transactions on Signal Processing*, vol. 64, p. 2269–2283, May 2016.

[50] B. Li, G. Zhou, and A. Cichocki, "Two efficient algorithms for approximately orthogonal nonnegative matrix factorization," *IEEE Signal Processing Letters*, vol. 22, p. 843–846, Jul 2015.

[51] Z. Yang and E. Oja, "Linear and nonlinear projective nonnegative matrix factorization," *IEEE Transactions on Neural Networks*, vol. 21, p. 734–749, May 2010.

[52] X. Chen, Y. He, and Z. Zhang, "Tight error bounds for the sign-constrained stiefel manifold," *SIAM Journal on Optimization*, vol. 35, p. 302–329, Jan 2025.

[53] J. Wen, I. M. Nasrallah, A. Abdulkadir, T. D. Satterthwaite, Z. Yang, G. Erus, T. Robert-Fitzgerald, A. Singh, A. Sotiras, A. Boquet-Pujadas, E. Mamourian, J. Doshi, Y. Cui, D. Srinivasan, I. Skampardoni, J. Chen, G. Hwang, M. Bergman, J. Bao, Y. Veturi, Z. Zhou, S. Yang, P. Dazzan, R. S. Kahn, H. G. Schnack, M. V. Zanetti, E. Meisenzahl, G. F. Busatto, B. Crespo-Facorro, C. Pantelis, S. J. Wood, C. Zhuo, R. T. Shinohara, R. C. Gur, R. E. Gur, N. Koutsouleris, D. H. Wolf, A. J. Saykin, M. D. Ritchie, L. Shen, P. M. Thompson, O. Colliot, K. Wittfeld, H. J. Grabe, D. Tosun, M. Bilgel, Y. An, D. S. Marcus, P. LaMontagne, S. R. Heckbert, T. R. Austin, L. J. Launer, M. Espeland, C. L. Masters, P. Maruff, J. Fripp, S. C. Johnson, J. C. Morris, M. S. Albert, R. N. Bryan, S. M. Resnick, Y. Fan, M. Habes, D. Wolk, H. Shou, and C. Davatzikos, "Genomic loci influence patterns of structural covariance in the human brain," *Proceedings of the National Academy of Sciences*, vol. 120, no. 52, p. e2300842120, 2023.

[54] S. M. Ha, A. Bani, and A. Sotiras, "Scalable nmf via linearly optimized data compression," *Medical Imaging 2023: Image Processing*, p. 26, Apr 2023.

[55] P. Zhu and A. Knyazev, "Angles between subspaces and their tangents," *Journal of Numerical Mathematics*, vol. 21, Jan 2013.

Glossary

List of Acronyms

IMS	Imaging Mass Spectrometry
MALDI	Matrix-assisted Laser Desorption/Ionization
DESI	Desorption Electrospray Ionization
SIMS	Secondary Ion Mass Spectrometry
LDI	Laser Desorption Ionization
SVD	Singular Value Decomposition
NMF	Nonnegative Matrix Factorization
AONMF	Approximately Orthogonal NMF
PNMF	Projective NMF
HALS	Hierarchical Alternating Least Squares
MU	Multiplicative Update
LS	Large Scale
RandHALS	Randomized HALS
RandMU	Randomized MU

List of Symbols

Abbreviations

$\mathbb{R}_+^{M\times N}$	Set of nonnegative real matrices of size M by N
$\mathcal{P}_+(X)$	Projection of X onto the nonnegative orthant
μ	Regularization parameter for orthogonality in AONMF and Proposed Methods
$\sum_{k=1}^{K}$	Summation from 1 to K with an index k
$\tilde{W}$	Compressed spatial factor matrix
$f(W,H)$	Objective function in NMF optimization with optimizers W and H
$H \in \mathbb{R}^{N\times K}$	Spectral Signature Factor Matrix
K	Fixed rank K specified in advance for rank-K factorization
L	Sketch Dimension $L = K + P$
P	Oversampling Parameter to increase sketch size slightly above fixed rank K, $L = K + P$
Q	Orthonormal Basis to the range space of sample matrix Y
Q_{AONMF}	Nonnegative Approximately Orthogonal Basis computed from AONMF
Q_{AONMF}	Nonnegative Approximately Orthonormal Basis computed from the Proposed Method
Q_{prop}	Nonnegative Basis from Proposed Method
Q_{qr}	Orthonormal Basis computed from QR decomposition
$W \in \mathbb{R}^{M\times K}$	Spatial Abundance Factor Matrix
$X \approx WH^T$	NMF decomposition of matrix X
$X \geq 0$	Element-wise nonnegative constraint on data matrix X
$X \in \mathbb{R}^{M\times N}$	Matrix X with M rows and N columns
$X_{\geq 0}$	Element-wise nonnegative Matrix
$X_{\ngeq 0}$	Matrix is not (entirely) element-wise nonnegative
X_+	Element-wise nonnegative projection: $\max(X, 0)$
X_-	Negative Entries in X