DISSERTATION

A GEOMETRIC DATA ANALYSIS APPROACH TO DIMENSION REDUCTION IN MACHINE LEARNING AND DATA MINING IN MEDICAL AND BIOLOGICAL SENSING

Submitted by

Tegan Halley Emerson

Department of Mathematics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: Michael Kirby
Co-Advisor: Chris Peterson

Margaret Chenney
Jennifer Nyborg

ABSTRACT

A GEOMETRIC DATA ANALYSIS APPROACH TO DIMENSION REDUCTION IN MACHINE LEARNING AND DATA MINING IN MEDICAL AND BIOLOGICAL SENSING

Geometric data analysis seeks to uncover and leverage structure in data for tasks in machine learning when data is visualized as points in some dimensional, abstract space. This dissertation considers data which is high dimensional with respect to varied notions of dimension. Algorithms developed herein seek to reduce or estimate dimension while preserving the ability to perform a specific task in detection, identification, or classification. In some of the applications the only property considered important to be preserved under dimension reduction is the ability to perform the indicated machine learning task while in others strictly geometric relationships between data points are required to be preserved or minimized. First presented is the development of a numerical representation of images of rare circulating cells in immunofluorescent images. This representation is paired with a support vector machine and is able to identify differentiating cell structure between cell populations under consideration. Moreover, this differentiating information can be visualized through inversion of the representation and was found to be consistent with classification criterion used by clinically trained pathologists. Considered second is the task of identification and tracking of aerosolized bioagents via a multispectral lidar system. A nonnegative matrix factorization problem arised out of this data mining task which can be solved in several ways including a ℓ_1-norm regularized, convex but nondifferentiable optimization problem. Exisiting methodologies achieve excellent results when internal matrix factor dimension is known but fail or can be computationally prohibitive when this dimension is not known. A modified optimization problem is proposed that may help reveal the appropriate internal factoring dimension based on the sparsity of averages of nonnegative values. Third, we present an algorithmic framework for reducing dimension in the linear mixing model. The mean-squared error of a statistical estimator of a component of the lin-

ear mixing model can be considered as a function of the rank of different estimating matrices. We seek to minimize mean squared error as a function of the rank of the appropriate estimating matrix and yield interesting order determination rules and improved results, relative to full rank counterparts, in applications in matched subspace detection and generalized modal analysis. Finally, the culminating work of this dissertation explores the existence of nearly isometric, dimension reducing mappings between special manifolds characterized by different dimensions. Understanding the analogous problem between Euclidean spaces provides insights into potential challenges and pitfalls one could encounter in proving the existence of such mappings. Most significant of the contributions is the statement and proof of a theorem establishing a connection between packing problems on Grassmannian manifolds and nearly isometric mappings between Grassmannians. The frameworks and algorithms constructed and developed in this doctoral research consider multiple manifestations of the notion of dimension. Across applications arising from varied areas of medical and biological sensing we have shown there to be great benefits to taking a geometric perspective on challenges in machine learning and data mining.

ACKNOWLEDGEMENTS

I would like to acknowledge my advisors Dr. Michael Kirby, Dr. Chris Peterson, and Dr. Louis Scharf. The opportunities you all afforded me in terms of research topics, conference travel, and access to your collective knowledge allowed me to grow as a scientist, positioned me to be competitive on the job market, and gain invaluable life experiences. I cannot thank you enough for these opportunities, your patience, and your support over the last several years.

My research advisor during my visiting research experience at MIT Lincoln Laboratory, Dr. Dimitris Manolakis, who taught me a great deal about digital signal processing and helped create the opportunity to work in a new and exciting environment.

My parents, Dawn and Bruce Emerson, who provided a seemingly endless source of emotional support throughout the rollercoaster that is the graduate student experience and proofread so many papers, abstracts, posters, and research statements for me.

My friends Julien Chaput, Tim Marrinan, Meghan Kahnle, Rosemary Grace, Sebastian McNab, Aaron Hill, Alison Aristoff, Logan Munoz, Maureen Kudola, Eric Truslow, Audrey May, Nick Landry, Sarah Lyons, and Lauren Nagel who provided me with support and community and ensured that I maintained something that resembled a life outside of graduate school. Thank you all for loving, laughing, cooking, drinking, eating, climbing, dancing, traveling, and sharing yourselves with me. To Julien, in particular, for so calmly and gracefully dealing with my crazy over the last year and a half.

My boyfriend, Mat Blodgett, who brought a shot of happiness and inspiration into my life that helped me push through to the end. Thank you for loving, supporting, and encouraging me.

Funding Acknowledgements: This doctoral research was partially supported by the National Science Foundation under Grants No. DMS-1228308, DMS-1322508, DMS-1115668, and DMS-1412674. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

Chapter 1
Introduction

What is data? Upon Googling the word there are a few different definitions that appear. The first definition is "facts and statistics collected for reference or analysis." Second, a definition arising from the field of computing pops up: "the quantities, characters, or symbols on which operations are performed by a computer, being stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media." Also for our consideration is the more philosphical definition:"things known or assumed as facts, making the basis of reasoning or calculation." As mathematicians, we often think about data from a perspective involving all three of the above definitions. Data is a collection of numerical representations which can be analyzed using mathematical techniques to reveal facts and patterns on which reasonable, and useful, models can be built.

Mathematicians can use numerical representations to visualize almost anything as a point in an abstract space. For example, a standard color image that people are familiar with can be thought of as a collection of points in a 3-dimensional space. That is to say a standard color image can be thought of as a collection of points with three coordinates: (x, y, z). Each pixel corresponds to a single point and the value of each coordinate is determined by the percent of red, green, and blue, respectively, in the pixel. By considering data as points existing in different spaces we can look for mathematical relationships between these points to better understand the data and extract information from it.

There are many different types of mathematical relationships one can consider. The methods of this doctoral research are rooted in looking for geometric relationships in data when considered as points in some space. Geometric Data Analysis (GDA) refers to the identification and exploitation of geometric structure in data to aid in machine learning tasks and knowledge discovery. Pattern analysis/recognition, computer vision, and manifold learning can all be considered as subsets of this overarching research area. When most people hear the word "geometry" they often start to picture triangles, squares, and circles that they may have encountered in a high school geome-

try course. The mathematical abstraction of these "shapes" is actually described using *topology*. Throughout this duration the term *geometric* will be used to describe the understanding of the ambient space/environment in which data exists. Questions aiming at understanding the structure of the space (e.g. "Is there a manifold underlying the data?"), relationships between data points (e.g. "Is most of the data close together in the space or scattered throughout?"), and dimension (e.g. "Are all the dimensions/coordinates of the data needed to perform the task of interest?") are all questions we consider to be geometric in nature. Questions regarding dimension are the focus of the research contained in this dissertation.

Techniques from GDA are frequently applied to big data. The expression "big data" has become a hot topic across many disciplines. From health care data including gene expression data, digital images from various medical imaging modalities, and patient history with lab results to commercial data including purchase history and internet click data, the need for techniques to mine "big data" has grown rapidly. There are two standard ways of describing the big-ness of data. First, the "big" can refer to the number of samples. Second, the "big" can refer to the dimensionality of the data. For example, many medical data sets like genomic data often have a small number of samples but with many different values attributed to each sample; small sample size but large dimension creating a "big data" set.

The "curse of dimensionality" refers to the challenges researchers face when working with data that is big with respect to dimension. These challenges manifest in a multitude of ways including data storage, insufficient sample sizes for statistical inference, data visualization and interpretation, and computational tractability. In order to combat the curse of dimensionality people look to the subfield of dimensionality reduction within the field of GDA. Dimensionality reduction can refer to techniques for visualizing data in lower dimensional spaces, transformations mapping data into a lower dimensional space to perform machine learning tasks, or identification of important dimensions (aka feature selection). The applications of this doctoral research are woven together by the thread of dimensionality reduction.

The core work comprising this dissertation is covered in Chapters 2-5 with one chapter dedicated to each of four different projects. Chapter 2 contains a project which used dimensionality reduction to identify and visualize differentiating structure between cell populations of interest. Chapter 4 describes a structured approach for dimensionality reduction by controlling bias and variance in statistical estimators by reducing the rank of matrices which produce the desired estimators. In Chapter 3 the challenge of nonnegative matrix factorization in the context of Frequency Agile Lidar (FAL) is described and a method for identifying a critical dimension is proposed. Lastly, in Chapter 5 the problem of producing nearly distance preserving maps between Grassmannian manifolds characterized by different/reduced dimensions is studied. A brief summary of each of these four projects concludes this introduction.

Feature Selection Based Visualization in Cancer Cell Detection: The task underlying this research was to develop a rotationally invariant feature vector for images of cells generated by immunofluoresent microscopy that would allow for classification of cell populations of interest. By exploiting the rotational invariance of the Fourier transform, we were able to generate a representation that produced high classification rates. Feature selection was performed by generating multiple classification models and looking at the relative weights of each feature across the models. Features that were frequently assigned high weights were selected. Through this feature selection and by leveraging the design of the feature vectors it was possible to reconstruct images of individual cells only using the selected features [1]. In this way, visualization of differentiating structure between cell types was obtained and these results could be interpreted and compared to manual pathologist classification criteria. Visualization through dimensionality reduction, in this case, was able to make direct connections between human and automated classification which is crucial for the broader acceptance of computer aided diagnostics in our society. This project was an analysis of data produced as a result of the masters work [2] and appeared as [1] in the journal "Computerized Medical Imaging and Graphics."

Nonegative Matrix Factorization Dimension Estimation in Frequency Agile Lidar: Frequency Agile Lidar (FAL) aims to passively detect and identify aerosolized bio-agents. The use of

aersolized bio-agents, like anthrax, in terrorism is considered to be a serious threat. Aerosolized particles are an optimal size for absorption into the lungs and many bio-agents capable of being aerosolized have no effective treatments or antidotes. The FAL system was designed to obtain measurements of the power of backscattered light of different frequency lasers based on the assumption that different aerosols will have characteristic spectral feature vectors (we say feature vector instead of signature as there are no established spectral signature libraries for aerosols). In 2012, a nonnegative matrix factorization problem was posed for this data that decomposes the FAL data into the product of two matrices; one containing information about the feature vector of each aerosol and another containing concentration information of each aerosol as a function of range [3]. The nonnegative matrix factorization problem from 2012 was based on minimizing the sum of squared error plus two ℓ_1-norm regularization terms: each of the matrix factors are ℓ_1-norm regularized. While conducting visiting graduate research at MIT Lincoln Laboratory, it was determined that modification of the objective function to exclude one of the ℓ_1-norm regularization terms, that of the feature vector matrix, yielded more desirable results. Additionally, we considered an extension of this problem wherein the goal was to determine the internal factoring dimension of the FAL model, i.e. the true number of aerosols present in the data. This internal dimension determination problem is a significant roadblock in this system becoming a online tool but is also present in many versions of the nonnegative matrix factorization problem. An approach to determine this dimension is presented which is based on the introduction of an additional regularization term penalizing for the ℓ_1-norm of a vector containing matrix row averages.

Reduced Dimension Estimators in Signal Processing Applications: A prominent tool in signal processing, due to its simple yet widely applicaple form, is the linear mixing model. In its general form, the linear mixing model assumes that data can be written as a linear combination of a known signal space, structured interference, and white noise. Previous attempts to apply dimensionality reduction to the linear mixing model have been done in two ways. First, some approaches focus on reducing the dimension of the data directly (mapping it to a lower dimensional space) and then applying existing algorithms within the new space. Second, looking for rank reduced matrices

to produce estimators have been considered when you are looking for optimality (as measured by some function) for a fixed reduced rank. Unlike these approaches, which have an ad hoc flavor to them, we developed a structured framework for dimensionality reduction based on reducing the dimension of estimators of components of the linear mixing model. The mean-squared error of a given estimator is formulated as a function of the rank of a particular estimating matrix and is minimized over all possible ranks. The resultant objective function yields strict and interpretable order determination rules for producing reduced dimension estimators by controling bias and variance through rank reduction. Consequently, unlike previous approaches which may in fact yield superior performance for a fixed reduced rank, our approach determines which rank is optimal for our particular performance metric. Application of this framework for classification of ground cover from hyperspectral imaging won a best paper award at the Workshop on Hyperspectral Imaging and Signal Processing: Evolutions in Remote Sensing conference [4]. New results using the framework in the context of generalized modal analysis is in preparation for resubmission to IEEE Transactions on Signal Processing[5]. In both applications, significant improvement in performance is achieved using a reduced dimension estimator as determined by our order determination rules.

Nearly Isometric, Dimension-Reducing Mappings Between Grassmannian Manifolds: A geometric property of data that people are often interested in preserving during dimension reduction is the set of pairwise distances between data points. Maps that preserve these pairwise distances exactly are called isometric mappings. Alternatively, we call a mapping nearly isometric when the distances are preserved up to a user specified distortion factor. A lot of theory has been developed related to the existence of these maps and how much you can reduce the dimension if you seek to preserve distances. Preservation of pairwise distances connects to the desire to have well posed mappings and prevention of collapsing data. One of the most well known lemmas related to the existence of these maps is the Johnson-Lindenstrauss Lemma [6]. This lemma guarantees the existence of a nearly isometric embedding into a Euclidean space of a (reduced) dimension as a function of the dimension of the distortion factor and the number of points in the

data set. A statistically based proof of the Johnson-Lindenstrauss Lemma was presented in [7] and has been recreated with added detail and explanation for the reader.

The final topic of this dissertation is an exploration and development of theory for the existence of such mappings where the data is not sampled from a Euclidean space but rather a special manifold. In particular, let $Gr(N, K)$ be the Grassmannian manifold corresponding to the set of all K-dimensional linear subspaces of $\mathbb{R}^N$. Given a sampling of points from $Gr(N, K)$ can we find nearly isometric mappings (with respect to metrics on the manifold) to $Gr(N, k)$ with $k < K$, $Gr(n, K)$ with $n < N$, or $Gr(n, k)$ with $n < N$ and $k < K$? If we are successful in producing such mappings, the potential applications that could benefit from this approach are wide and varied. For example, a set of medical values acquired at multiple times for a single patient can associate a subspace to the patient. Thus, comparison of different patients can be performed on $Gr(N, K)$ for some N, K. Mapping into a Grassmannian manifold characterized by lower values could provide insight into both critical times and critical medical values/measurements. Outside the biomedical arena, the two dimensions characterizing a Grassmannian manifold could be interpreted as spatial and temporal resolutions and reducing these dimensions could be considered resolution reduction. Exploratory computational results are presented and the development of the theory is structured as sequences of analagous statements to exisiting related proofs an problems. Of particular significance is establishment of a connection between the Grassmannian packing problem and nearly distance preserving maps as well as a proof of a theorem that could be critical in developing a proof, by construction, of a nearly distance preserving map through modification of an algorithm presented in [8].

Chapter 2
Fourier-Ring Descriptor to Characterize Rare Circulating Cells from Images Generated Using Immunofluorescence Microscopy

This chapter contains a continuation and analysis of work begun for my Master's thesis [2]. Detailed derivations and introductions to properties of the Fourier transform, which play a critical role in the numerical representation of the cells, can be found therein. What follows is a slightly modified version of the paper produced from this work entitled "Fourier-Ring Descriptors to Characterize Rare Circulating Cells from Images Generated Using Immunofluorescence Microscopy" that appeared in the journal *Computerized Medical Imaging and Graphics*[1].

This work highlights the use of dimensionality reduction via feature selection for interpretation of results and enrichment of cross disciplinary communications. Features, at their core, are numerical values used to represent an object. Sometimes these numerical values are highly interpretable (e.g. representing a person's face by a set of measurements like the distance between their eyes), but are more often built based on abstract mathematical properties (i.e. texture, curvature, gradients, etc.). In the age of big data, when we talk about "big-ness" with respect to dimension, we assign a very large number of features to each object under consideration. While this allows highly accurate, and often complex, models to be built, these models lack one very desirable property: interpretability. How does one interpret a model that computes a complex, non-linear combination of thousands of abstract values? The answer is simple: you don't. As mathematicians this can seem completely reasonable. However, as a patient receiving a diagnosis from such a model this may not seem as reasonable. The use of mathematics in biomedical applications, including computer aided diagnostics, must keep this in mind at all times.

Using *Fourier-Ring Descriptors* (FRDs) as a numerical representation of cells, we built several linear models to separate out the populations of cells of interest. Features were determined to be important if they were leveraged in all instances of the models that were trained. This reduced the number of features (feature selection) that were considered in the final model. Although these features were generated based on mathematical theory, they are interpretable and moreover can be

inverted to visualize what a cell would look like based on the selected subset of features. Selected feature based visualization allowed for fruitful cross discipline discussion and instilled confidence in the model based on its interpretability and agreement with expert pathologist based decision making processes.

The remainder of this chapter is formatted in agreement with [1]. Section 2.1 provides the bological context and motivation for the problem considered. Section 2.2 describes the materials, including data acquisition, and methods involved in the experiment described in Section 2.3. Neccessary theory and calculations are presented in Section 2.4. The results of the experiment described as well as a discussion of these results are provided in Sections 2.5 and 2.6, respectively.

2.1 Background

Counts of *Circulating Tumor Cells* (CTCs) have been correlated with outcomes in patients with tumors of epithelial origin including breast cancer, colorectal cancer, non-small cell lung cancer, and prostate cancer [9, 10, 11, 12, 13]. For these reasons, research has grown over the last decade. Clinically useful CTC counts have almost exclusively been generated using CellSearch^{TM} [14], which is an FDA approved enumeration method that relies on sample enrichment prior to analysis. Recent research has shown, however, that enrichment prior to analysis results in lower counts of CTCs and the potential loss of entire classes of disease-derived rare cells [15]. This deficiency diminishes the CTC count accuracy and hence lessens its impact as a diagnostic tool. On the other hand, newer methods that do not enrich prior to analysis have shown the potential to identify a larger and maybe more complete set of candidate cells but require significant computational analysis as well as human verification. In particular, the class of data generated without enrichment provides for the ability to optimize computational methods because of the large amount of data available, over twelve million cells per CTC test. Therefore, there is motivation to improve the level of automation of detection of CTCs on data generated without enrichment. Increasing the automation of detection of CTCs in this setting is the aim of the research presented here.

Over the last 20 years there has been a steady increase in the number of research papers published surrounding morphological cell image analysis [16]. It has been asserted that the five most significant roles of morphological cell analysis in medical imaging are malignant cell identification and cancer detection, following morphological changes during a cell cycle, cell classification, changes in morphology due to treatments, and morphometrical studies [16]. These roles are all strongly related to identifying a set of features that allow us to categorize candidate cells into these rules.

Cellular features are generally related to the geometry of the cell and properties of the interior of the cell, and carry a high-level of interpretability. A review of shape representations and description techniques for general images can be found in [17]. Geometric features of cells commonly include area, radii, perimeter, circularity, eccentricity, and irregularity [16, 18]. Internal features of a cell include the intensity, texture, and regularity of the nucleus. The combination of these geometrical and internal features have been employed in many pathological tasks, especially related to human epithelial cells [19, 20]. These features can be extracted from images and also visually inspected by a pathologist. However, due to the subjectivity involved in visual inspection, there has been an increase in the efforts to quantify these features using computers and to employ additional computer methods to support pathologists. A layout of one such image-guided decision support function is described in [21].

The set of features described above work well in many pathologist tasks, however interpretable features can have significant drawbacks. A major drawback is the reliance on segmentation. Geometric features are only as accurate as the ability to identify the comprehensive boundary of the cell. There have been several methods proposed for improving the automated segmentation of cell events in blood and tissue [22, 23, 24]. A second drawback can be the limitation of these features to differentiate cells. For example, there have not been successful classification methods for differentiating between CTCs in patients of different clinical status using these interpretable features alone. Additional discussion of technical and mathematical challenges of automated screening of epithelial cells can be found in [25]. Features that are generated without concern for interpretabil-

ity and are mathematical in nature sit in the realm of bioinformatics and are referred to as low-level features. Low-level feature methods have been employed in various cell classification tasks and perform with varying degrees of success. A comparison of different feature types together with varying classifiers is discussed in [26].

We have identified a set of low-level features that contain strongly differentiating structural information to separate CTC events of interest from white blood cells as well as identify distinct sub populations. Fourier-Ring descriptors (FRDs) extract features from cells by combining techniques from bioinformatics and computer vision. Additionally, FRDs contain interpretable information as well as indirectly containing many geometric features.

In the following sections we will first present the data acquisition process together with the composition of the data. Then, we will describe methods that we have considered, and highlight some of their challenges. Next, we will present our image representation technique and our classification structure for classifying a cell.

2.2 Material and Methods

2.2 Materials

Our analysis has been performed on images generated using the high definition circulating tumor cell (HD-CTC) assay developed by the Scripps Physical Sciences in Oncology Center (Scripps PSOC) as described in [15]. First, we will briefly describe how the data is acquired. Next, we will explain the current two-phase semi-automated algorithm employed at the Scripps PSOC to detect cells of interest. Last, we will define the types of cell populations included in the data set analyzed herein.

2.2 Data acquisition.

The assay developed at the Scripps PSOC produces images using automated fluorescence microscopy. Data evaluated in this paper was generated using three immunofluorescent stains: $Alexa555$ and pan cytokeratin to highlight epithelial cells, $Alexa647$ conjugated to an anti-CD45

antibody for WBC detection, and $DAPI(385nm)$ to stain for a nucleus. What separates this assay from many others is the lack of sample enrichment prior to imaging. Images are taken of the entire slide at 10X resolution by an inverted microscope. A set of images corresponding to a slide contains approximately three million nucleated cells, of which less than 0.01 percent are cells of high interest. Cells of high interest are currently detected using a two part semi-automated algorithm.

2.2 Phase I: Automated Detection.

Phase one of the current algorithm is computer automated and utilizes the medical imaging software ImageJ [27]. Cell centers are determined by detecting on the nuclear image channel and computing a center of mass described in [15]. A set of candidate cells of interest are output, based primarily on the intensity of the cytokeratin and CD-45 channels for any measured cell compared to all cells on the slide. Candidate cells are then passed to a technician or pathologist to manually classify.

2.2 Phase II of detection employed at Scripps PSOC: Manual Classification.

The second phase of the algorithm consists of the manual classification of candidate events of interest into one of six groups. The four groups of cells of interest are defined as follows:

1. **CTC-Candidate.** Cells that appear to have a high likelihood of being a CTC. Characterized by bright cytokeratin stain, an intact nucleus, and no CD-45 signal. Cells must be morphologically distinct from surrounding WBCs. This morphological difference typically manifests as a larger nucleus than the neighboring WBCs. These events will then be evaluated by a pathologist to confirm or reject this classification.

2. **CTC-small.** There exists a population of cells that has appropriate levels of cytokeratin expression to be considered a CTC but has an insufficient nuclear size relative to its surrounding white blood cells (WBCs). This population is considered to be a marginal population of CTCs.

3. **CTC-dim.** This population accounts for cells which have insufficient levels of cytokeratin expression to be considered CTCs but do have a nuclei that are significantly larger than neighboring cells. This population is considered to be a marginal population of CTCs.

4. **CTC-Ap.** A last marginal population of CTCs comprised of cells that appear to be apoptotic by identification of nuclear fragmentation or cytoplasmic blebbing.

The two remaining populations of cells are WBCs and imaging noise. This naming scheme is consistent with the cell populations presented in [11]. It is the goal of ongoing research to automate this phase of classification using the methods and experimental design presented here.

2.2 Composition of the data set.

The analysis herein was performed on a data set of cells which were hand selected for work done in [2]. This data set consists of one thousand cells: five hundred cells of interest and five hundred white blood cells. The five hundred cells of interest contains two hundred CTC-Candidate and one hundred each of CTC-Ap, CTC-dim, and CTC-Small. Cells in the data set were taken from 39 patients with diagnosed lung, breast, and prostate cancers (25 Lung, 7 breast, 7 prostate). This data set is by no means comprehensive, however, it does provide a proof-of-concept to advance this avenue of research.

2.2 Methods

We accomplish this classification by identifying a set of low-level features that can structurally differentiate between cell populations of interest. We will present the challenges of the discussed classification tasks and present our solution.

In the current assay used by the Scripps PSOC a single slide is imaged at 10X resolution. Although a technician can manually classify a cell at 10X resolution, the low resolution fails to capture textural variation within each cell channel. This lack of texture, in turn, makes the implementation of gradient-based feature extraction methods impractical. It is impractical to increase the resolution of the images since an n-fold increase in resolution increases the number of images

generated from each channel by a factor of n^2. Additionally, low resolution results in s single cell comprising a small area within an image. The small size of the object makes the extraction of patch-based features unnecessary. Alternatively, treating the image of a single cell as a sole patch requires accurate segmentation of the cell. However, because of the frequency of cell overlap and noise from staining in the cytokeratin channel and CD-45 channel it is challenging to determine comprehensive cell boundaries. Furthermore, the cells themselves carry associated challenges. First, a representation technique which is interpretable and captures characteristics of the cells (including size of the nucleus, levels of cytokeratin expression, circularity, and uniformity of the nucleus, for example), is ideal. Next, a cell has no “correct" orientation which makes it imperative to have an image representation which is rotationally invariant.

Concentric circles have been proposed to handle this issue of rotational invariance in other tasks. Circles have been employed to detect objects in multiple rotations of complex colored images [28]. Additionally, in [29] we see the use of concentric rings combined with wavelet transforms for pattern matching. In the method described in [29], a single representative value for each ring is computed, then a wavelet transform of the series of representative values is computed. More recently we see another method involving the sampling of image values along concentric rings, which was developed with pathological tasks in mind [30]. Although the method being applied to these pathological tasks has performed very well, it requires an exhaustive codebook search method for pattern matching [31]. Thus, we have chosen to combine the desirable components of each method and employ concentric rings together with a transformation while not overly restricting the quantity of information taken from each ring.

The Fourier-Ring Descriptor (FRD) is based on the Fourier transformations of concentric rings about the nucleus of a cell. FRDs treat the image of a cell as a single patch, have limited reliance on segmentation, are rotationally invariant, and the features can be visualized and carry a high level of interpretability. Our solution does not require comprehensive cell boundaries in all channels, but rather only the computation of the center of mass of the cell nucleus. Due to the nuclear channel being the cleanest of the monochrome images, this greatly reduces the reliance on whole

cell segmentation. Also, the amplitude spectrum of a Fourier transform is rotationally invariant, and thus the amplitude spectrum of the Fourier transform of a ring is also rotationally invariant[1]. Furthermore, by using concentric rings about the nucleus of the cell we are able to obtain size and morphological information for a cell based on the presence, or lack there of, of features on rings of particular radii. Details of the generation of FRD follow in the Experimental section.

2.3 Experimental

There are two components of our experiment that will be discussed in this section. First, we will discuss our image representation, the Fourier-Ring Descriptor, in detail. Next, we will discuss our classification structure and how we implement the structure together with our image representation technique.

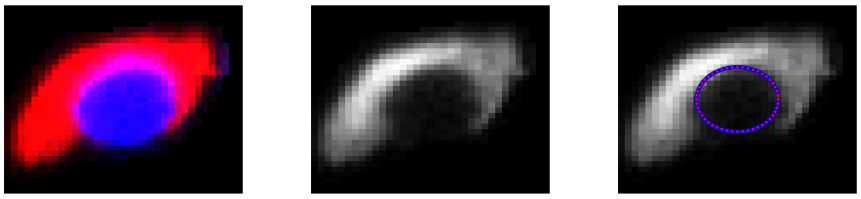
2.3 Fourier-Ring Descriptor

The Fourier-Ring Descriptor (FRD) is an image representation technique based on the Fourier transform of concentric rings constructed about the center of the nucleus of the cell. We first determine a cell center by computing a center-of-mass for each nucleus that is detected in the DAPI image channel using the medical image processing software ImageJ [27]. Once we have identified the center of a cell, we begin to construct concentric rings that are centered on the cell center.

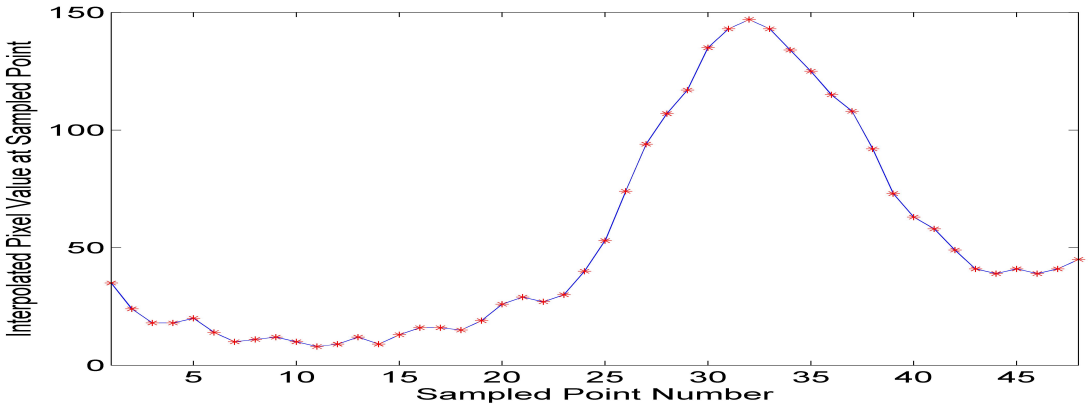
We use sixteen concentric rings to generate our image representation. The number of rings was determined by experimentation as discussed in [2]. The steps to generate our image representation are as follows:

i) We sample a number of evenly-spaced points along a ring of a given radius. An example of the location of these points on a ring of radius six pixels is shown in Figure 2.3.1(c).

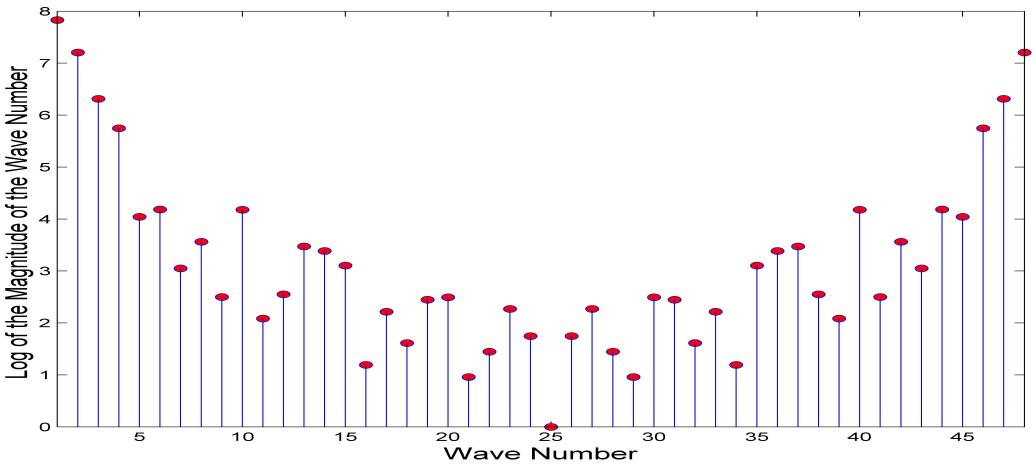
[1] Additional information about the rotational invariance can be found in the Theory/Calculations section



(a) **(b)** **(c)**



(d)



(e)

Figure 2.3.1: (a) Shows a composite image of a CTC-Candidate. (b) Shows the monochromatic cytokeratin channel image of the CTC-Candidate in (a).(c) Shows the monochromatic cytokeratin channel image of a circulating tumor cell with the ring of radius six pixels, oriented around the center of the nucleus of the cell, is overlaid in blue. The locations of the sampled points are evenly spaced along the indicated ring and are shown in red. There are 48 points sampled along this ring. (d) Shows a plot of the interpolated image values at the sampling points along the ring of radius six pixels shown (c). Each red point shows an interpolated image value. The image values can be between 0 and 255 as they are taken from $8-$bit JPEGs. Lastly, (e) is the amplitude spectrum of the Fourier transform of the curve shown in (d).

ii) Image values at each of the points shown in Figure 2.3.1(c) are interpolated. This set of interpolated image values determines a periodic curve as shown in Figure 2.3.1(d).

iii) Given the periodic curve of image values, we then perform a Discrete Fourier Transform (DFT) of the image values and keep the magnitude of the Fourier transform. An example of the amplitude spectrum of a single ring is shown in Figure 2.3.1(e).

iv) The process of sampling points along a ring and computing the DFT of the interpolated image values at the sampled points is repeated for the sixteen different concentric rings in a single monochrome image channel. Examples of the amplitude spectrums for the sixteen different rings in a channel can be seen in Figures 2.3.2, 2.3.3, and 2.3.4.

v) The amplitude spectrums of all rings from a single channel are then concatenated into a single vector according to increasing radius. This process is then repeated in the remaining image channels and the concatenated amplitude spectrums from within a channel are concatenated with the other channels in the following order: Cytokeratin, CD-45, DAPI.

The number of points sampled along a ring is scaled linearly in accordance with the increase in circumference. Eight points were selected for the ring of radius one based on the maximum number of distinct pixels a ring of radius one could encounter. Image values at the sampled locations are determined using cubic interpolation which causes highly-associated values across neighboring rings, but this redundancy is minimized by enforcing sparse feature selection in the classifier. Given that we start with eight sampled points on the inner most ring and scale linearly over the sixteen rings, we obtain a total of 1088 features from a single image channel, and a total of 3264 features over all three channels. Each feature of the FRD corresponds to the magnitude of a single frequency along a specific ring.

2.3 Classification Structure

The data analyzed for this paper contains five populations of cells. There are white blood cells, CTC-candidates, CTC-Aps, CTC-dims, and CTC-small. We refer to CTC-candidates, CTC-

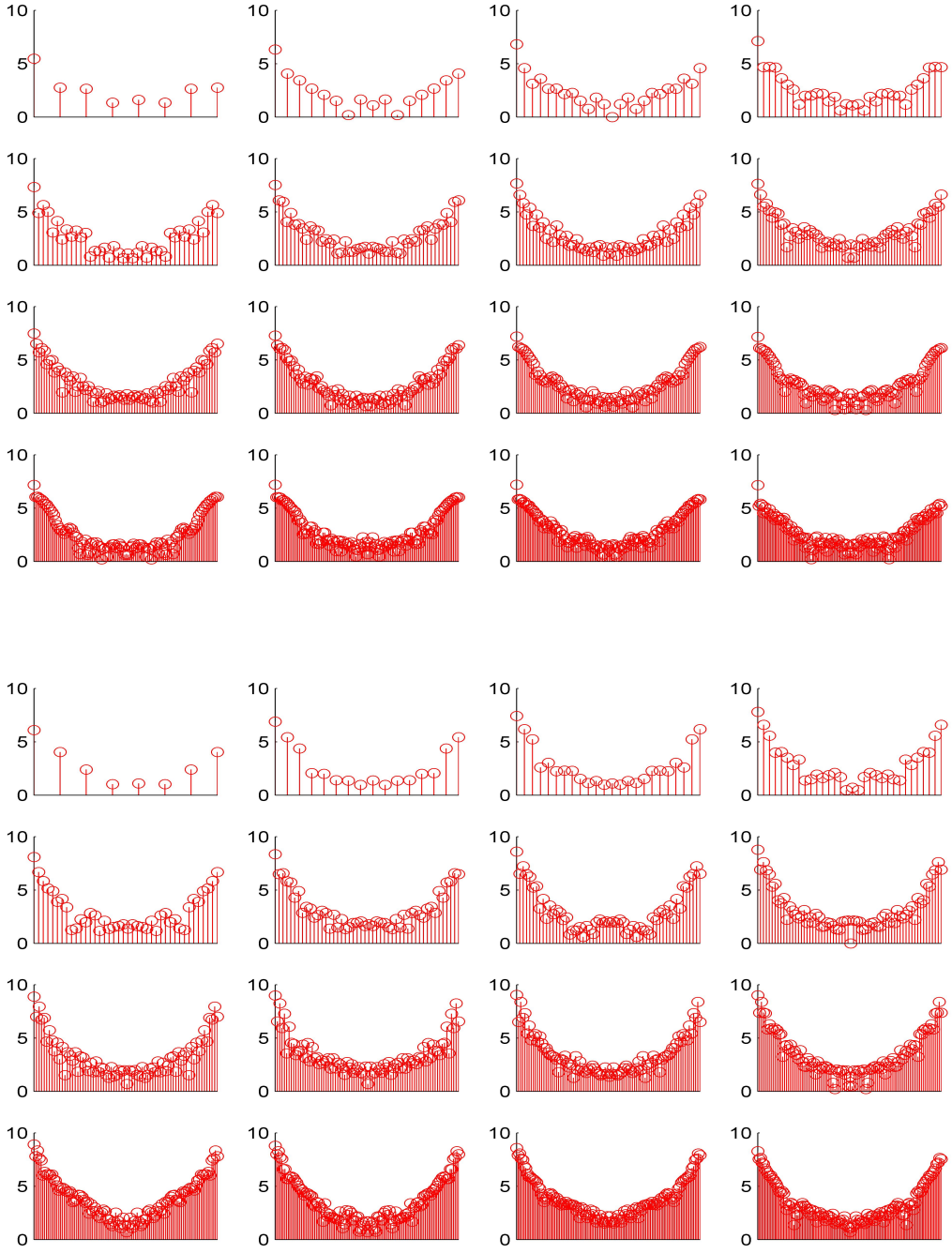


Figure 2.3.2: The top figure shows all 16 rings in the cytokeratin channel for the closest-to-average CTC-Candidate in our data set while the bottom figure shows all 16 rings for the farthest-from-average CTC-Candidate in our data set. The $y - axis$ is the log of the amplitude and the $x - axis$ corresponds to the wavenumber. A image value curve with m points will generate an amplitude spectrum for m frequencies and we recall for the ring of radius r there will be $8r$ image values sampled. The radius increases left to right and top to bottom.

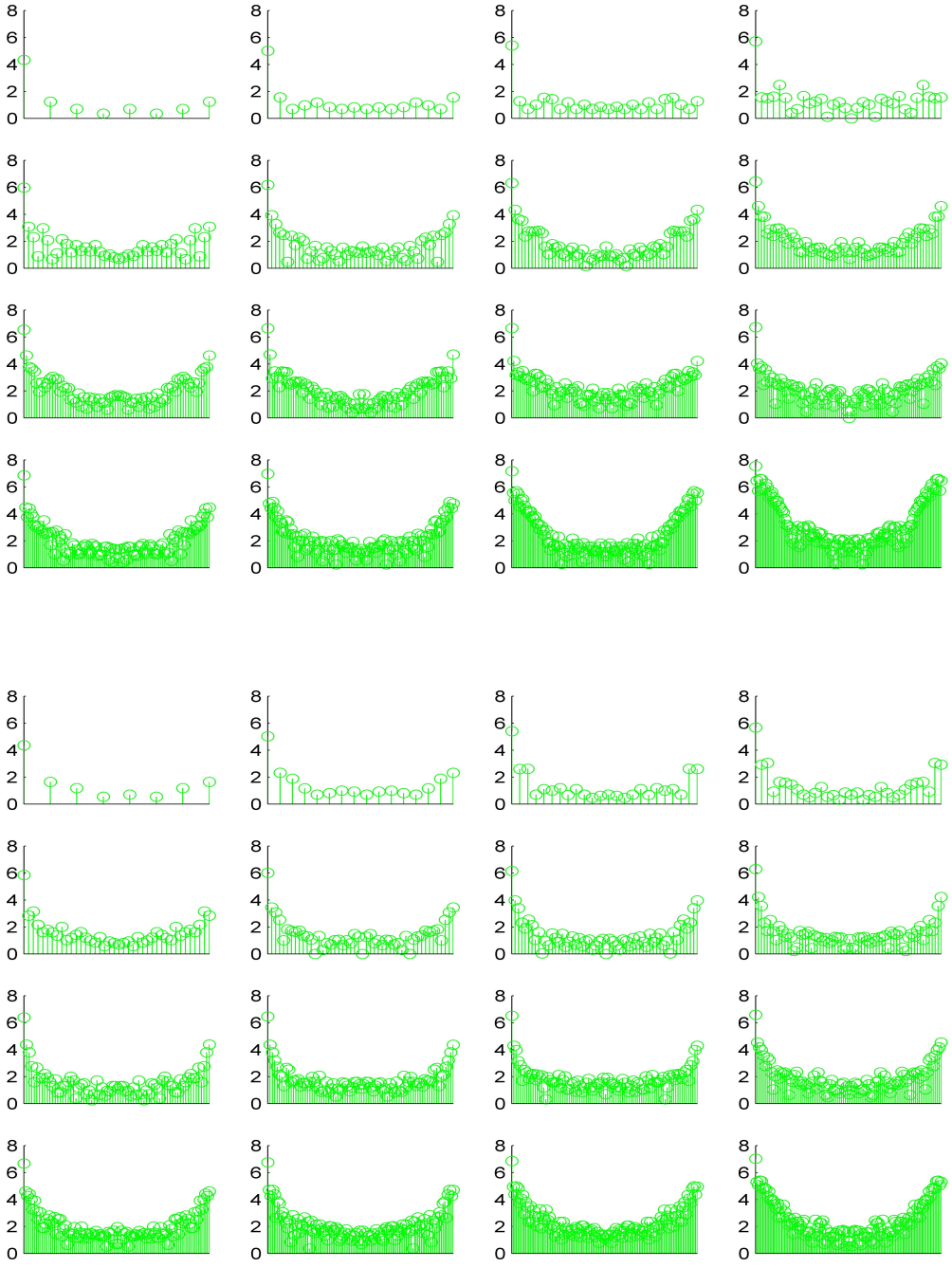


Figure 2.3.3: The top figure shows all 16 rings in the CD-45 channel for the closest-to-average CTC-Candidate in our data set while the bottom figure shows all 16 rings for the farthest-from-average CTC-Candidate in our data set. The $y - axis$ is the log of the amplitude and the $x - axis$ corresponds to the wavenumber. A image value curve with m points will generate an amplitude spectrum for m frequencies and we recall for the ring of radius r there will be $8r$ image values sampled. The radius increases left to right and top to bottom.

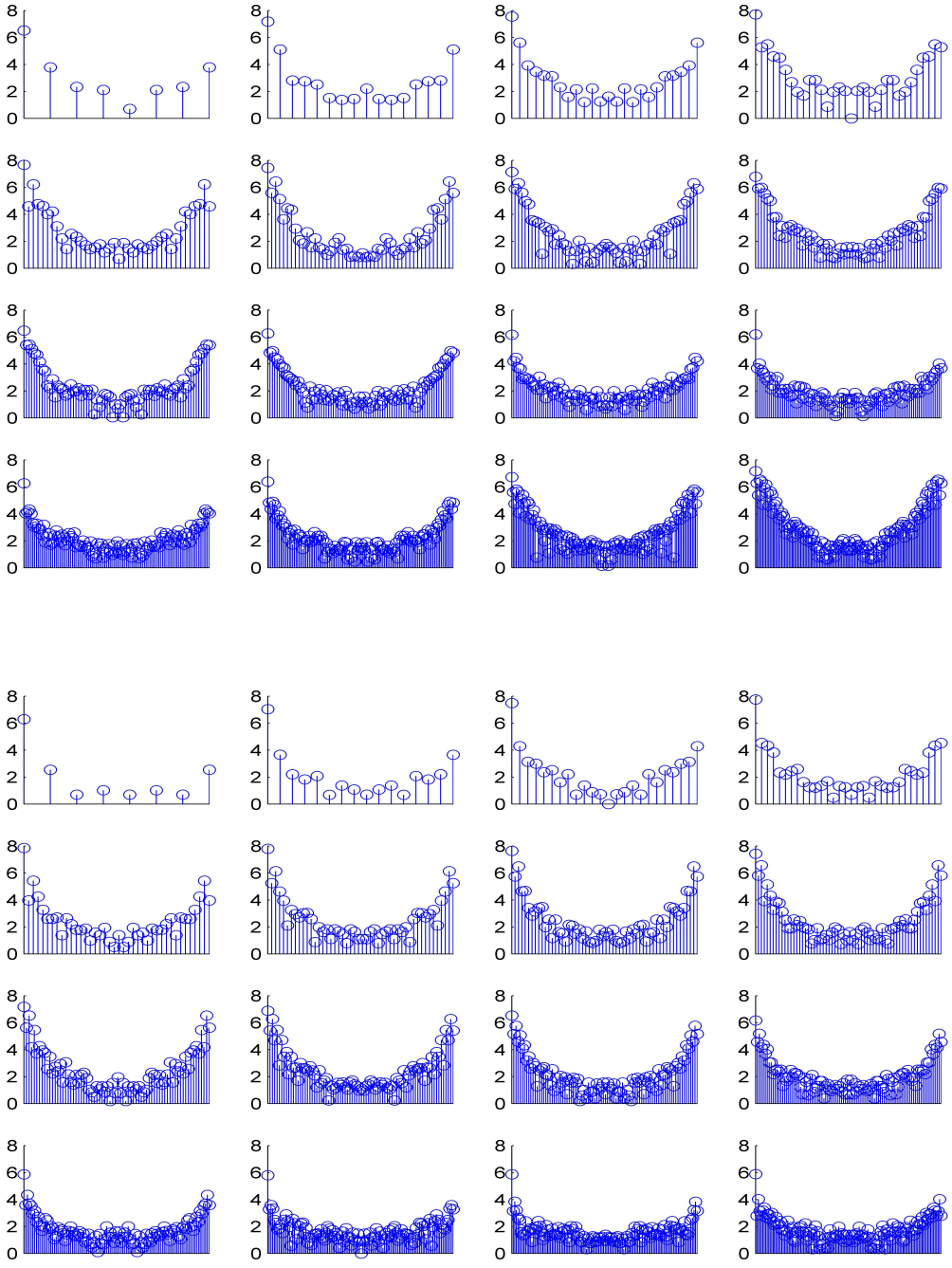


Figure 2.3.4: The top figure shows all 16 rings in the DAPI channel for the closest-to-average CTC-Candidate in our data set while the bottom figure shows all 16 rings for the farthest-from-average CTC-Candidate in our data set. The $y - axis$ is the log of the amplitude and the $x - axis$ corresponds to the wavenumber. A image value curve with m points will generate an amplitude spectrum for m frequencies and we recall for the ring of radius r there will be $8r$ image values sampled. The radius increases left to right and top to bottom.

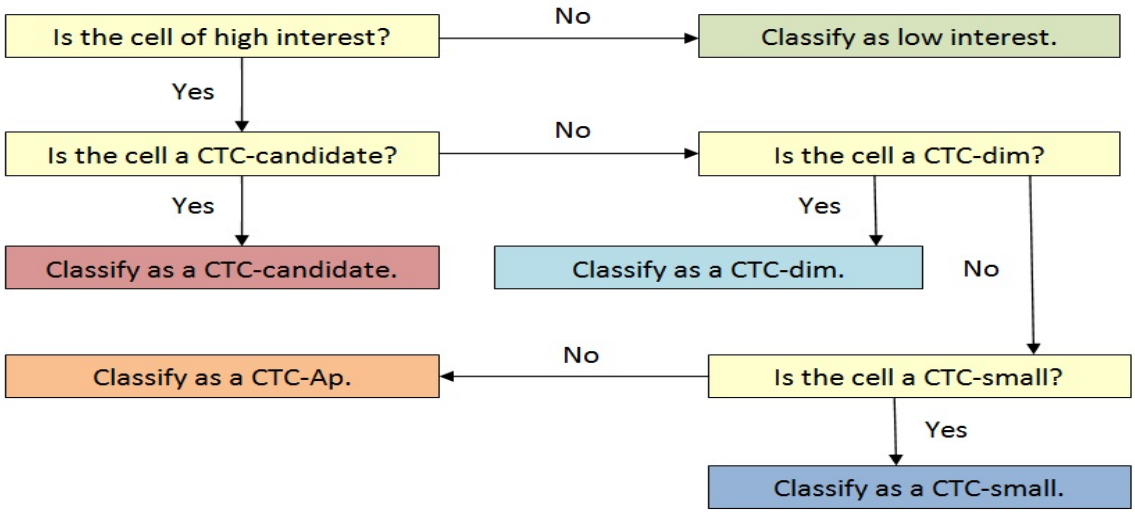


Figure 2.3.5: Schematic of the decision tree classification structure we are building based on strongly-differentiating structural features of cells. The answer to each question is determined by a support vector machine classifier.

Aps, CTC-dims, and CTC-smalls as populations of interest. As previously described, we consider CTC-Aps, CTC-dims, and CTC-smalls to be marginal populations of CTC-candidates. Initially we want to be able to separate the populations of interest from white blood cells. Separation of white blood cells from populations of interest can first be whittled-down using empirical property values obtained by the Scripps PSOC. Once the population of cells we are looking at has been reduced we can implement a decision tree classification structure. In this decision tree we first want to say with confidence whether or not a given cell is interesting. Once it has been identified as interesting we then ask whether or not the cell is a CTC-Candidate, our most important population of interest. We then proceed to subdivide the populations according to the schematic shown in Figure 2.3.5.

We have chosen to use a decision function determined by an l_1 regularized, l_2 loss function linear support vector machine classifier as implemented by LIBLINEAR [32]. A discussion of the decision function can be found in Section **??**. By implementing an l_1 regularized support vector machine we force sparsity of features required for classification. It is desirable for us to encourage sparsity both to limit the number of features we must interpret as well as to account for the potential

Figure 2.4.1: This figure shows the result of rotating a cell 60 degrees counterclockwise from the original orientation of the cell. The image on the left the cell in its original orientation on the slide image with the ring of radius 8 pixels overlaid while the right image is the numerically rotated image of the original cell generated using bicubic interpolation, also with the ring of radius 8 pixels overlaid.

of over-sampling in the generation of the FRD. By using a support vector machine classifier, we have a way of communicating our confidence in the classification of an event. Cells which are closer to a separating hyperplane may be misclassified while cells far away from the hyperplane may be extremal cases within their class.

2.4 Theory/Calculations

In this section we will address two concepts. First, we have stated that the FRD method is rotationally invariant. We will provide a graphical proof of the rotational invariance of a single FRD which can then be extended to show the invariance of the entire descriptor. Next, we will discuss the type of the decision function that we have implemented at each step in our decision tree classification structure. A description of the optimization problem to be solved and the parameters and variables involved are also included.

2.4 Rotational Invariance of FRDs

Here we present graphical proof of the rotational invariance of the amplitude spectrum of the Fourier transform in the context of our application. A rigorous proof of the rotational invariance of the Fourier transform can be found in [33] or [2].

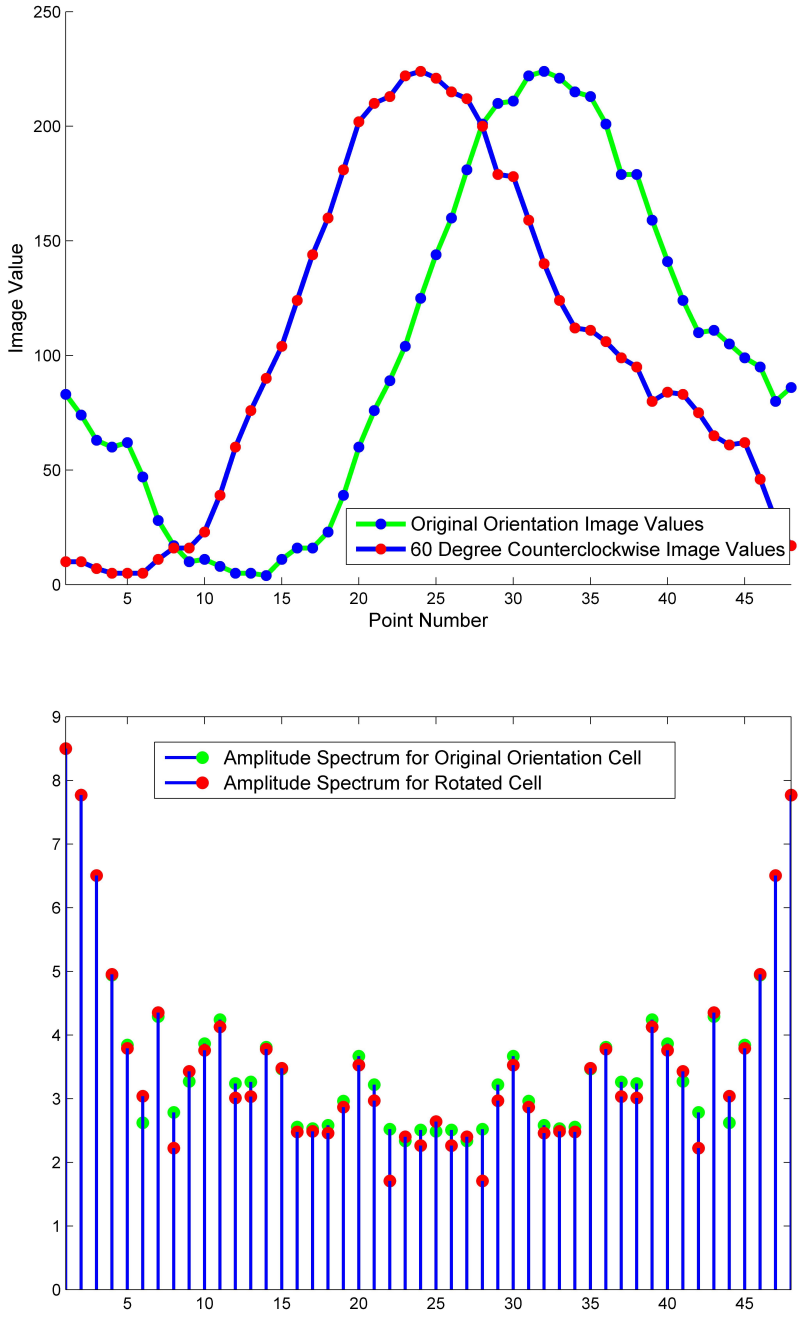


Figure 2.4.2: These figures demonstrate the rotational invariance of the amplitude spectrum. On the top we see the set of image values interpolated along the ring of radius 8 pixels for both the rotated and unrotated versions of the cell. The bottom figure shows the log of the amplitude spectrums for each of the two sets of image values.

Figure 2.4.1 provides us with two images: the first of a cell represented in the cytokeratin channel in its original orientation on the slide image, the second shows the same cell in the cytokeratin channel rotated 60 degrees counterclockwise. For this didactic example, we employed a numerical rotation function that utilizes bicubic interpolation. Next, in Figure 2.4.2(a), we see the curves of interpolated image values along the ring of radius 8 pixels for both the rotated and unrotated versions of the cell. We see that the two image value curves are simply shifted versions of one another, minus slight numerical error. Furthermore, in Figure 2.4.2(b) we see the amplitude spectrums of the two image value curves. From Figure 2.4.2(b), we see that the two amplitude spectrums are nearly identical minus the effect of numerical artifacts present from the rotation. The combination of Figures 2.4.1 and 2.4.2 demonstrates that a single FRD generated for a cell is invariant to the orientation of the cell on the slide. The full FRD representation of a cell amounts to stacking the amplitude spectrums for each of the rings inside a channel and then concatenating the channels. Thus, for two orientations of the cell we have shown that the amplitude spectrum for each ring is rotationally invariant and consequently the stacking of those amplitude spectrums will result in the same channel representation for a cell no matter the rotation.

2.4 Decision Function

A linear support vector machine (SVM) aims to define a decision function based on a hyperplane that separates data into two half spaces. Data is classified by which side of the separating hyperplane a datum falls on. We define the separating hyperplane by the equation $y(x) = w^T x + b$. In a classification task involving two classes of data with the standard ± 1 labeling, the classification of a novel data point $\hat{x}$ is determined by

$$D(\hat{x}) = \begin{cases} \hat{x} \in C^+ & if \quad w^T\hat{x} + b \geq 0 \\ \hat{x} \in C^- & if \quad w^T\hat{x} + b \leq 0 \end{cases}$$

where C^+ corresponds to the class of positive samples and C^+ corresponds to the class of negative samples.

In all of the classification tasks reported here we have implemented the linear, l_1 regularized, l_2 loss function SVM as implemented by LIBLINEAR [32]. The l_1 regularized, l_2 loss function linear SVM is defined by minimizing

$$\min_w \|w\|_1 + C \sum_{i=1}^{n} \max(0, 1 - y_i w^T x_i)^2. \tag{2.4.1}$$

In the cost function shown in (1) there are several terms to be defined. First, *w* is defined to be the normal vector to the linear separating surface. Next, x_i is a data point with corresponding label y_i. When trying to separate two groups of data, the standard convention is to label one class as +1 and the other with -1. Lastly, in (1) we see the variable *C* which is a parameter that determines the contribution of the loss function to the cost function. All accuracies presented here in have been generated using *C*= $\frac{1}{2}$. We refer to this formulation as an "l_2 loss function SVM" based on the second component of the cost function referred to as the loss function. A loss function penalizes for generating a hyperplane that misclassifies events. In particular, an l_2 loss function penalizes more for points that are strongly misclassified and less for those that hover near the hyperplane. This formulation of an SVM is referred to as l_1 regularized because of computing the 1-norm on the vector *w* inside the cost function. Using the 1-norm of *w* encourages sparsity in the normal vector. This, in turn, promotes sparsity of features used in a classification task due to the definition of the decision function.

2.5 Results

Our results can be divided into two categories: quantitative and qualitative. In our quantitative results section we will discuss the accuracy of the different decision functions for each branch of our previously-discussed decision tree. Next, in the qualitative results section we will provide reconstructions of cells using features selected by the decision functions in 95% of trials for a given classification task.

2.5 Quantitative Results

In accordance with the decision tree structure outlined in the Classification Structure section, we need to determine decision functions that can first differentiate between white blood cells (WBCs) and events of interest (EOI). Next, we want to be able to separate the EOI into CTC-Candidate and all marginal populations. Third, we want to separate CTC-dim from CTC-Ap and CTC-Small within the marginal populations. Last, we want to separate CTC-Small from CTC-Ap. In Table 2.1 we see that we obtain the accuracy for each decision function in the classification tree to be $99.52 \pm 0.48\%$, $92.16 \pm 2.70\%$, $89.76 \pm 3.89\%$, and $82.48 \pm 5.11\%$ respectively. The classification tasks become more challenging and have larger standard deviations the farther down the decision tree we move. Of particular interest is the classification task separating CTC-Ap from CTC-Small which suggests higher levels of overlap between these two populations which concures with the Scipps PSOC visually determined definitions of these two cell populations. Given this result, we decided to also explore all other pairwise classification tasks, the results of which are also shown in Table 2.1. A discussion of these results and their connection to pathologist experience is included in the discussion section.

Also shown in Table 2.1 are the number of features that were selected as important for classification by enforcing sparsity in 95% of the trials run. In each trial the data for each class in the classification task is randomly partitioned to have 75% used for training and the remaining 25% for testing, as well as matching the size of the two groups. There does not appear to be a connection between the accuracy of a given classification task and the number of features selected in 95% of the trials. Table 2.1 also highlights an important fact about the CTC-Ap cells. All of the pairwise classification tasks separating CTC-Aps from another cell of interest have the lowest pairwise classification accuracies. This reflects the biological property of this class. CTC-Aps are cells undergoing apoptosis. Apoptosis involves many stages and consequently has the largest variation within the populations of interest as the cells likey represent different events in the apoptotic cascade.

In evaluating the significance of the information provided in Table 2.1, it is important to acknowledge what the number of reoccurring features means. A large number of reoccurring features means that the classification task is more robust to the training set, while a small set of overlapping features means there is more variation in the classifiers. The features that do reoccur with high frequency, however, provide information as to what are consistently important features. To evaluate what these results reveal, we have broken down the distribution of the reoccurring features across each channel for each channel for all classification tasks in Tables 2.2, 2.3, and 2.4. In these tables, we have separated a cell into four regions. First, the inner region refers to the area contained within the first four rings, a circle of radius 4 pixels oriented about the cell center. Second, the inner-center of a cell corresponds to the area after the fourth ring up to, and including, the eighth ring. Next, the outer-center of the cell is the area after the eighth ring up to, and including, the twelfth ring. Last, we refer to the area after the twelfth ring up to, and including, the sixteenth as the outer cell.

Table 2.2 highlights some very interesting observations. First, from the first row we note that the distinguishing information between CTC-Candidates and other cells of interest, in the cytokeratin channel, is contained outside of the the inner-region of the cell. A second observation from this table is that in all of the pairwise classification tasks including CTC-Candidates, the outer-regions (outer-central and outer) contain a higher number of important features for distinguishing CTC-Candidates than the inner-regions. Third, we notice that the pairwise classification tasks involving CTC-Ap have the highest number of cytokeratin features, and that they occur in significant quantity in all of the cell regions. Fourth, we see that no classification task involving CTC-dim as an individual population have a frequently reoccurring cytokeratin feature in the inner-cell. Next, we note that the distribution of reoccurring features are nearly identical for the CTC-Candidate versus CTC-Ap and CTC-Candidate versus CTC-small classification tasks. Additionally, we see that the last four rows of the table contain the smallest number of cytokeratin features reoccurring.

From Table 2.3 we can draw additional conclusions. Of particular note is the non-trivial nature of this table. This table shows that, despite what we initially thought, the CD-45 channel plays a role in many of the classification tasks involving cells of interest. For example, in the row describ-

ing CTC-Ap vs. CTC-Small we see the highest number of CD-45 features. This could suggest one of several hypotheses. First, since CD-45 is a characteristic of white blood cells, does the presence of outer-region CD-45 features suggest frequent overlap with white blood cells for one of the two populations? Second, does the presence of the CD-45 suggest that our method is capturing features that support the knowledge that apoptotic cells non-specifically bind to antibodies as dying cells may have more epitopes exposed? Third, are there existing sub-populations of these cells of interest which express non-trivially in the CD-45 channel?

Finally, we consider the significance of DAPI features, as shown in Table 2.4. Similar to Table 2.2, we can immediately note the small number of features located in the inner-cell. Next, we note that when separating CTC-Candidates from all other cells of interest, or from either CTC-Ap or CTC-Small in pairwise classification, the bulk of the reoccurring features occur in the inner-central region of the cell. Also, we observe that CTC-dim versus all other marginal and versus CTC-Small show heavily weighted central-regions of the cells. Furthermore, we see that the outer-region of the cell, in the DAPI channel, is the most heavily weighted region in three classification tasks: CTC-Candidate vs. CTC-dim, CTC-Ap vs. CTC-dim, and CTC-Ap vs, CTC-Small.

The above observations reinforce many of the characteristics currently considered in manual classification. CTC-Candidates, for example, are identified, in part, by the intensity of cytokeratin, size, and cytokeratin which encompasses the over sized nucleus. This is reflected in the significance of the inner-central DAPI features together with the heavily weighted outer-region cytokeratin features. Additionally, CTC-dims are characterized by their large nucleus and low cytokeratin expression. Thus, in the the CTC-Candidate versus CTC-dim classification the presence of heavily weighted outer-region DAPI and cytokeratin features is in agreement with manual classification. A CTC-Candidate will have have cytokeratin in the outer-region of the cell but limited DAPI, while a CTC-dim will have limited cytokeratin expression overall but strong DAPI expression in the outer-region of the cell. These reflections of characterizations used in manual classification show that FRDs capture both interpretable, significant and biologically relevant structural information pertaining to different classification tasks.

Table 2.1: The first column of the table states the classification task of interest. Each classification task was run 25 times where 75% of the data of each class was randomly selected and used for training, while the remaining 25% was used for testing. In the event that the size of the classes differ, we randomly select samples from the larger class to match the size of the smaller class and then separate into 75/25 partitions. The second column shows the number of features (NOF), out of 3264, that are selected as important for the classification task in 95% of the trials for the given task. The final column gives the overall accuracy on a given classification task.

Classification Task	No. of Features	Accuracy
WBC vs Events of Interest	85	99.52± 0.48
CTC-Candidate vs. All other	122	92.16±2.70
CTC-Candidate vs. CTC-Ap	63	92.64±3.35
CTC-Candidate vs. CTC-dim	84	93.76±2.79
CTC-Candidate vs. CTC-Small	63	93.20±3.16
CTC-Ap vs. All Other	44	81.20 ± 5.26
CTC-Ap vs. CTC-dim	110	89.50±3.60
CTC-Ap vs. CTC-Small	167	82.48±5.11
CTC-dim vs. All Other	75	86.40 ±4.32
CTC-dim vs. Other Marginal	55	89.76±3.89
CTC-dim vs. CTC-Small	76	96.00±2.77
CTC-Small vs. All Other	48	87.12±5.42

Table 2.2: This table shows the break down of the location of the reoccurring features in a given classification task within the cytokeratin channel as a fraction of the total number of reoccurring features. Inner refers to occurring within the first four rings. Inner-central refers to occurring within the fifth through eighth rings. Outer-central refers to occurring within the ninth to twelfth rings. Lastly, outer refers to occurring within the thirteenth to sixteenth rings.

Classification Task	Inner	Inner-Central	Outer-Central	Outer
CTC-Candidate vs. All other	1/122	13/122	16/122	22/122
CTC-Candidate vs. CTC-Ap	0/63	4/63	11/63	10/63
CTC-Candidate vs. CTC-dim	0/84	5/84	19/84	12/84
CTC-Candidate vs. CTC-Small	0/63	3/63	12/63	7/63
CTC-Ap vs. All Other	3/44	9/44	8/44	4/44
CTC-Ap vs. CTC-dim	10/110	29/110	25/110	13/110
CTC-Ap vs. CTC-Small	12/167	24/167	16/167	15/167
CTC-dim vs. All Other	0/75	8/75	6/75	5/75
CTC-dim vs. Other Marginal	0/55	5/55	3/55	4/55
CTC-dim vs. CTC-Small	0/76	1/76	8/76	8/76
CTC-Small vs. All Other	2/48	6/48	2/48	7/48

Table 2.3: This table shows the break down of the location of the reoccurring features in a given classification task within the CD-45 channel as a fraction of the total number of reoccurring features. Inner refers to occurring within the first four rings. Inner-central refers to occurring within the fifth through eighth rings. Outer-central refers to occurring within the ninth to twelfth rings. Lastly, outer refers to occurring within the thirteenth to sixteenth rings.

Classification Task	Inner	Inner-Central	Outer-Central	Outer
CTC-Candidate vs. All other	4/122	6/122	11/122	16/122
CTC-Candidate vs. CTC-Ap	3/63	4/63	4/63	5/63
CTC-Candidate vs. CTC-dim	4/84	4/84	5/84	6/84
CTC-Candidate vs. CTC-Small	2/63	4/63	4/63	7/63
CTC-Ap vs. All Other	0/44	0/44	5/44	1/44
CTC-Ap vs. CTC-dim	0/110	0/110	5/110	5/110
CTC-Ap vs. CTC-Small	0/167	9/167	16/167	15/167
CTC-dim vs. All Other	0/75	0/75	4/75	9/75
CTC-dim vs. Other Marginal	0/55	0/55	2/55	4/55
CTC-dim vs. CTC-Small	0/76	0/76	5/76	17/76
CTC-Small vs. All Other	1/48	0/48	1/48	8/48

2.5 Qualitative Results

Given the high accuracy of each of our decision functions, and the insights afforded to us by determining reoccurring features, we want to visualize the reoccurring features and observe if these visualizations could be interpreted by pathologists in a meaningful way. For each classification task there were 25 trials run. In each trial, the data was randomly-partitioned for training and testing using the standard 75/25 partitioning. Within each trial, a decision function was built using the LIBLINEAR l_1 regularized, l_2 loss function linear support vector machine. In each trial, a subset of the features are selected as important for the given classification task. After running the trials we can identify the features that are selected in 95% of the trials, the number of which can be seen in Table 2.1. The design of the FRD allows us to invert the descriptor and visualize individual features or observe a single cell reconstructed using a specific set of features. We have reconstructed specific cells using the set of features selected in 95% of trials for various classification tasks. In this way we can visualize our results and further understand the differentiating structure for an indicated classification task.

We define two different cells of each class. The cell with the prefix “closest" refers to the cell of a given class which had the FRD closest to the average FRD across that class. Similarly, we define

Table 2.4: This table shows the break down of the location of the reoccurring features in a given classification task within the DAPI channel as a fraction of the total number of reoccurring features. Inner refers to occurring within the first four rings. Inner-central refers to occurring within the fifth through eighth rings. Outer-central refers to occurring within the ninth to twelfth rings. Lastly, outer refers to occurring within the thirteenth to sixteenth rings.

Classification Task	Inner	Inner-Central	Outer-Central	Outer
CTC-Candidate vs. All other	1/122	18/122	7/122	7/122
CTC-Candidate vs. CTC-Ap	1/63	14/63	3/63	4/63
CTC-Candidate vs. CTC-dim	0/84	8/84	6/84	15/84
CTC-Candidate vs. CTC-Small	0/63	17/63	3/63	4/63
CTC-Ap vs. All Other	0/44	8/44	1/44	5/44
CTC-Ap vs. CTC-dim	1/110	3/110	7/110	12/110
CTC-Ap vs. CTC-Small	0/167	17/167	8/167	35/167
CTC-dim vs. All Other	3/75	11/75	18/75	11/75
CTC-dim vs. Other Marginal	2/55	14/55	14/55	7/55
CTC-dim vs. CTC-Small	0/76	13/76	13/76	1/76
CTC-Small vs. All Other	6/48	3/48	9/48	3/48

the prefix “farthest" to mean the cell of class who’s FRD was the farthest from the average FRD across that class. We determine the closest and farthest from average FRD using the Euclidean distance measure. The following figures show the reconstructions of the closest and farthest from average cells of each class involved in the indicated classification task. In Figures 2.5.1, 2.5.2, 2.5.3, 2.5.4, 2.5.5, 2.5.6, 2.5.7, 2.5.8, 2.5.9, 2.5.10, and 2.5.11 we see reconstructions of cells using features from different classification tasks. For each of these figures, the image directly below a labeled cell image is the reconstruction of the labeled cell for the given feature set. Additionally, all images have been shown on using a color axis bounded between 0 and 255. Thus, though some of the images may appear dim the intensity of the colors in the reconstruction are not artifacts but rather capture important differences.

First, contained in Figure 2.5.1 we see reconstructions of representatives of each cell type using the reoccurring features from the CTC-Candidate versus all other cells of interest classification task. We see, in this figure, that CTC-Candidates are distinguished from the other cells of interest based on the uniformity and thickness of cytokeratin in the central-cell together with greater amounts of inner-central DAPI present. Within this figure we can also observe the similarities between the reconstructions for each of the farthest-from-average cells. The similarities between the

farthest-from-average cells for these cell types illustrate the high-level of variation in the current labeling of data. Furthermore, we observe the lack of features/structure from the inner cell. This lack of inner structure suggests that at the given resolution, the central cell/center of the nucleus contains limited structurally differentiating information. Biologically, this is significant since research has shown that there is differentiating information contained in the nuclei and that at the current resolution this information is not being be captured.

Next, Figure 2.5.2 captures discriminating information between CTC-Candidates and CTC-Aps. Based on the reconstructions shown, we can infer that CTC-Candidates can be separated from CTC-Aps based on a more uniform and intense nucleus, more uniform cytokeratin extending into the outer-center region of the cell, and a less circular shape. The first inference is based on the bright inner cell DAPI. Our second inference is based on the purple hue of the outer regions of the closest-to-average CTC-Candidate, as purple suggests the presence of both DAPI and cytokeratin. Additionally, the purple in the closest-to-average CTC-Candidate appears uniform in its intensity suggesting uniformity in the two contributing channels.

According to Figure 2.5.3 the combination of outer-central cytokeratin expression and outer DAPI expression contribute to the differential structure between CTC-Candidates and CTC-dims. Visually, the category of CTC-dim is more uniform than many of the others, but still contains size variation in the nucleus of the cells. Thus, since CTC-dims typically have larger nuclei than CTC-Candidates, the presence of outer cell DAPI features for discrimination is in agreement with visual classification criterion.

Figure 2.5.4 illustrates the differentiating structure between CTC-Candidates and CTC-Smalls. Of note in this figure are the inner-central cytokeratin features and outer DAPI features. CTC-Smalls are biologically classified based on their high cytokeratin expression, but small nucleus and overall cell size. Thus, the presence of DAPI in the outer cell highlights the size variation between these two cell populations, as does the location of the cytokeratin which would not extend beyond the inner-central rings in a cell of small size. We also note, in this figure, the presence of outer

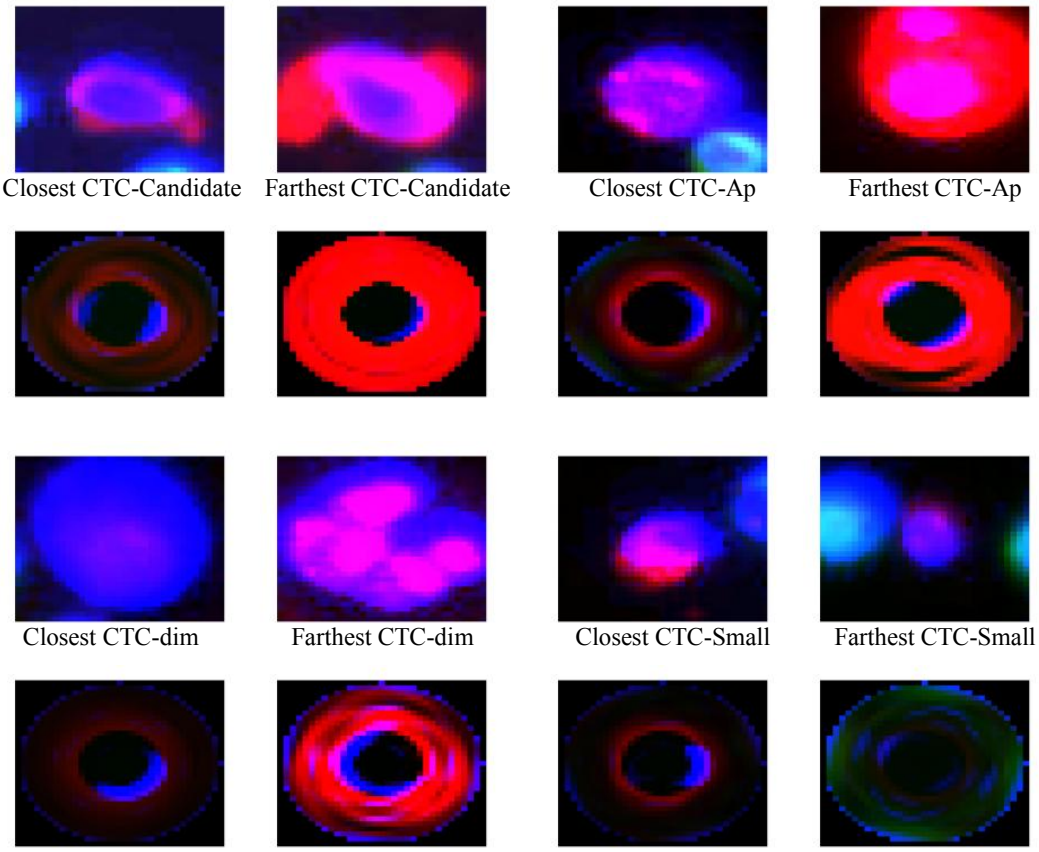


Figure 2.5.1: This figure provides a visualization of the differential structure used to separate CTC-Candidates from all other cells of interest.

DAPI features in the CTC-Small shown can be attributed to the closeness of the the neighboring cells in both the closest-to-average and farthest-from-average cells.

The visualizations shown in Figure 2.5.5 make the low performance of the CTC-Ap versus all other cells of interest classification task very understandable. From Table 2.1 we see that the CTC-Ap one-versus-all classification task has the lowest accuracy of all one-versus-all tasks. Visually, there appears to be much less structurally differentiating information contained in this classification task. Both the closest-to-average and farthest-from-average cells of each cell type are very similar to those of the CTC-Aps. As mentioned in the qualitative results section, CTC-Aps are cells undergoing apoptosis. This figure suggests that the labeling of CTC-Aps, based on visual inspection, likely has captured cells of the other cell populations which may be undergoing apop-

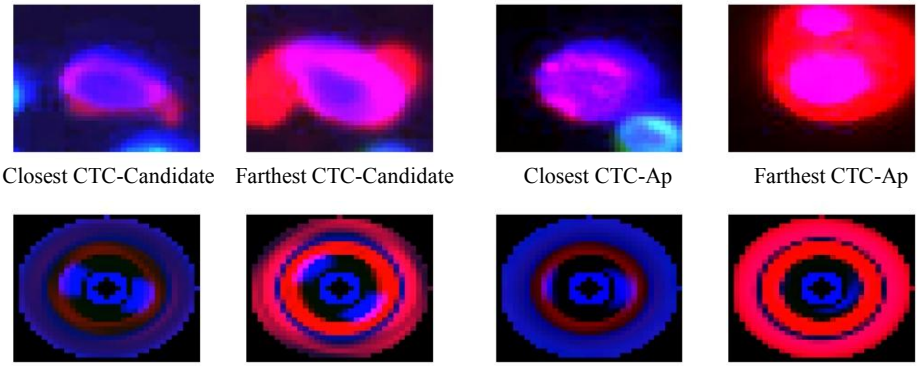


Figure 2.5.2: This figure provides a visualization of the differential structure used to separate CTC-Candidates from CTC-Aps.

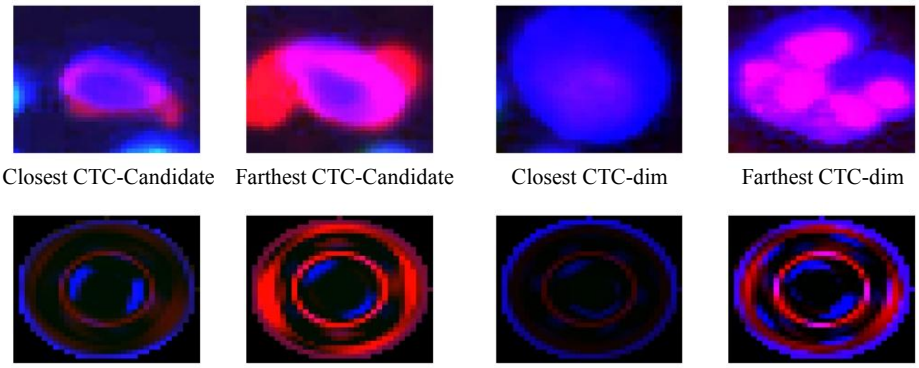


Figure 2.5.3: This figure provides a visualization of the differential structure used to separate CTC-Candidates from CTC-dims.

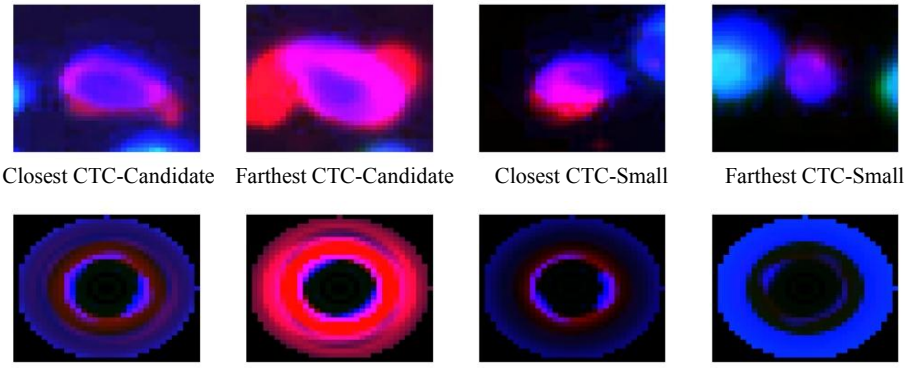


Figure 2.5.4: This figure provides a visualization of the differential structure used to separate CTC-Candidates from CTC-Smalls.

tosis. Additionally, we note that this classification task had the fewest number of reoccurring features suggesting that the classifier is highly dependent on the set of cells used for training. A consequence of the small number of reoccurring features is a less complete reconstruction.

Despite the lack of differential structure for separating CTC-Aps from all other cells of interest, there is differential structure in the remaining CTC-Ap pairwise classifications. For example, in Figure 2.5.6 we see that inner cytokeratin and non-uniform outer DAPI are involved in separating CTC-Aps from CTC-dims. In fact, this classification task is the only task with apparent cytokeratin in the inner cell. Next, in Figure 2.5.7 we again see the differential structure returning to the outer regions of the cell to differentiate between CTC-Aps and CTC-Smalls. Although there is a weaker level of separation between CTC-Aps and CTC-Smalls, according to Table 2.1, by observing the reconstructions, there are clear differences between the cell populations involved. These differences are especially pronounced between the farthest-from-average cells, although the farthest-from-average CTC-Small closely resembles the closest-to-average CTC-Ap. In many of the classification tasks shown, we see greater similarity between the farthest-from-average cells of different types. Biologically, the patters of similarity in the CTC-Ap versus CTC-Small classification task could suggest a greater overlap of these two populations.

Moving to Figure 2.5.8 we see our first case of a heavily-weighted inner DAPI, and the DAPI channel being used almost exclusively, for separating cell populations. Figure 2.5.8 shows reconstructions based on the classification task of separating CTC-dims from all other cell populations. Observing the reconstructions we see DAPI features, the other channel features are overwhelmed, and note the size, circularity, and uniformity of the DAPI in the CTC-dim reconstructions. The other populations of interest, however, have more asymmetry and smaller areas of intense DAPI expression. Again, this mirrors the characterization used in visual classification in which a CTC-dim is characterized largely by the size of the nucleus and limited to nonexistent cytokeratin.

The remaining classification tasks for distinguishing CTC-dims from other populations are visualized in Figures 2.5.9 and 2.5.10. In both the classification task separating CTC-dims from the remaining marginal populations and separating CTC-dims from CTC-Smalls, we again see the

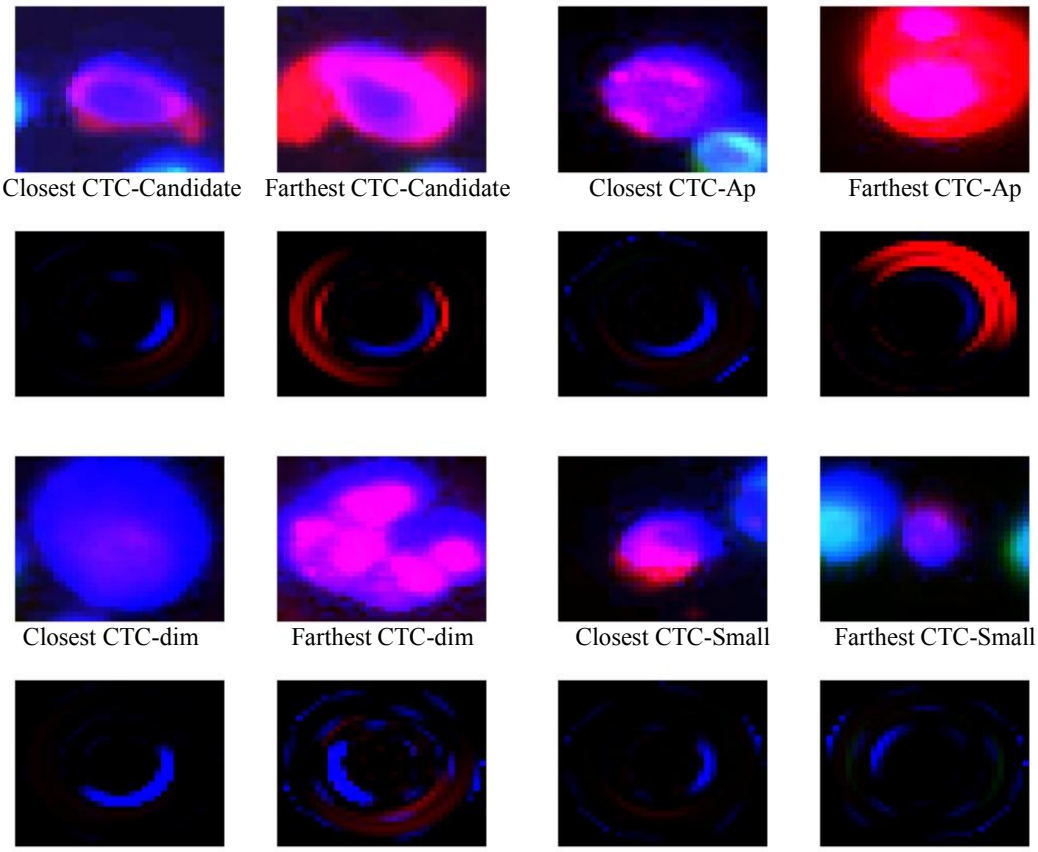


Figure 2.5.5: This figure provides a visualization of the differential structure used to separate CTC-Aps from all other cells of interest.

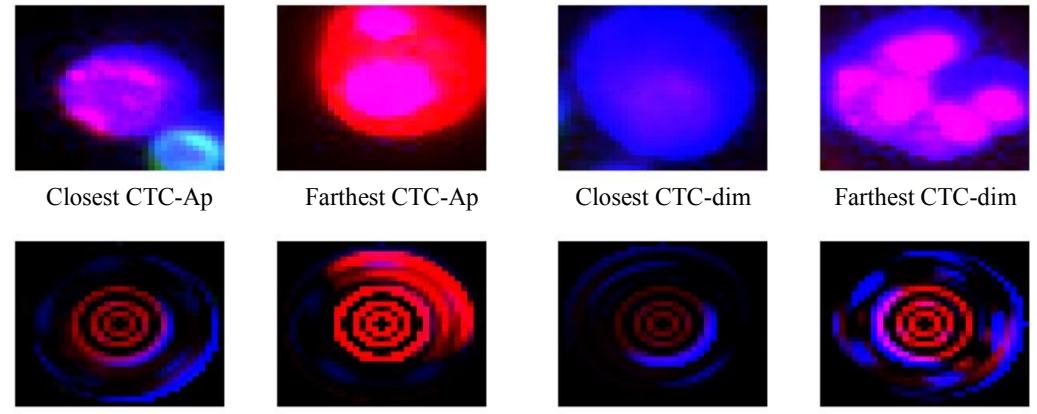


Figure 2.5.6: This figure provides a visualization of the differential structure used to separate CTC-Aps from CTC-dims.

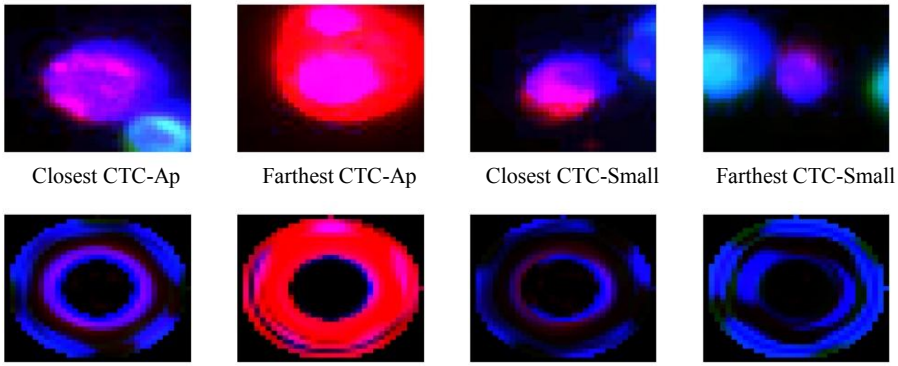


Figure 2.5.7: This figure provides a visualization of the differential structure used to separate CTC-Aps from CTC-Smalls.

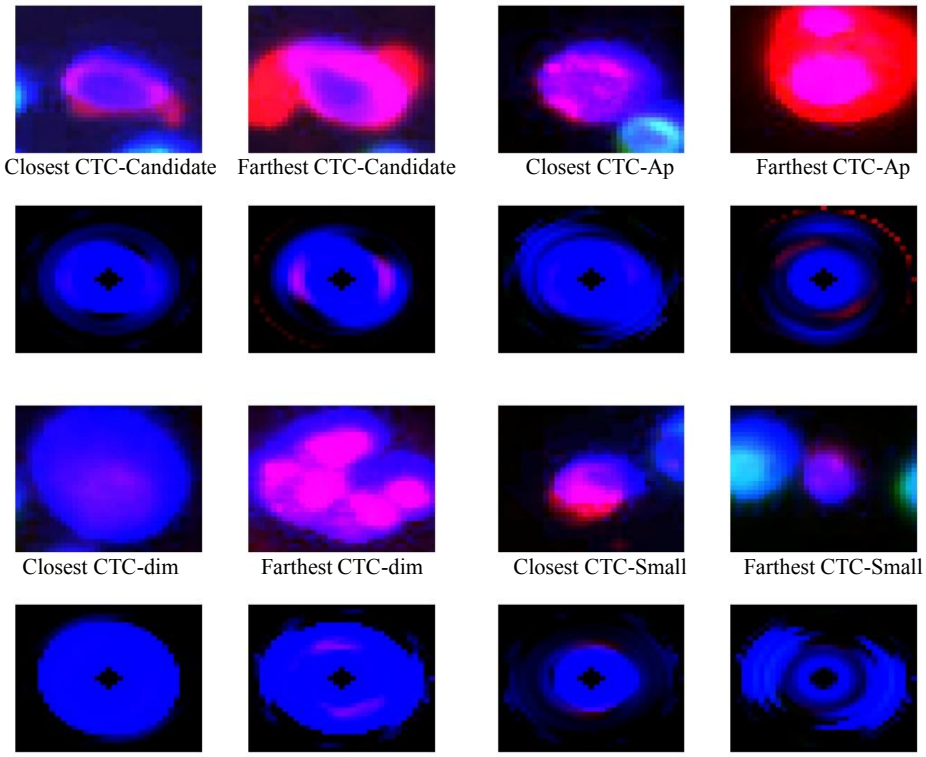


Figure 2.5.8: This figure provides a visualization of the differential structure used to separate CTC-dims from all other cells of interest.

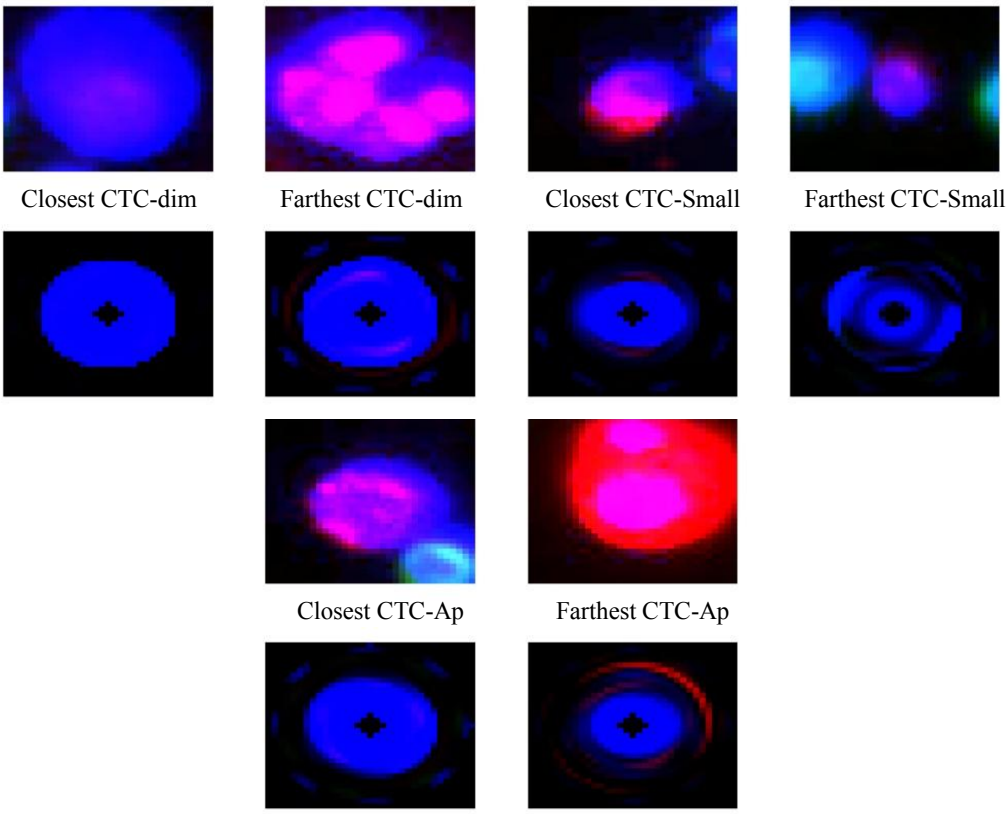


Figure 2.5.9: This figure provides a visualization of the differential structure used to separate CTC-dims from other marginal cell populations.

majority of the features coming from the DAPI channel. Also, we again see that the reconstructions of CTC-dims show a larger region of uniform DAPI intensity which are more uniform and circular than the reconstructions of the other cells.

Finally, we consider Figure 2.5.11 which contains the reconstructions of all cells using the reoccurring features selected to separate CTC-Smalls from all other cells of interest. This classification task also uses the second fewest number of reoccurring features for reconstruction, suggesting, like CTC-Ap versus all other cells of interest, that the classifier is highly dependent on the training set. We do, however, see that the inner and inner-central DAPI and cytokeratin play a role in differentiating between the cells. Thus, we are again capturing the size information of the CTC-Smalls.

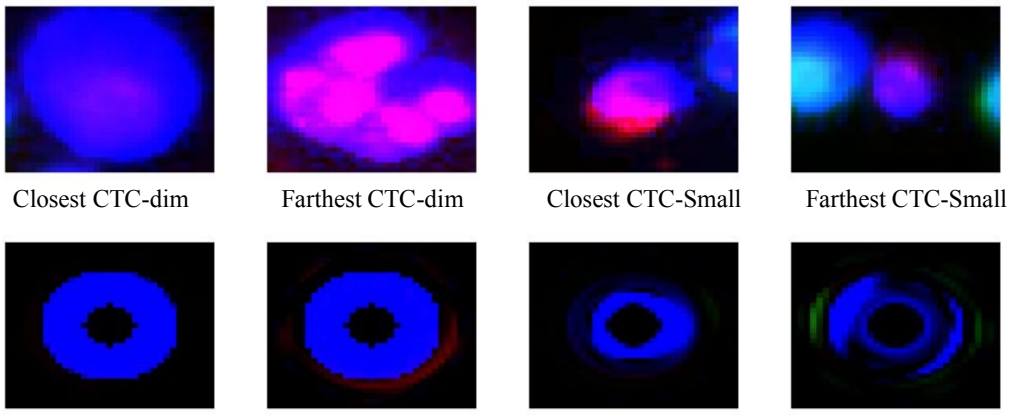


Figure 2.5.10: This figure provides a visualization of the differential structure used to separate CTC-dims from all CTC-Small.

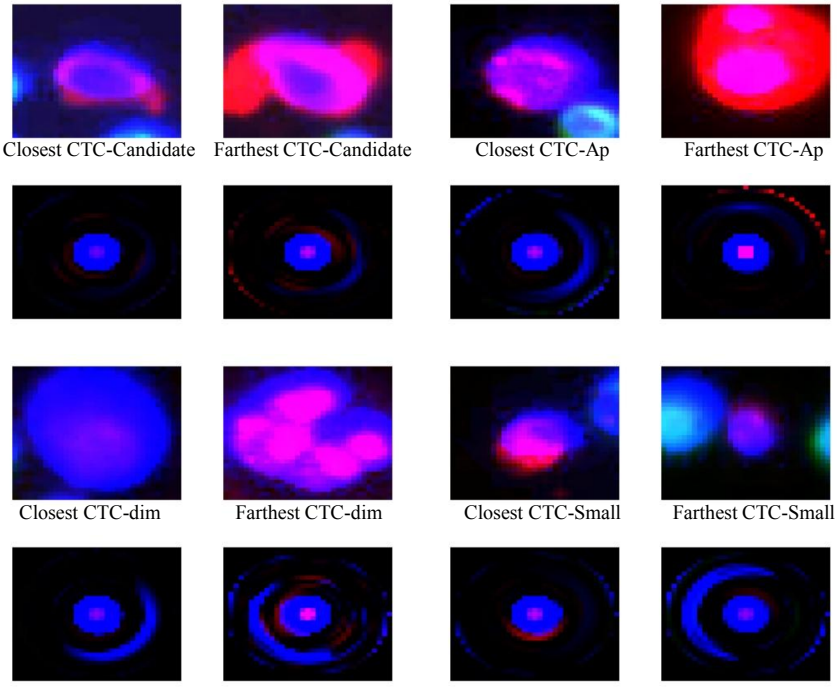


Figure 2.5.11: This figure provides a visualization of the differential structure used to separate CTC-Smalls from all other cells of interest.

The culmination of these visualizations allow us to describe the features that are consistently evaluated at each branch of our decision tree. First, we classify according to the features shown in Figure 2.5.1. Next, we classify based on the features shown in Figure 2.5.9. Finally, we separate based on the features shown in Figure 2.5.7. This amounts to first looking at the combination of central cell cytokertain and inner cell DAPI, then looking for at the uniformity, size, and circularity of the DAPI features from the inner cell to the outer-central cell, and then finally looking at the combination of outer-central to outer DAPI features and outer-central cytokeratin.

The variation across reconstructions of a single cell show that there is strongly differentiating structure within a cell that changes depending on the classification the reconstruction is based on. By observing the various reconstructions we can highlight tasks where there is more or less differentiating structure. Again, this knowledge allows us to understand the classifications and define hard lines between what we look for in a cell in order to belong to a certain cell population.

While many of our features can be connected to features that are currently used in visual classification, there are several benefits to computer automated classification. First, in addition to being objective an automated classification method, when properly optimized, could reduce valuable labor hours needed to evaluate a single patient sample. Second, due to the construction of the FRD, we are evaluating information extracted from a pixel level which cannot be accomplished by the human eye. Thus, by using an automated method, we have the ability to use all available information in a cell to classify all events. Finally, and most importantly, by using a numerical approach to classification we are able to express the unbiased confidence of our classification of a single event. Using a support vector machine classifier, we determine the class of a cell in a training set based on the sign of the decision function value. For data which is linearly separable, the decision values of one class will all be greater than or equal to one, while all of the decision values of the other class will be less than or equal to one. Thus, for data points in the training set, the larger the magnitude of the decision function value, the greater our confidence in the classification of the event. For example, in Figure 2.5.12 we see a plot of the decision function values for all the points in a training set for one trial of the CTC-Candidate versus all other cells of interest classifier. In this example

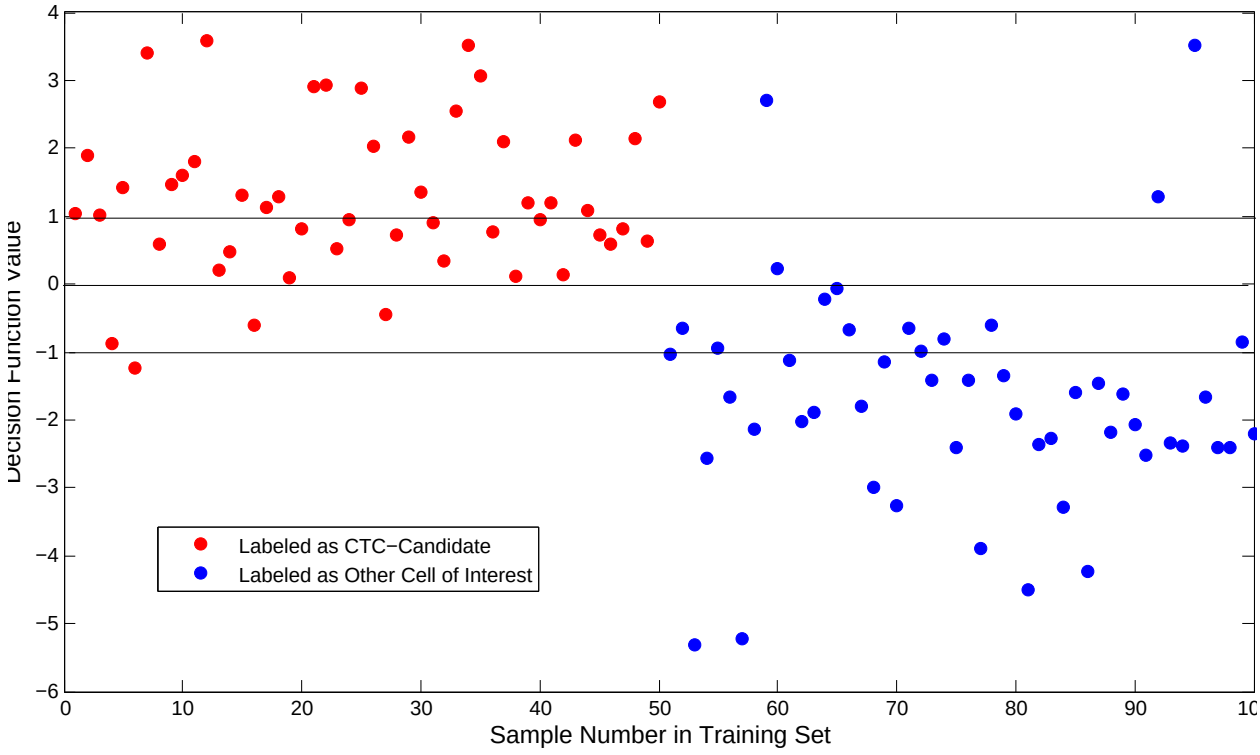


Figure 2.5.12: This figure shows a plot of the decision function values for the samples in the training set for one trial of the CTC-Candidate versus all other cells of interest classification task.

we consider the points that have been visually classified as CTC-Candidates to be the positive (red) class, while the cells which have been visually classified as other cells of interest are the negative (blue) class. Therefore, any point with a decision value greater than zero has been classified by our classifier as a CTC-Candidate, and any with a decision value less than zero has been classified as another type of cell of interest. For any point with a decision value magnitude greater than one, we have confidence in the classification (more confidence the larger the magnitude). However, points with decision values between one and negative one, we are less confident in the classification. In this way we can quantify the confidence of a cell's classification, which cannot easily be routinely accomplished using visual classification. The results here show that perhaps some cells visually classified into certain cell populations may actually fit better within a different cell population.

To conclude, from the images provided, one can readily observe a significant amount of variation within a single cell class. This is important to be able to quantify for several reasons. Of primary concern is ongoing research connecting quantities of circulating tumor cells to cancer status. If these correlations between quantity of circulating tumor cells and cancer status are based on highly subjective classifications, then there will be doubt as to the validity of those claims. However, if using these visualization techniques and classification methods allow us to not only confidently classify, but also capture all cell types of interest, then we will have confidence in any connections that we may find in the future. While the variation we observe is over a small dataset, it allows us to understand the challenges in generating a set of classifiers that can be readily applied to many patient samples across various cancer types. In supervised learning methods, it is very important to select the right set of data to train your classifiers. From these results, we see that we will need to systematically scale our training set to produce a more universal set of classifiers both within a cancer type and across all cancers.

2.6 Discussion

The current methods for CTC isolation and characterization are both representative of traditional pathology practice of visual evaluation but also lend themselves to sophisticated and rigor-

ous mathematical frameworks as the data sets are all digital by nature. Human inspection, while of high quality, will always be qualitative and limited to certain scales of data sets leading to a self-evident risk of false negatives. The FRD method, when applied to large datasets, could provide more confidence around having more completely interrogated a patient's sample. There are significant advantages in using computational methods in that patterns can be identified more rapidly. For example, no single parameter currently exists that would distinguish CTCs originating from different organs but a large scale computational approach might indeed be able to do so. However, the more detailed and comprehensive method described herein may lend further insight into distinguishing cancer types and stages of disease.

The Fourier-Ring Descriptor (FRD) methods developed in this paper are the first that we know of that exploit the rotational structure of cells identified in peripheral blood samples from patients with metastatic breast, prostate and lung cancer cells. We have used these FRDs, along with a linear support vector machine decision tree classifier to exploit the size variations and morphological distinctions among the cell populations to obtain reasonable and quantifiable accuracy benchmarks. While we are not proposing that this automated technique be used in place of visual inspection in making clinical decisions, we do see our methods as offering clinically relevant quantifiable support in their decision processes. Hence, we view this as a first step in developing a useful computer-vision based clinical support tool for the active monitoring of cancer patients based on peripheral blood samples.

Chapter 3
Nonnegative Matrix Factorization Using the Split Bregman Algorithm and an Application in Frequency Agile Lidar

During the first half of 2015 I was a visiting researcher at MIT Lincoln Laboratory under Dr. Dimitris Manolakis of the Sensor Technology Group. Unlike the other projects presented so far, this project was motivated by a specific application: *Frequency Agile Lidar* (FAL). The FAL system was designed as a laser radar (lidar) system that could be used to identify and track aerosolized bio-agents.

In 2012 [34] presented an approach to analyze the data obtained by the FAL system that was successful in identification and tracking of aerosolized agents. An evaluation of the approach proposed in [34] (which relies heavily on a preprocessing algorithm presented in [35]) was requested by the laboratory so that it could objectively determine if continued investment in the system was advisable.

A background introduction to FAL as well as to the Split Bregman algorithm (the primary tool used in producing the information on which identification and tracking are performed) are presented in Sections 3.1 and 3.2, respectively. Next, in Section 3.3 the nonnegative matrix factorization problem is described. Nonnegative matrix factorization arises within the context of FAL and is solvable using the Split Bregman Algorithm. The issue of internal dimension identification within nonnegative matrix factorization, connecting this application to the mathematical theme of this doctoral research, is explained in Section 3.5.2. In Section 3.4 the Split Bregman algorithm is applied to FAL to produce results similar to those in [34]. This chapter concludes with a discussion section, Section 3.5, that presents two novel insights. Section 3.5.1 is a discussion of the results of the application of Split Bregman to FAL which contains the results of a modification of the method proposed in [34] which produces superior results. We return to the issue of dimension in Section 3.5.2 with the proposal of another modification of an objective function that might enable estimation of the internal factoring dimension.

3.1 Frequency Agile Lidar

Bio-toxins, viruses, and bacteria can be used as dangerous weapons of terror in an aerosolized form. Aerosolized molecules of size $1 - 5\mu m$ are typically used because of how readily they are absorbed into the lungs. As a consequence, agencies would like to be able to rapidly detect, locate, and identify (discriminate various agents) biological agents to provide warning of threats. One method for reaching these objectives is to use elastic backscatter lidar. From a lidar system one can compute the concentration of an aerosol at a given range and the backscatter cross section of the aerosol at a particular wavelength. Concentration can be used to track an aerosol cloud while the backscatter cross section, at multiple wavelengths, can be used too identify the aerosol. We describe one such standoff lidar system designed to meet the goals of detection, location, and identification of aerosols: the frequency agile lidar (FAL).

Figure 3.1.1 provides a schematic of the FAL system. A FAL system is based on a finely tunable CO_2 laser that can be rapidly tuned to 60, the number of distinct molecular states, different frequencies. The range of wavelengths to which the laser can be tuned classifies the system as a long wave infrared lidar. FAL is a form of elastic backscatter lidar meaning that the photons measured at the receiving detector are photons of the same wavelength of the transmitted laser scattered by the aerosol molecules at an angle of 180 degrees (back along the line of transmission). In many ways FAL can be thought of as a blind differential absorption lidar. Essentially, the system is designed based on the assumption that each aerosol has a Rayleigh backscatter cross-section signature. The optical devices used in the FAL system are there to reduce beam divergence of the transmitted photons as well as to ensure that all transmitted energy is enclosed in the receiver field of view.

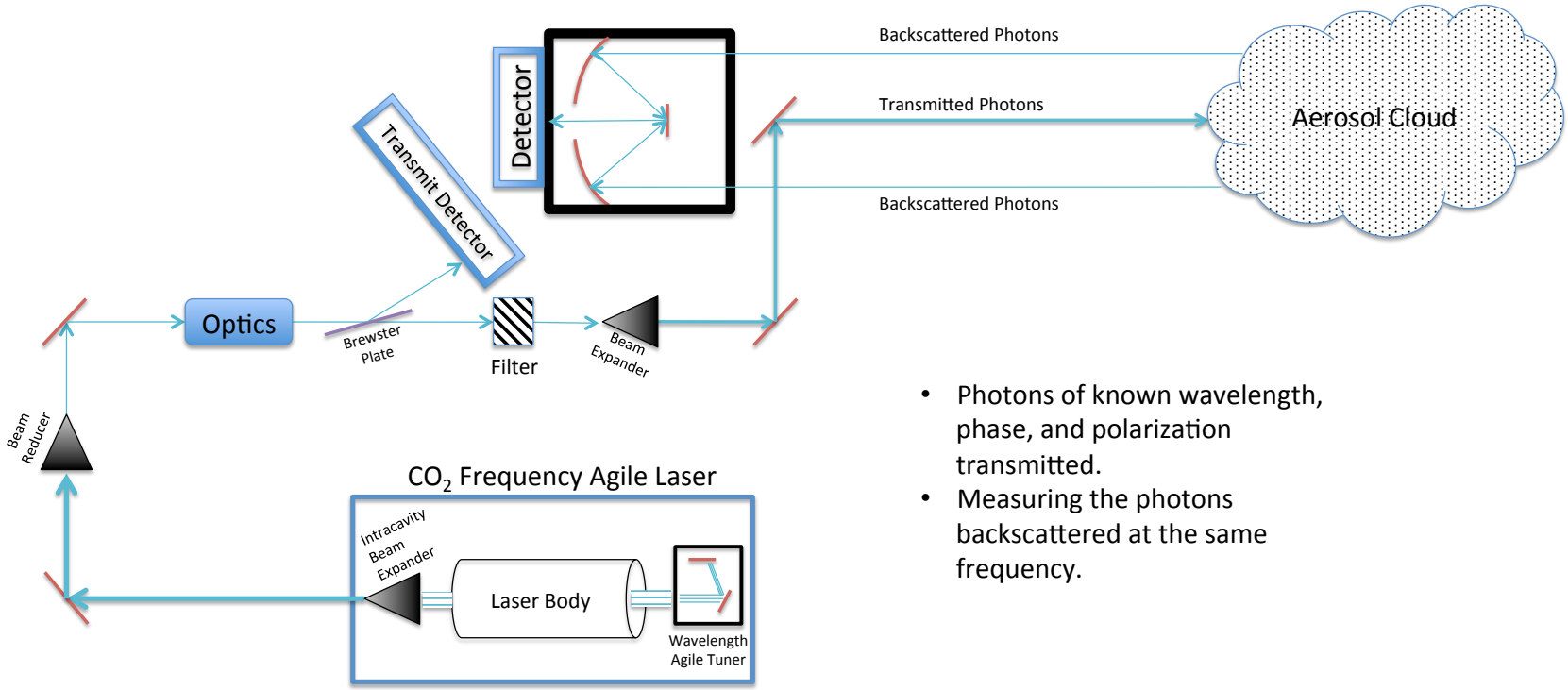


Figure 3.1.1: A schematic of the FAL system.

In all data evaluated, 19 different wavelengths are leveraged (although 60 are available). A *burst* is a fixed pattern of these 19 individual *pulses*. At each burst index one cycle of the pattern is transmitted (a portion of the transmitted waveform is deflected for measurement) and the backscatter from each pulse is measured. Some important specifications of the system are summarized in the following list:

- Pulsed CO_2 laser,
- Automated tuning to 60 discrete wavelengths between 9.2μm and 10.7μm,
- Uses a pseudo-resonant galvanometer with a flat copper mirror that is aligned with a fixed grating for the automatic tuning of wavelength (selection based on angular position of mirror w.r.t. fixed grating),
- Max number of pulses per burst is 20 (multi-pulse bursts),
- Wavelength/pulse switching time $\tau_\lambda = 5$ms
- τ_B depends on number of pulses in burst
- Output pulses are 120ns full-width half-max spikes followed by nitrogen tails.

An example of two different bursts, from real data, are shown in Figure 3.1.2. Figure 3.1.2 aptly shows that the maximum transmitted energy of each wavelength is not uniform across pulses, as well as variations in the nitrogen tails of each pulse. For this reason, as well the presence of electromagnetic interference and other artifacts, steps of preprocessing of the data must occur on a pulse-by-pulse basis.

Moreover, in Figure 3.1.3 is an example of a transmitted and received waveform for a single pulse within a burst corresponding to aerosol being present. The jumping of the received waveform, highlighted in Figure 3.1.4, at the far left is electromagnetic interference due to the lack of calibration in the system. Additionally, the first little bump in the received waveform occurring at the same range index of the transmitted waveform peak, highlighted in Figure 3.1.5, is a consequence of transmitted photons backscattering off of optical equipment into the detector. This is the

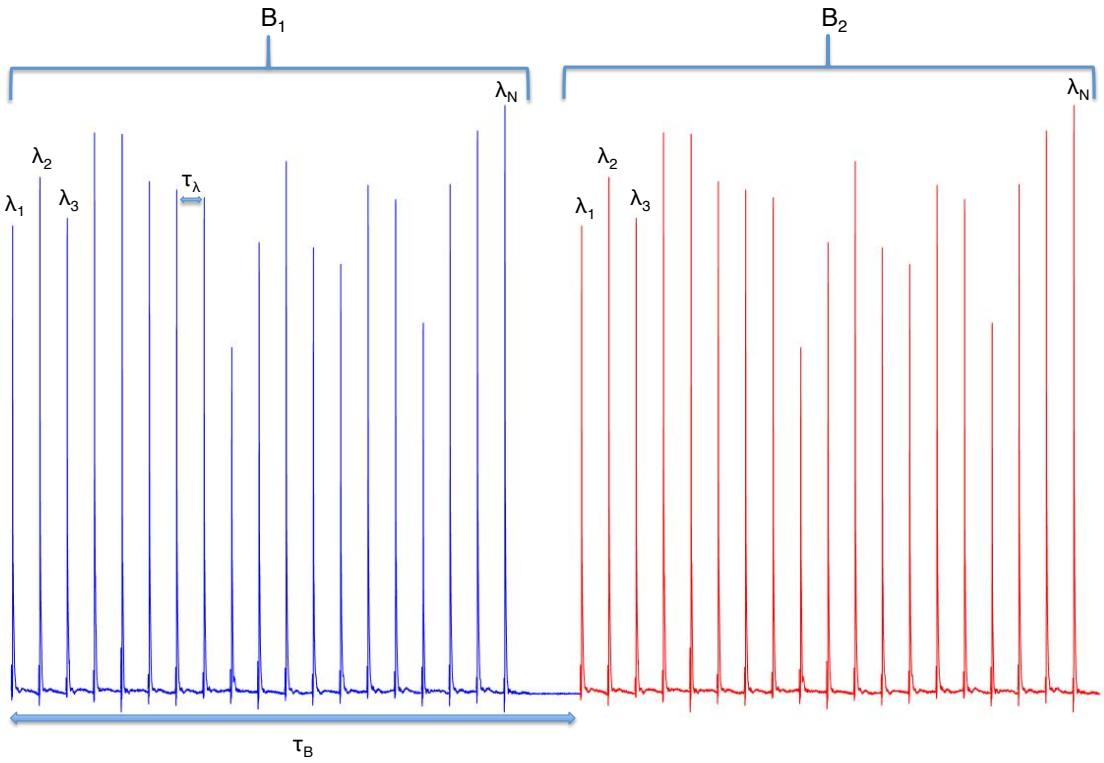


Figure 3.1.2: Example of two transmitted bursts.

result of an inability to isolate the received and transmitted waveform. The bump near the right end of the received waveform, highlighted in Figure 3.1.6, corresponds to the backscatter from the aerosol cloud. Notice the similarity in shape between the transmitted waveform and this bump; the heavy tail. The heavy tail, highlighted in Figure 3.1.7, is referred to as the heavy nitrogen tail of a CO_2 laser. The extended duration of the transmitted pulse results in a need to deconvolve the transmitted waveform shape from the received waveform. Finally, the received waveform does not fall exactly to zero over the range measured which is due to characteristics of the lidar system parameters and setup.

3.1 The Lidar Equation

Mie scattering is the scattering of photons off of particles of a size larger than the wavelength of the photons being scattered. Rayleigh scattering is for particles of a smaller size than the wave-

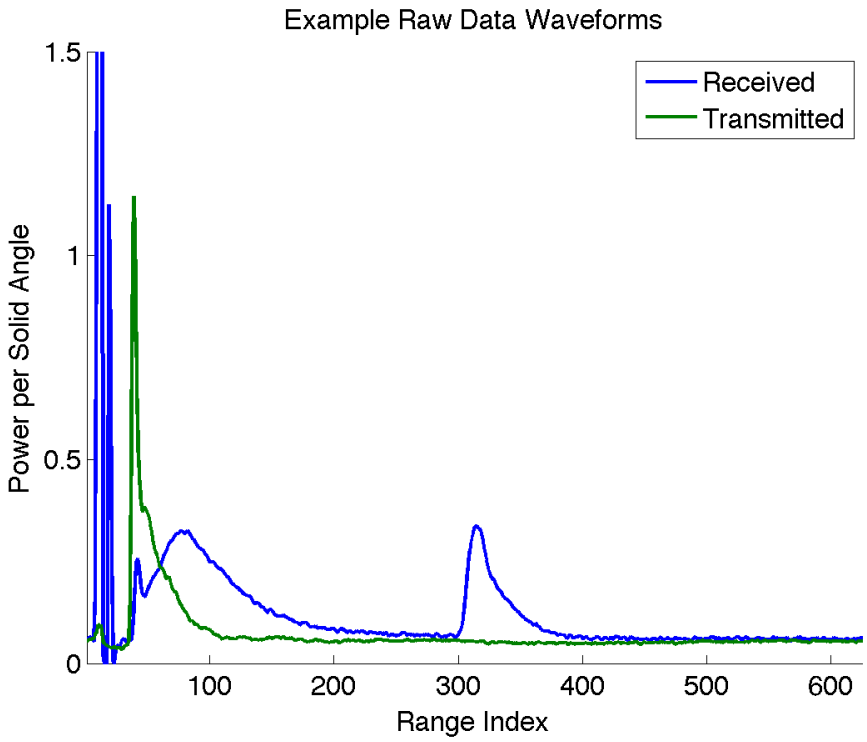


Figure 3.1.3: Example of raw data for a single pulse in a burst index corresponding to presence of aerosol.

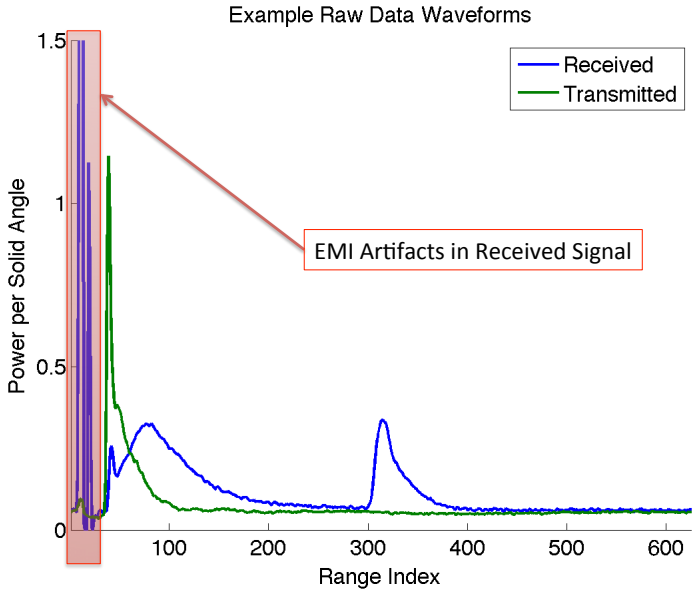


Figure 3.1.4: Example of raw data for a single pulse in a burst index in with a positive presence of aerosol.

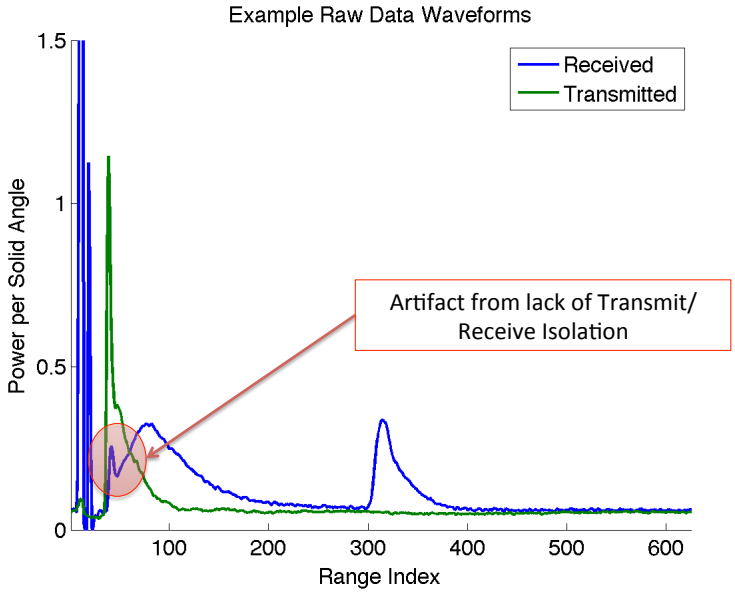


Figure 3.1.5: The shaded region highlights the artifact that appears as a consequence of inability to isolate the transmitted and received data.

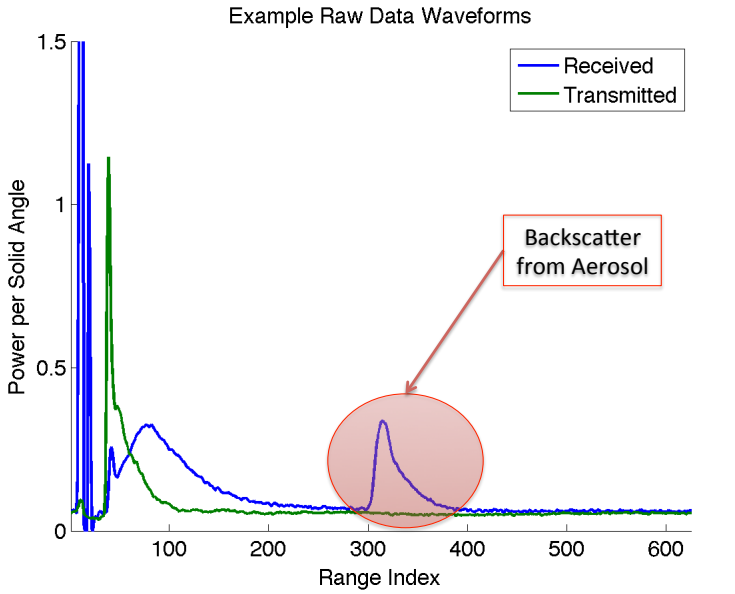


Figure 3.1.6: The shaded region highlights the portion of the received, raw waveform that comes from backscatter off the aerosol cloud.

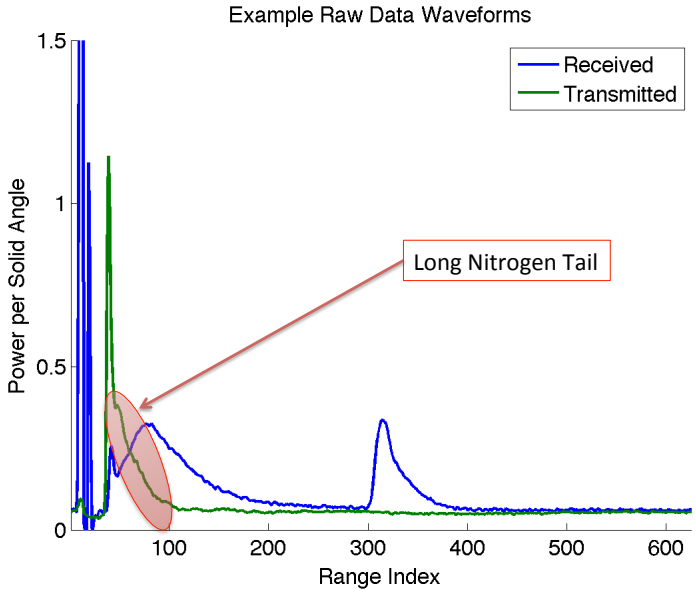


Figure 3.1.7: The shaded region highlights the long nitrogen tail of the CO_2 laser which is the primary cause for deconvolution in preprocessing.

length being scattered. Since aerosolized molecules are often $1-5\mu m$ in size, and the wavelengths used in FAL span $9.2-10.7\mu m$, we are looking at Reighly scattering. The lidar equation relates characteristic parameters of a molecule to the measured power per steradian. We write the lidar equation as

$$P(R,\lambda) = (\frac{P_0}{R^2})(\frac{\tau Ac}{2})(T(\lambda))(\beta(\lambda))\exp\left[-2\int_0^R \alpha(r)dr\right] \tag{3.1.1}$$

where

- R and λ are distance and wavelength, respectively
- P_0 is the laser power, so the P_0/R^2 term accounts for intensity falling off over distance.
- τ is the duration of the pulse, A is the area of the receiver, c is the speed of light, and $\tau Ac/2$ is the volume of the area receiving backscatter from.
- $T(\lambda)$ is the optical efficiency of the Lidar at a given wavelength

- $\beta(R)$ is the backscatter coefficient and $\beta(R) = \beta_{atm}(R) + \beta_{aer}(R)$.
- $\alpha(R)$ is the extinction coefficient and $\alpha(R) = \alpha_{atm}(R) + \alpha_{are}(R)$.

The following are key assumptions made by Warren, *et al*:

1. Natural atmosphere and aerosol only sources of backscatter,
2. Optically thin plume ($\alpha_{aer} = 0$),
3. Natural atmosphere transmission must be known,
4. Know the time-steps/ burst indices which are aerosol free ($i = 1, ..., T_b$),
5. The backscatter coefficient of aerosol (β_{aer}) can be written as the product of the efficiency of backscatter/ backscatter cross section and the concentration/ number density of the aerosol.

Under these assumptions we can break the lidar equation into two pieces: one coming from the atmosphere, and another coming from the aerosol. This yields the following modified version of the lidar equation

$$P(R,\lambda) = \left[(\frac{P_0}{R^2})(\frac{\tau A c}{2})(T(\lambda))(\beta_{atm}(R,\lambda)) \exp\left[-2 \int_0^R \alpha_{atm}(r,\lambda) dr \right] \right] \tag{3.1.2}$$

$$+ \left[(\frac{P_0}{R^2})(\frac{\tau A c}{2})(T(\lambda))(\beta_{aer}(R,\lambda)) \exp\left[-2 \int_0^R \alpha_{atm}(r,\lambda) dr \right] \right]. \tag{3.1.3}$$

The term shown in Equation 3.1.2 comes from the atmosphere while the term shown as Equation 3.1.3 is from the aerosol. The forms of the lidar equation we have presented so far use the standard notation from within the physics community and we note that it is dissimilar to the notation used by Warren, *et al*. This form of the lidar equation is used at each of the different wavelengths.

3.1 Pre-Processing of the FAL data

In order to utilize the lidar equation to determine the backscatter cross section for a given wavelength of a given aerosol and the concentration of the aerosol, we first need remove artifacts in the

data by preprocessing. We describe and implement the preprocessing as proposed in [35] and refer the reader to that paper for additional detail. Additionally, preprocessing examples are generated primarily through the use of preprocessing code provided by the authors of [35]. Preprocessing consists of 3 (or 4) steps. Those steps are as follows

1. Create zero baseline using the tail of a pulse,

2. Remove electromagnetic interference artifacts in data (optional),

3. Subtraction of the mean of the natural atmosphere (obtained by averaging the received waveforms over all aerosol-free time-steps), *This amounts to estimating the term in Equation 3.1.2.*

4. Deconvolve the transmitted waveform from the received waveform using a Wiener-Helstrom Filter.

Each of these preprocessing steps will be described separately in in the remainder of this section.

Zero Baselining

Zero baselining is done on a pulse-by-pulse basis for both the transmitted and received waveforms, separately. Baselining is done to make averaging meaningful. The zero baseline is obtained by averaging the values at the tail end of each pulse and uniformly subtracting this value from all of the measured values for that pulse. An example of the result of baselining can be seen in Figure 3.1.8.

Remove Electromagnetic Interference Artifacts

This is an optional step that is not implemented by Warren, *et al*. This step consists of identifying where the transmit signal is sent out and removing data from before this point. This effectively removes much of the variance of the data based on these artifacts. In the next step, background mean subtracting, if these artifacts are not removed or lessened, there will be greater variance in the data being fed into deconvolution. The authors of [35] avoid this by truncating the data based on user specified distances of interest . One such way to lessen the effect of the interference artifacts

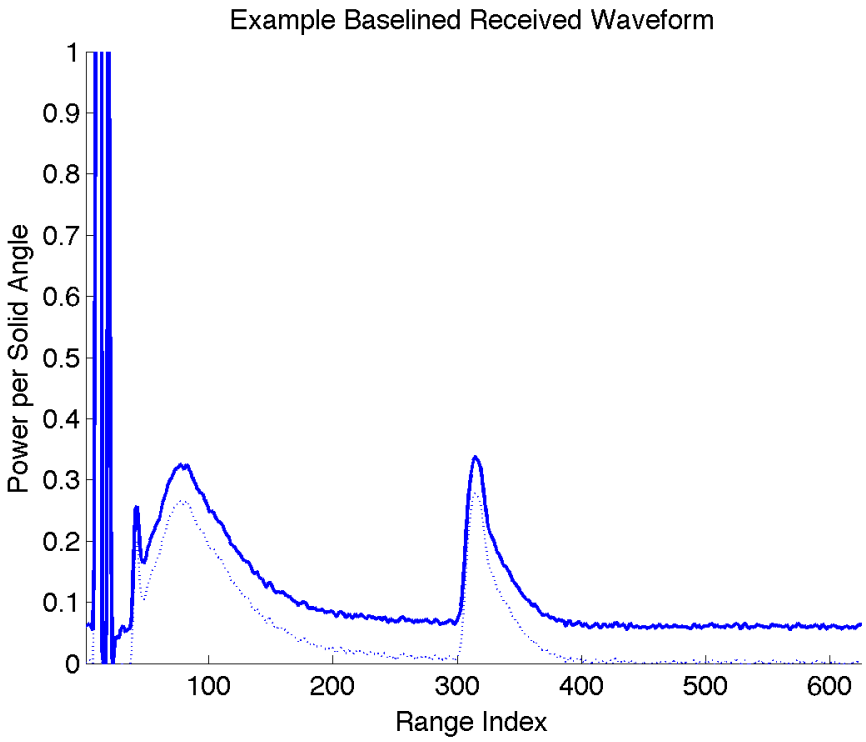


Figure 3.1.8: Result of zero baselining on a received waveform for a single pulse in a burst containing aerosol.

is to base truncation on the derivative of the transmitted waveform. For example, on a pulse-by-pulse basis we can look at the average pairwise change in the transmitted waveform and determine the starting index based on the first jump greater than three standard deviations above the mean jump. Since the delay to triggering the laser is different depending on the pulse we can zero pad to equal length vectors across pulses. Zero padding is made an appropriate option based on the zero-baselining.

We note that when implementing author provided code, a minimum distance of interest equal to 0km versus 0.5km produces a drastic difference in the final processing algorithm outputs. The thickness of the band corresponding to the response from the aerosol shown in Figure 3.1.10 is much narrower than that of Figure 3.1.12. This corresponds to a greater location uncertainty of the cloud when you are interested in the near field of view. Furthermore, we see a larger range of values occurring in the data before deconvolution when we are interested in the near field. This is

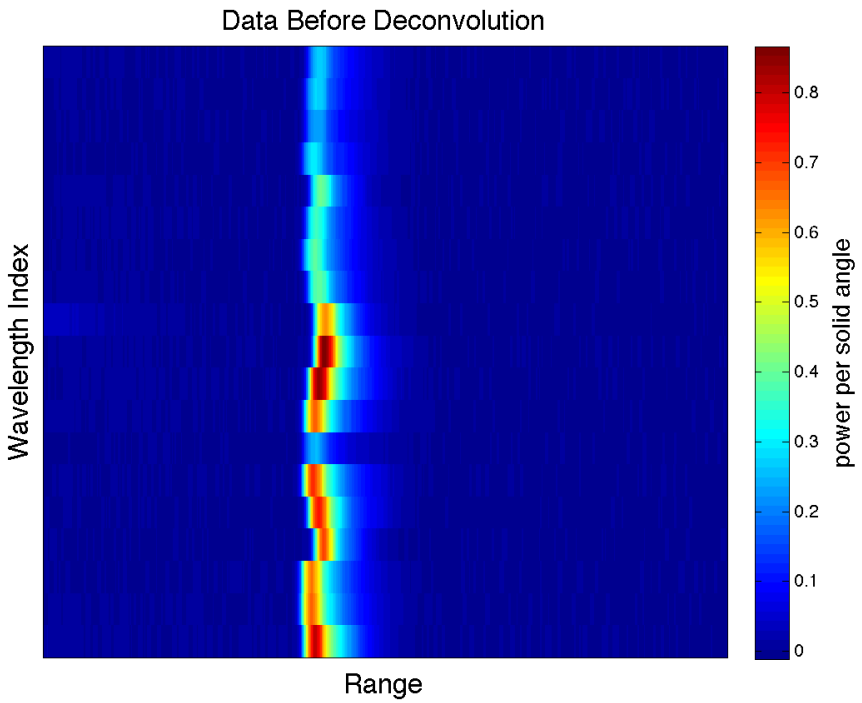


Figure 3.1.9: An example of all data for a single burst before deconvolution with a minimum range of interest of 0.5km.

a consequence of the averaging of the EMI artifacts. Examples of the differences in data before deconvolution can be seen in Figures 3.1.9 and 3.1.11.

Mean Natural Atmosphere Subtraction

The preprocessing occurring at this step requires knowledge of the burst indices corresponding to no aerosol present and will use the zero baselined, and artifact free if user has chosen to remove these artifacts, data for a single wavelength/pulse over all background burst indices. For a single wavelength, the zero baselined received waveforms over all background burst indices will be averaged to estimate the term in the lidar equation coming from the atmosphere, Equation 3.1.2. This average will then be subtracted off of that pulse from all bursts. An example of the result of mean natural atmosphere subtraction can be seen in Figure 3.1.13.

As you can see in Figure 3.1.13, after the subtraction we now have an estimate for the lidar response specifically from the aerosol that has been convolved with the transmitted waveform for that pulse. We feel it important to note that mean natural atmosphere subtraction is a centering of

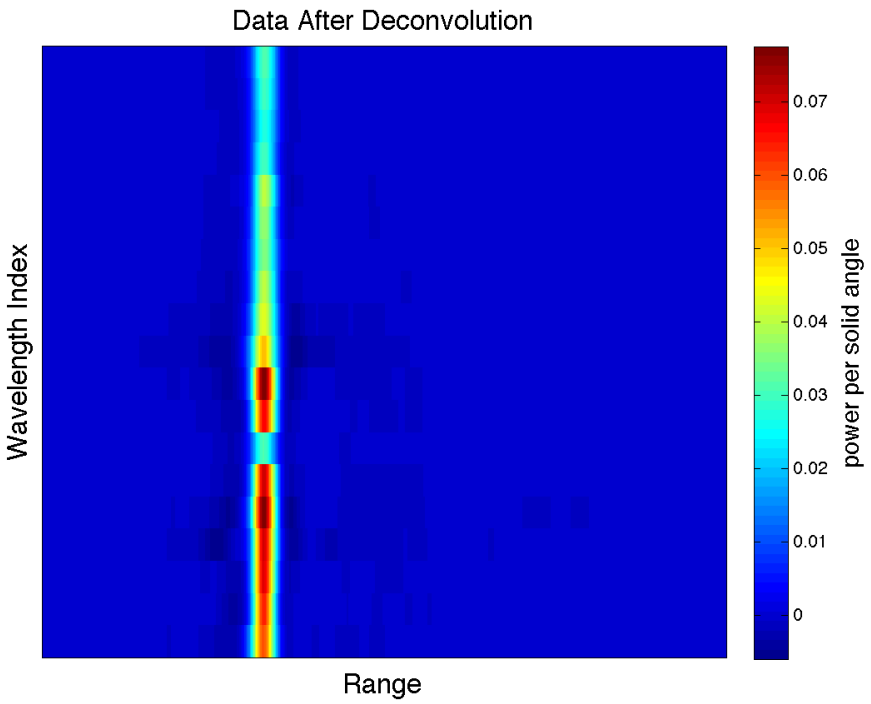


Figure 3.1.10: An example of a complete data matrix for a single burst after the completion of pulse-by-pulse deconvolution with a minimum range of interest of 0.5km.

the data with respect to the clutter mean, but it does not completely remove it as the atmosphere is not static.

Wiener Filter Deconvolution

A Wiener filter is used to deconvolve the transmitted waveform for a pulse from the zero baselined, mean natural atmosphere subtracted data. Weiner filter deconvolution requires knowledge of the power spectral density (PSD) of both the noise and the signal. The noise PSD can be estimated from pulses in background bursts. However, the signal PSD is estimated based on an assumption of Gaussianity and ad hoc parameter selection. The rationale for this seems to be unclear and/or not in alignment with the associated phenomenology. Furthermore, some of the techniques implemented in the provided code are nonstandard for the proposed method of deconvolution. An example of a deconvolved waveform is shown in Figure 3.1.14.

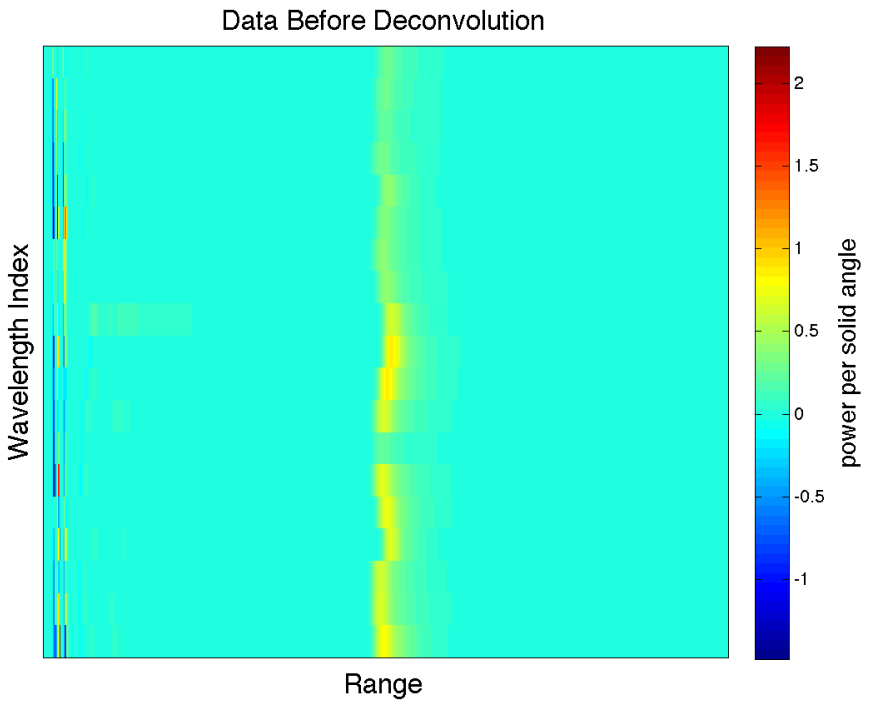


Figure 3.1.11: An example of all data for a single burst before deconvolution with a minimum range of interest of 0.0km.

3.1 Data After Preprocessing

After preprocessing of the data we have something equivalent to a noisy version of the lidar equation aerosol term, Equation 3.1.3. Deconvolution of the data is used to register the data. The effectiveness of this is highlighted in Figure 3.1.9 and Figure 3.1.10 which show a data matrix before and after deconvolution, respectively. Data after deconvolution, shown in Figure 3.1.10 is the starting input data for the processing pipelines. Furthermore, we can see the form of the data after deconvolution for a single pulse within a burst in Figure 3.1.14.

Based on the assumptions of the model we have that the lidar response, at a given wavelength, to the aerosol will be proportional to the backscatter coefficient of the aerosol at that wavelength. Both processing approaches assume that this constant of proportionality is equal to 1. Under this assumption being placed on the constant of proportionality the mathematical model for the preprocessed data becomes

$$\boldsymbol{G} = \boldsymbol{\rho C} + \boldsymbol{\nu},$$

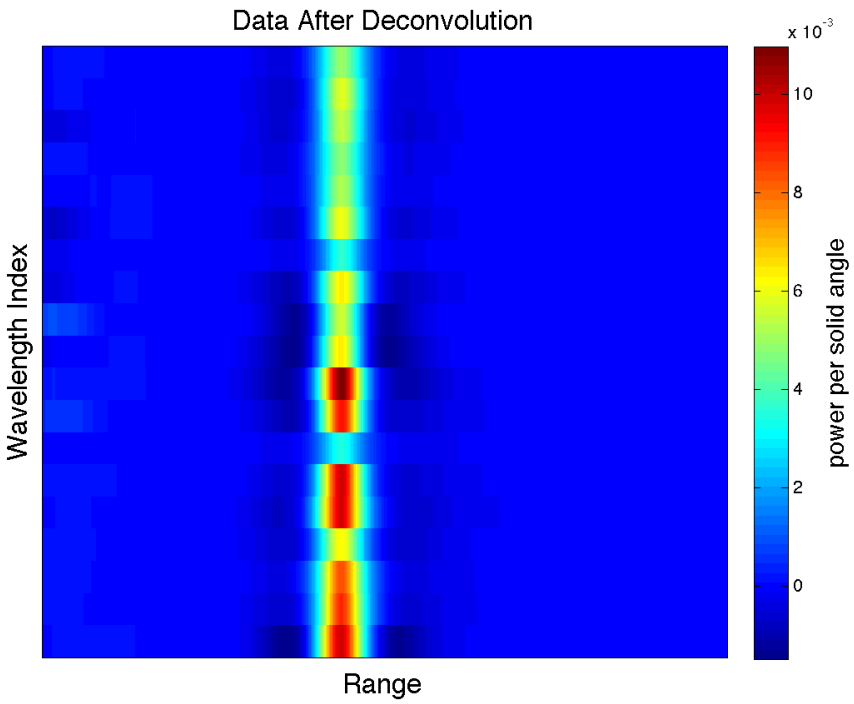


Figure 3.1.12: An example of a complete data matrix for a single burst after the completion of pulse-by-pulse deconvolution with a minimum range of interest of 0.0km.

where

- ρ is the matrix of aerosol feature vectors as columns,
- C is the matrix of aerosol concentrations as a function of range as rows,
- and ν is a noise matrix.

Figure 3.1.15 shows a more detailed schematic of the mathematical model for the preprocessed data. A consequence of the constant of proportionality assumption is that all units of measurement become arbitrary and one cannot establish a clear mapping between the scales of the values that would be obtained in a properly calibrated laboratory environment.

At a high level, the entire process of from system to end goals can be represented as in Figure 3.1.16. We will be discussing the final box in the system: processing of the data.

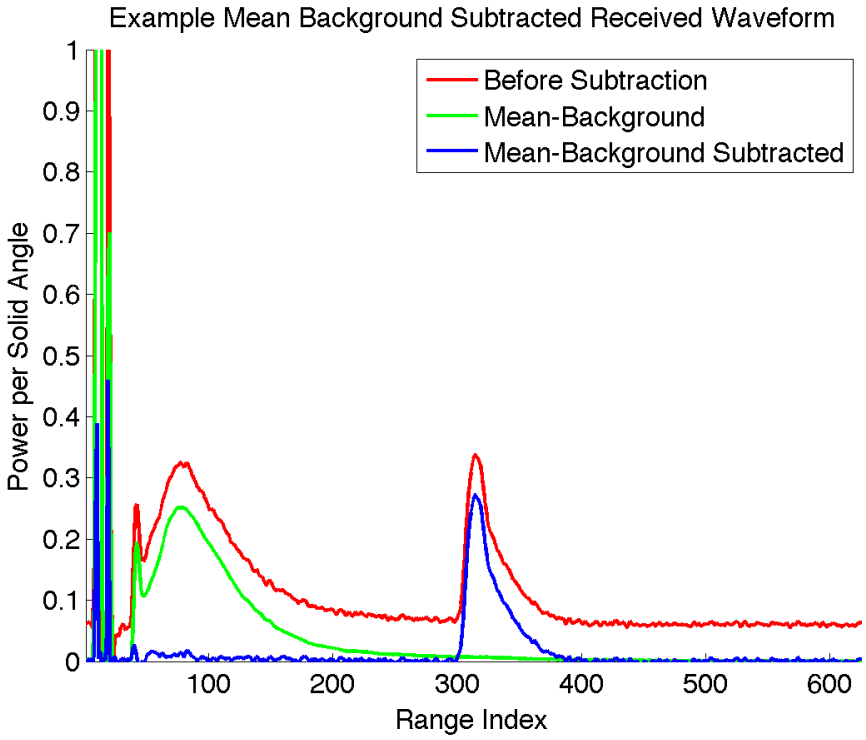


Figure 3.1.13: Example of an artifact free, zero baselined, mean natural atmosphere subtracted received pulse for a burst with aerosol present.

3.2 Split Bregman Algorithm

The Split Bregman algorithm [36] is an ingenuitive combination of classical techniques from convex optimization. As such, it is also a method for convex optimization. The approach is based on splitting variables by introduction of auxiliary variables as well as alternating primal and dual variable updates. The method has been shown to be very efficient under a variety of conditions.

In Figure 3.2.1 we can see how the Split Bregman algorithm fits into the greater landscape of convex optimization. In the network shown, red nodes correspond to classical methods while blue nodes correspond to more modern approaches. The color of the line connecting two nodes is associated to a piece of literature existing establishing a connection between those nodes. This network aims to highlight the rich and complex connections between many of the approaches to solving convex optimization problems. Figure 3.2.2 shows only the connections of the Split Bregman node from the larger network.

Split Bregman shows its greatest utility for problems of the following form:

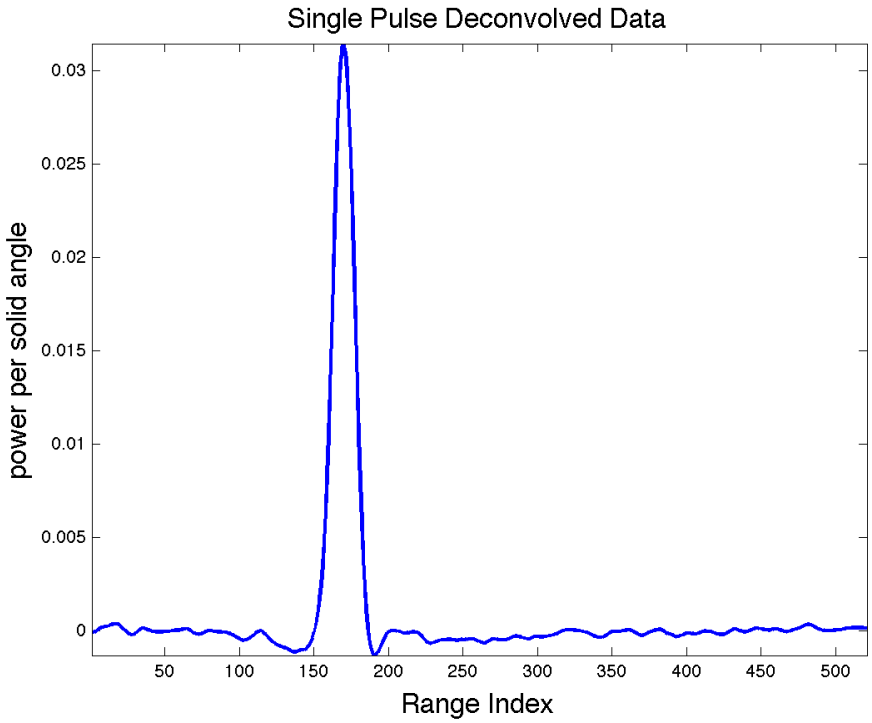


Figure 3.1.14: A single deconvolved pulse from within a burst using a starting distance of interest equal to 0.5km. Produced using Warren's code.

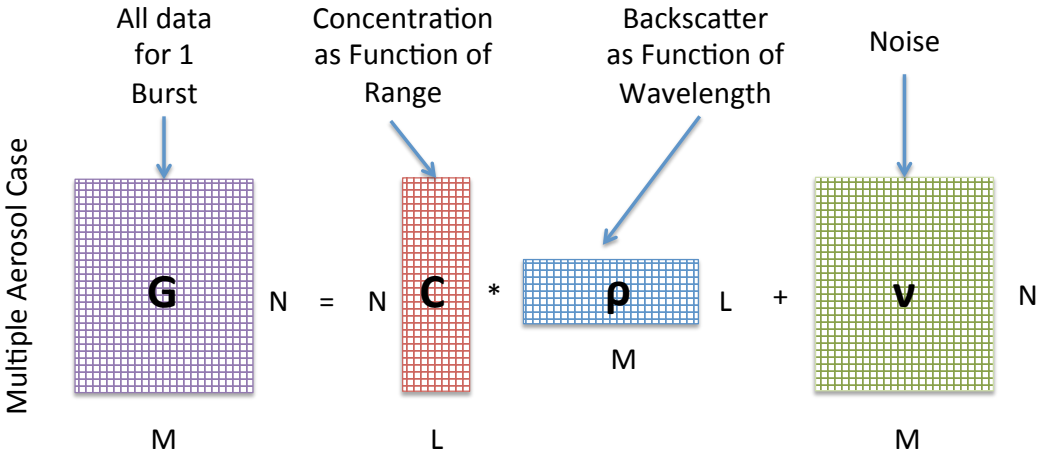


Figure 3.1.15: A representation of the mathematical model. We define M as the number of wavelengths, N as the number of range cells, and L as the number of aerosols.

- p^t_{ij} : Transmitted voltage/photons as a function of time at a fixed burst and wavelength.
- p^r_{ij} : Received voltage/backscattered photons as a function of time/range at a fixed burst and wavelength.
- G : Preprocessed data estimate of the lidar response from the aerosols as a function of range, for all wavelengths, at a single burst.
- C : Matrix of concentration vectors (as a function of range) of all aerosols for a fixed burst. *Used for tracking aerosol*.
- ρ : Matrix of backscatter coefficient vectors (as a function of wavelength) of all aerosols for a fixed burst. *Feature vectors used for aerosol discrimination and detection are taken from this matrix*.

Figure 3.1.16

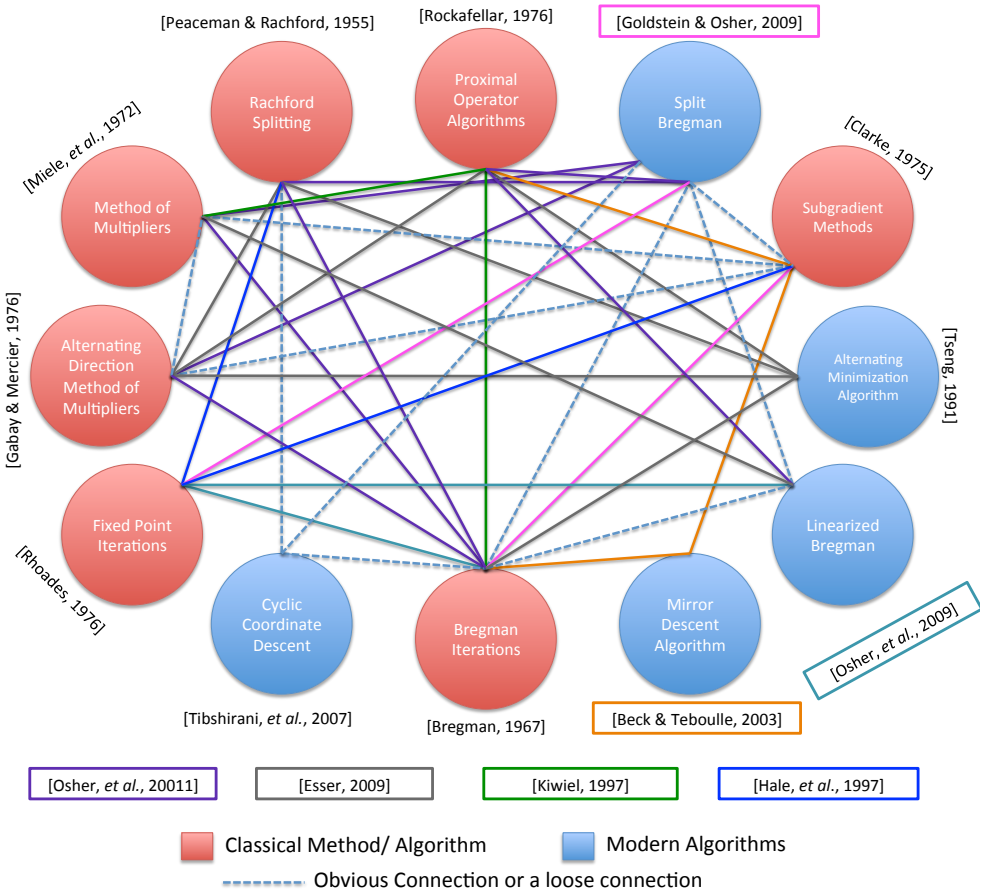


Figure 3.2.1: A network illustrating how Split Bregman fits into the larger landscape of convex optimization

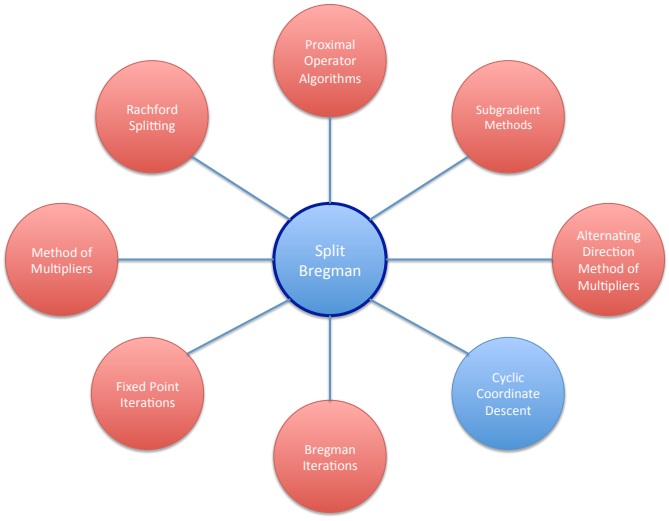


Figure 3.2.2: A Split Bregman centric network of convex optimization methods.

$$\underset{u}{\text{minimize}} \quad f(\phi(u)) + g(u),$$

where

1. $\phi(u)$ convex and differentiable,
2. f convex and non-differentiable, and
3. g convex and differentiable.

In many of the well known applications of Split Bregman, f is a 1-norm or the BV-norm, though the framework has been successfully adapted to other formulations.

The Split Bregman algorithm begins with the unconstrained problem

$$\underset{u}{\text{minimize}} \quad f(\phi(u)) + g(u).$$

From this unconstrained problem a new constrained problem is created through introduction of a new variable as

$$\underset{u,y}{\text{minimize}} \quad f(y) + g(u) \quad \text{subject to} \quad \phi(u) = y.$$

Once we have a constrained optimization problem we can create the augmented Lagrangian

$$L(u, y, b) = f(y) + g(u) + \frac{\lambda}{2}||\phi(u) - y - b||_2^2.$$

Next we alternate between primal and dual variable updates according to

$$\begin{aligned} y^{n+1} &= \underset{y}{\text{argmin}} \quad L(u^n, y, b^n), \\ u^{n+1} &= \underset{u}{\text{argmin}} \quad L(u, y^{n+1}, b^n), \\ b^{n+1} &= b^n + y^{n+1} - \phi(u^{n+1}). \end{aligned}$$

As previously stated, many of the well known applications of Split Bregman involve a 1-norm or BV-norm. Our discussion will focus on the former. Two well known problems that arise in signal processing that can be solved by Split Bregman are basis pursuit (also known as LASSO) and sparse nonnegative matrix factorization. The basis pursuit problem can be stated as

$$\underset{s}{\text{minimize}} \quad \frac{1}{2}||Ks - f||_2^2 + \mu||s||_1,$$

where K is a known matrix, f is a vector of data, and s is the signal you wish to recover. Alternatively, sparse nonnegative matrix factorization is formulated as

$$\underset{A,B}{\text{minimize}} \quad \frac{1}{2}||AB - F||_F^2 + \mu_A|A| + \mu_B|B|,$$

where A, B are unknown sparse matrices, F is a matrix of data, and we wish to recover both A and B.

The shift to using 1-norm regularization in such problems, as opposed to the classical 2-norm regularization, is generally motivated by three things:

1. forces sparsity,
2. variable selection, and
3. allows easier interpretation.

In order to understand how these desirable traits are obtained, we look to Figure 3.2.3. Figure 3.2.3 shows the set up of a 2-norm regularized regression problem on the left and a 1-norm regularized regression problem on the right. The shaded regions correspond to the feasible region based on the norm of the vector. A 2-norm feasible region is smooth due while the 1-norm feasible region has four discontinuities. The ellipses are level sets of the objective function and the optimal solution (the red dot) occurs at the first point where a level set is tangent to the feasible region. It is clear to see that there is a higher likelihood, overall, for the level sets of a variety of objective functions will hit at a sparse solution (in this case having one zero) for the 1-norm feasible region.

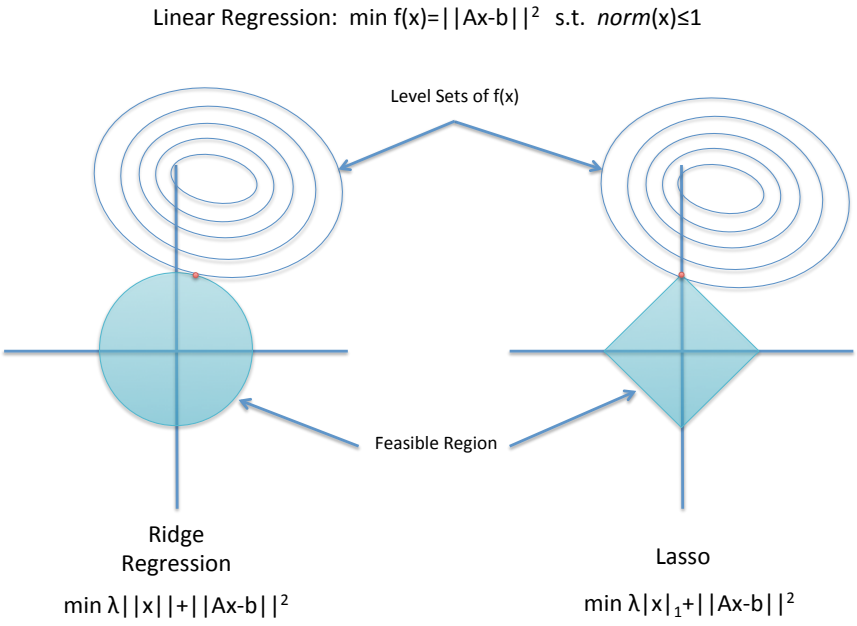


Figure 3.2.3: Comparison of 2-norm and 1-norm regression.

Additionally, one can see why often a 2-norm regularized problem has small but non-zero variable values.

It is important to note that despite the desirable attributes of 1-norm regularization, there are some short comings to consider. Throughout the literature it has been shown that 1-norm regularization performs more poorly on data sets that contain groups of correlated variables [37, 38]. In the case of variable correlation, 1-norm regularized problems will generally select a single feature from the group while 2-norm regularization will give equal importance to each feature in that group. This particular characteristic will be mentioned again in the context of a real world application in a later section. We conclude this section by presenting a toy example solved by Split Bregman.

3.2 Toy Example: Basis Pursuit

Basis pursuit has the end goal of recovering a one dimensional signal when the matrix linear transformation matrix associated to the linear system is known. We will begin with a true signal

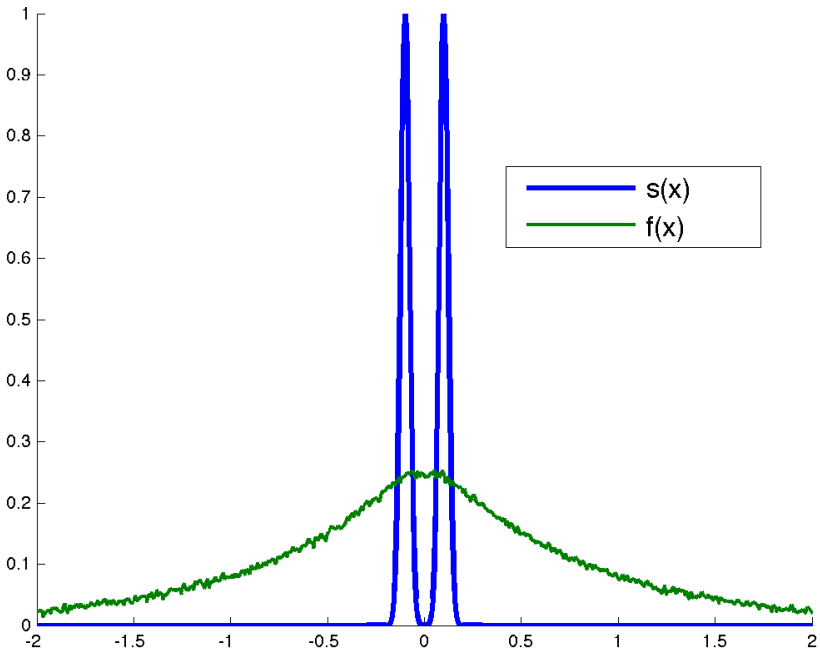


Figure 3.2.4: The true and distorted signals for the basis pursuit toy problem.

$s(x)$ that is the sum of two Gaussians:

$$s(x) = \exp[-\frac{(x+0.1)^2}{0.001}] + \exp[-\frac{(x-0.1)^2}{0.001}].$$

Next we will simulate a linear system by blurring the true signal with a Toeplitz matrix given by

$$K_{nm} = r^{|n-m|}, \ r = 0.99.$$

Finally, we complete the simulation of a linear system by adding random white noise, with standard deviation chosen to be 0.004, to obtain our final observed/distorted signal $f(x)$. Our goal will be to recover $s(x)$ from $f(x)$. Figure 3.2.4 shows a plot of both $s(x)$ and $f(x)$.

The basis pursuit optimization problem is written as

$$\underset{s}{\text{minimize}} \quad \frac{1}{2}||Ks - f||_2^2 + \mu||s||_1.$$

Split Bregman aims to solve the equivalent constrained optimization problem

$$\begin{aligned} \underset{s,d}{\text{minimize}} \quad & \frac{1}{2}||Ks - f||_2^2 + \mu||d||_1 \\ \text{subject to} \quad & s = d, \end{aligned}$$

which has a scaled dual form augmented Lagrangian defined to be

$$L(s, d, b) = \frac{1}{2}||Ks - f||_2^2 + \mu||d||_1 + \frac{\lambda}{2}||s - d + b||_2^2.$$

Following the steps previously stated for Split Bregman, we complete the updates of each variable according to

$$\begin{aligned} s^{k+1} &= \underset{s}{\text{argmin}} \quad \frac{1}{2}||Ks - f||_2^2 + \frac{\lambda}{2}||s - d^k + b^k||_2^2, \\ d^{k+1} &= \underset{d}{\text{argmin}} \quad \mu||d||_1 + \frac{\lambda}{2}||s^{k+1} - d + b^k||_2^2, \\ b^{k+1} &= b^k + s^{K+1} - d^{k+1}. \end{aligned}$$

The update for the variable s is based on an unconstrained minimization of a differentiable function. Consequently, we can differentiate and set equal to zero to find the closed form solution for the update to be

$$s^{k+1} = (K^t K + \lambda I)^{-1}(K^T f + \lambda(d^k - b^k)).$$

Next, we notice that the update for the variable d can be efficiently solved using the shrinkage operator. As a result we determine the update of d to be given by

$$d^{k+1} = \text{shrinkage}(s^{k+1} + b^k, \mu/\lambda).$$

Finally, we can update the dual variable b using a standard gradient ascent method to produce updates of the form

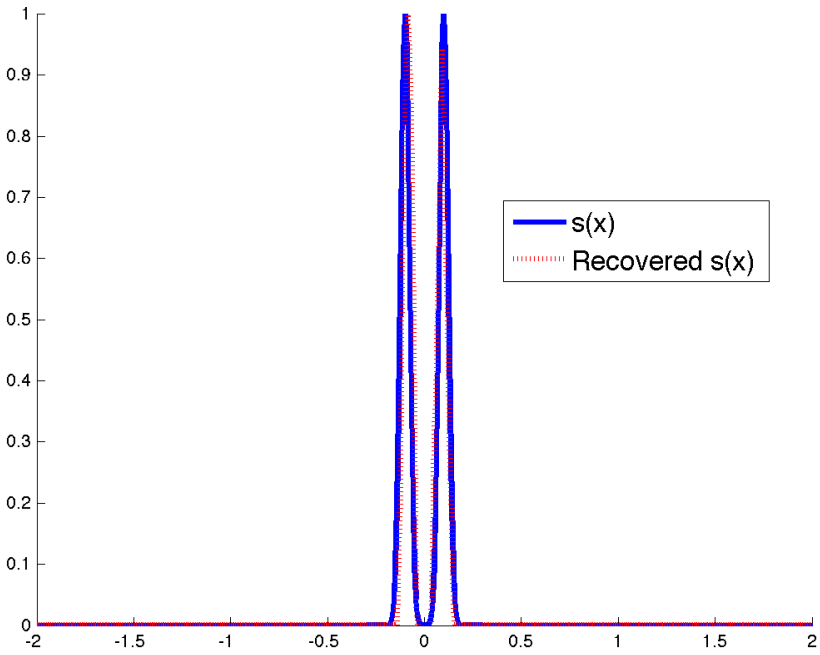


Figure 3.2.5: The true and recovered signals for the toy basis pursuit example

$$b^{k+1} = b^k + s^{k+1} - d^{k+1}.$$

Notice that there are two parameters involved in this system. The first parameter, μ, determines how strongly we seek to force sparsity on our recovered signal. Alternatively, the second parameter, λ, is associated with the dual variable and effectively controls the fidelity of our auxiliary variable to its corresponding original variable. Essentially, λ governs how much of an approximate solution to the true solution we will accept. In the context of our experiment we set $\mu, \lambda = 0.015$. Lastly, we need to implement a stopping criterion. For a stopping criterion given by $s^{k+1} - s^k < 0.001$, and the parameter values shown, we obtain the recovered signal shown in Figure 3.2.5.

3.3 Nonnegative Matrix Factorization

The setup for our *Nonnegative Matrix Factorization* (NMF) problem assumes that we have a known data matrix, $\mathbf{F}$, generated as product of 2 nonnegative matrices ($\mathbf{A}$, $\mathbf{B}$) plus white noise, as shown in Figure 3.3.1. In this context, nonnegativity refers specifically to the entries in the matrix

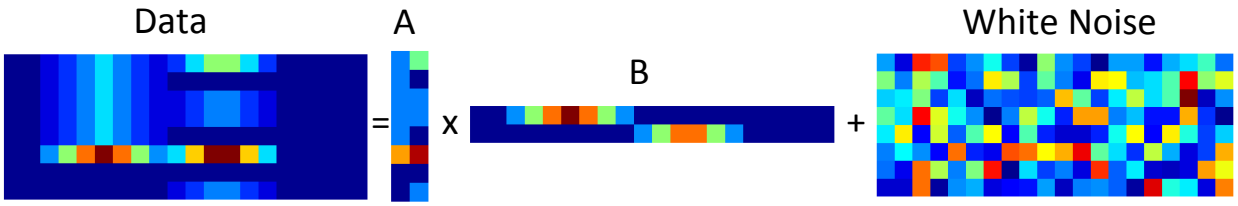


Figure 3.3.1: Schematic of how the blind matrix factorization toy data is generated.

factors $\mathbf{A}$ and $\mathbf{B}$ being nonnegative. The data matrix $\mathbf{F}$ may not always be nonnegative due to the presence of noise. NMF seeks to recover $\mathbf{A}$ and $\mathbf{B}$ from the data matrix $\mathbf{F}$.

Like standard numerical factorization, matrix factorization is not unique. In order to produce a "good" factorization, additional information is needed. For example, if you are told to factor the number 24 into two factors you could return one of several answers: 1x24, 2x12, 3x8, or 4x6. Each factorization produces the same final answer. If instead you are asked to factor 24 into two factors and additionally require that one factor has a magnitude equal to 3, you now only have the factorization 3x8 (which is equivalent to 8x3 given that multiplication is commutative). In matrix factorization, a standard approach to get at a unique solution is to implement an analogous magnitude constraint on one of the matrix factors. This is commonly done by either fixing the matrix norm of one factor to be equal to 1 or to require the rows (or columns) of one matrix factor to all be of unit length. Matrix multiplication is not commutative, but even with the added norm constraint the produced factorization is only unique up to permutation of the rows and columns of the two matrix factors.

To determine the semi-unique matrix factors to the problem described above, we can formulate the following optimization problem

$$\underset{\mathbf{A}_{ij}\geq 0, \mathbf{B}_{ij}\geq 0, ||\mathbf{B}_i||_2=1}{\text{minimize}} \quad \frac{1}{2}||\mathbf{AB}-\mathbf{F}||_F^2$$

to produce a nonnegative factorization. In this formulation, the objective function consists only of the reconstruction error as measured by the matrix Frobenius norm. Explicitly, $||\mathbf{M}||_F^2 =$

trace$(\mathbf{M}^T\mathbf{M})$. In some cases you may have intuition or phenomenological support for additional structure in your matrix factors. A common type of structure that occurs is sparsity: the notion that many entries in your matrix factors should not just be nonnegative, but exactly equal to zero. A desire to have sparse matrix factors can be mathematically implemented by modifying the objective function by adding matrix 1-norm regularization terms to promote general matrix sparsity. With this in mind, the optimization problem can be written as

$$\underset{\mathbf{A}_{ij}\geq 0, \mathbf{B}_{ij}\geq 0, ||\mathbf{B}_i||_2^2=1}{\text{minimize}} \quad \frac{1}{2}||\mathbf{AB}-\mathbf{F}||_F^2 + \mu_A|\mathbf{A}|_1 + \mu_B|\mathbf{B}|_1,$$

where μ_A and μ_B are the parameters weighting the importance of sparsity in the two matrix factors.

Sparse nonnegative matrix factorization has been studied extensively as it arises in many different applications. There are multiple techniques to solve this exact problem as well as a handful of slightly different formulations driving at the same problem. However, there is one issue that is rarely explicitly addressed. In the numerical factorization analogy the problem was posed as "factor 24 into *two* factors, one with magnitude 3" to arrive at the factorization 3x8. What if the specification for two factors is omitted? Again there is no unique solution even if only integer factors are considered! The analogous challenge in matrix factorization manifests as the shared dimension of the matrix factors: the internal factoring dimension.

Given a generic M×N matrix, it can forcibly be factored into an M×L and L×N matrix factor pair, for any L. We refer to this L as the *internal factoring dimension.* In some (limited) applications where the system producing the data is known perfectly and the internal dimension has a specific interpretation, one might know L ahead of time. How does one choose the correct L when not working with data arising in this ideal scenario? A heuristic approach might be to find the "optimal" factors for L=1, 2,... and so on and pick the optimal L based on a chosen metric (like matrix reconstruction error). This feels ad hoc and minimally satisfactory. Consequently, we propose an approach to address this issue in Section 3.5.2.

3.3 Split Bregman for Sparse Nonnegative Matrix Factorization

As a generic example, consider the sparse nonnegative matrix factorization as described. Let $\boldsymbol{F}$ be the MxN data matrix, $\boldsymbol{A}$ be the MxL matrix factor, and $\boldsymbol{B}$ be the LxN matrix factor, with both $\boldsymbol{A}$ and $\boldsymbol{B}$ sparse, and let L be known. For well-posedness we add the magnitude constraint to the matrix $\boldsymbol{B}$ as $||\boldsymbol{B}_i||_2 = 1$. With this added constraint our optimization problem becomes

$$\underset{\boldsymbol{A},\boldsymbol{B},||\boldsymbol{B}_i||_2=1}{\text{minimize}} \quad \frac{1}{2}||\boldsymbol{AB}-\boldsymbol{F}||_F^2 + \mu_A|\boldsymbol{A}|_1 + \mu_B|\boldsymbol{B}|_1,$$

where μ_A and μ_B are the parameters enforcing sparsity on the two matrices. Our first step for implementing Split Bregman is to write the equivalent constrained optimization problem

$$\begin{aligned} \underset{\boldsymbol{A},\boldsymbol{B},\boldsymbol{a},\boldsymbol{b}||\boldsymbol{b}_i||_2=1}{\text{minimize}} \quad & \frac{1}{2}||\boldsymbol{AB}-\boldsymbol{F}||_F^2 + \mu_A|\boldsymbol{a}|_1 + \mu_B|\boldsymbol{b}|_1 \\ \text{subject to} \quad & \boldsymbol{a} = \boldsymbol{A} \\ & \boldsymbol{b} = \boldsymbol{B}. \end{aligned}$$

This constrained problem has augmented Lagrangian given by

$$L(\boldsymbol{a},\boldsymbol{b},\boldsymbol{A},\boldsymbol{B},\boldsymbol{c}_A,\boldsymbol{c}_B) = \frac{1}{2}||\boldsymbol{AB}-\boldsymbol{F}||_F^2+\mu_A|\boldsymbol{a}|_1+\mu_B|\boldsymbol{b}|_1+\frac{\lambda_A}{2}||\boldsymbol{a}-\boldsymbol{A}+\boldsymbol{c}_A||_F^2+\frac{\lambda_B}{2}||\boldsymbol{b}-\boldsymbol{B}+\boldsymbol{c}_B||_F^2.$$

Next, we compute the following updates

$$
\begin{aligned}
\boldsymbol{a}^{n+1} =& \underset{\boldsymbol{a}}{\text{argmin}} \quad L(\boldsymbol{a}, \boldsymbol{b}^n, \boldsymbol{A}^n, \boldsymbol{B}^n, \boldsymbol{c}_A^n, \boldsymbol{c}_B^n), \\
\boldsymbol{b}^{n+1} =& \underset{\boldsymbol{b}, ||\boldsymbol{b}_i||_2=1}{\text{argmin}} \quad L(\boldsymbol{a}^{n+1}, \boldsymbol{b}, \boldsymbol{A}^n, \boldsymbol{B}^n, \boldsymbol{c}_A^n, \boldsymbol{c}_B^n), \\
\boldsymbol{A}^{n+1} =& \underset{\boldsymbol{A}}{\text{argmin}} \quad L(\boldsymbol{a}^{n+1}, \boldsymbol{b}^{n+1}, \boldsymbol{A}, \boldsymbol{B}^n, \boldsymbol{c}_A^n, \boldsymbol{c}_B^n), \\
\boldsymbol{B}^{n+1} =& \underset{\boldsymbol{B}}{\text{argmin}} \quad L(\boldsymbol{a}^{n+1}, \boldsymbol{b}^{n+1}, \boldsymbol{A}^{n+1}, \boldsymbol{B}, \boldsymbol{c}_A^n, \boldsymbol{c}_B^n), \\
\boldsymbol{c}_A^{n+1} =& \boldsymbol{c}_A^n + \boldsymbol{A}^{n+1} - \boldsymbol{a}^{n+1}, \\
\boldsymbol{c}_B^{n+1} =& \boldsymbol{c}_B^n + \boldsymbol{B}^{n+1} - \boldsymbol{b}^{n+1}.
\end{aligned}
$$

We note that there is more than one order in which the user can choose to update the variables although only one is shown.

Updating $\boldsymbol{a}$ and $\boldsymbol{b}$ can be done using the matrix shrinkage operator and then, for $\boldsymbol{b}$, normalizing with respect to the matrix norm. Resultantly, we have

$$
\begin{aligned}
\boldsymbol{a}^{n+1} &= \text{shrinkage}(\boldsymbol{A}^n + \boldsymbol{c}_A^n - \frac{\mu_A}{\lambda_A}), \\
\boldsymbol{b}^{n+1} &= \text{shrinkage}\frac{(\boldsymbol{B}^n + \boldsymbol{c}_B^n - \frac{\mu_B}{\lambda_B})}{||(\boldsymbol{B}^n + \boldsymbol{c}_B^n - \frac{\mu_B}{\lambda_B})||_F}.
\end{aligned}
$$

The updates for both $\boldsymbol{A}$ and $\boldsymbol{B}$ are solutions to unconstrained minimizations of differentiable functions and consequently have closed form solutions given by

$$
\begin{aligned}
\boldsymbol{A}^{n+1} &= (\boldsymbol{F}(\boldsymbol{B}^n)^T + \lambda_a(\boldsymbol{a}^{n+1} - \boldsymbol{c}_A^n))(((\boldsymbol{B}^n)(\boldsymbol{B}^n)^T + \lambda_A \boldsymbol{I})^{-1}, \\
\boldsymbol{B}^{n+1} &= (((\boldsymbol{A}^{n+1})^T(\boldsymbol{A}^{n+1}) + \lambda_B \boldsymbol{I})^{-1}((\boldsymbol{A}^{n+1})^T \boldsymbol{F} + \lambda_B(\boldsymbol{b}^{n+1} - \boldsymbol{c}_B^n)).
\end{aligned}
$$

Finally, the updates of the dual variables $\boldsymbol{c}_A$ and $\boldsymbol{c}_B$ are still completed using gradient ascent.

3.4 Split Bregman Applied to FAL

3.4 Writing the FAL Model for Split Bregman

Based on the mathematical model describing FAL, our goal becomes to estimate $\boldsymbol{\rho}$ and $\boldsymbol{C}$ at each burst while minimizing the sum of square error, $||\boldsymbol{G} - \boldsymbol{\rho C}||_2^2$. The columns of $\boldsymbol{\rho}$ will act as backscatter feature vectors to be used for the purpose of identification of aerosols, while the rows of $\boldsymbol{C}$ will be used for tracking each aerosol over time. We follow the problem formulation proposed in [34] for 1-norm regularized optimization[2]. The resultant optimization problem is

$$\underset{||\boldsymbol{\rho}_i||_2=1, \boldsymbol{\rho}_{ij}\geq 0, \boldsymbol{C}_{ij}\geq 0}{\text{minimize}} \quad \mu_\rho|\boldsymbol{\rho}|_1 + \mu_C|\boldsymbol{C}|_1 + \frac{1}{2}||\boldsymbol{G} - \boldsymbol{\rho C}||_F^2,$$

where $\boldsymbol{G}$ is the preprocessed data from a single burst. An example of $\boldsymbol{G}$ is the matrix shown in Figure 3.1.10. The non-negativity constraints on $\boldsymbol{\rho}$ and $\boldsymbol{C}$ come from the physical interpretation of the matrices: it does not make sense to have negative concentration or a negative backscatter cross section. This optimization problem is nonnegative matrix factorization and it can be efficiently solved using the Split Bregman algorithm.

As we have previously seen, the first step to implementation of the Split Bregman algorithm is to create the equivalent constrained optimization problem given by

$$\begin{aligned} \underset{\boldsymbol{\rho},\boldsymbol{C},r,c,||r_i||_2=1, r_{ij}\geq 0, c_{ij}\geq 0}{\text{minimize}} \quad & \mu_\rho|\boldsymbol{\rho}|_1 + \mu_C|\boldsymbol{C}|_1 + \frac{1}{2}||\boldsymbol{G} - \boldsymbol{\rho C}||_F^2 \\ \text{subject to} \quad & r = \boldsymbol{\rho} \\ & c = \boldsymbol{C}. \end{aligned}$$

The corresponding augmented Lagrangian is

[2]Implementation of the Split Bregman algorithm to produce results is based on code written by the author of this document and can be made available upon request.

$$L(r, \boldsymbol{\rho}, b_r, c, \boldsymbol{C}, b_c) = \mu_\rho |r| + \mu_C |c| + \frac{1}{2}||\boldsymbol{G} - \boldsymbol{\rho}\boldsymbol{C}||_F^2 + \frac{\lambda_\rho}{2}||r - \boldsymbol{\rho} - b_r||_F^2 + \frac{\lambda_C}{2}||c - \boldsymbol{C} - b_C||_F^2.$$

From the augmented Lagrangian we have the following updates

$$\begin{aligned}
\boldsymbol{C}^{n+1} =& \underset{\boldsymbol{C}}{\text{argmin}} \quad L(r^n, \boldsymbol{\rho}^n, b_r^n, c^n, \boldsymbol{C}, b_c^n), \\
c^{n+1} =& \underset{c, c_{ij} \geq 0}{\text{argmin}} \quad L(r^n, \boldsymbol{\rho}^n, b_r^n, c, \boldsymbol{C}^{n+1}, b_c^n), \\
b_C^{n+1} =& b_C^n + \boldsymbol{C}^{n+1} - c^{n+1}, \\
\boldsymbol{\rho}^{n+1} =& \underset{\boldsymbol{\rho}}{\text{argmin}} \quad L(r^n, \boldsymbol{\rho}, b_r^n, c^{n+1}, \boldsymbol{C}^{n+1}, b_c^{n+1}), \\
r^{n+1} =& \underset{r, r_{ij} \geq 0, ||r_i||_2 = 1}{\text{argmin}} \quad L(r, \boldsymbol{\rho}^{n+1}, b_r^n, c^{n+1}, \boldsymbol{C}^{n+1}, b_c^{n+1}), \\
b_\rho^{n+1} =& b_\rho^n + \boldsymbol{\rho}^{n+1} - r^{n+1}.
\end{aligned}$$

Updating $\boldsymbol{C}$ is, as we have seen before, an unconstrained minimization of a differentiable function yielding a closed form for the update given by

$$\boldsymbol{C}^{n+1} = (((\boldsymbol{\rho}^n)^T(\boldsymbol{\rho}^n) + \lambda_C I)^{-1}((\boldsymbol{\rho}^n)^T \boldsymbol{G} + \lambda_C(c^n - b_C^n)).$$

Due to the non-negativity constraint on c we can compute the update as the maximum of the shrinkage operator and zero to yield

$$c^{n+1} = \max(\boldsymbol{C}^{n+1} + b_C^n - \frac{\mu_C}{\lambda_C},\, 0).$$

The update of the dual variable b_C comes, again, from gradient ascent. Using the same rationale as for the updates of $\boldsymbol{C}$ and its associated variables, we obtain the updates for the $\boldsymbol{\rho}$ associated variables to be

$$\boldsymbol{\rho}^{n+1} = (\boldsymbol{G}(\boldsymbol{C}^{n+1})^T + \lambda_\rho(r^n - b_r^n))(((\boldsymbol{C}^{n+1})(\boldsymbol{C}^{n+1})^T + \lambda_\rho I)^{-1},$$
$$r_i^{n+1} = \frac{(\max(\boldsymbol{\rho}^{n+1} + b_r^n - \frac{\mu_C}{\lambda_C}, 0))_i}{||(\max(\boldsymbol{\rho}^{n+1} + b_r^n - \frac{\mu_C}{\lambda_C}, 0))_i||_2}.$$

Pseudocode for can be seen in [34]. Note that there is an error in the formula for the update of $\boldsymbol{\rho}$ in [34].

3.4 FAL Factorization Results

In the data we have, there are several data cubes containing a single labeled aerosol release with known background and release burst indices. Although only one release is labeled, there are in fact three releases per cube. Seven of the data cubes contain two non-overlapping labeled aerosols with known background and release burst indices. Due to the aerosols not overlapping in burst we can use a model in which only one aerosol is present at a time, i.e. $L = 1$.

Figure 3.4.1 shows the quality of data reconstruction for both the Split Bregman and Kalman filter approaches applied to a single data cube containing one labeled release of the bio-toxin simulant Ovalbumin. The order of magnitude of the residuals for both methods is 10^{-2} resulting in errors up to 10% of the size of the original scale. However, we note that the overall error for the Split Bregman approach [34] is smaller than that of the originally proposed Kalman filter approach in [35]. Unless otherwise stated, results were generated using parameter values set at

- $\lambda_\rho = \lambda_C = 0.05$,
- $\mu_C = \mu_\rho = \lambda_\rho^2$,
- with a minimum range of interest equal to 0.05km.

In Figure 3.4.2 we see all of the concentration and corresponding feature vectors for an entire data cube containing a single labeled release of Ovalbumin. We notice several things about these results. First, the angle of the aerosol cloud in the concentration image is consistent with the fact that aerosol is released at the end of a tunnel and blown, by a fan, towards the near end of the tunnel. Second, the appearance of concentration does not occur until later than the labeled starting burst

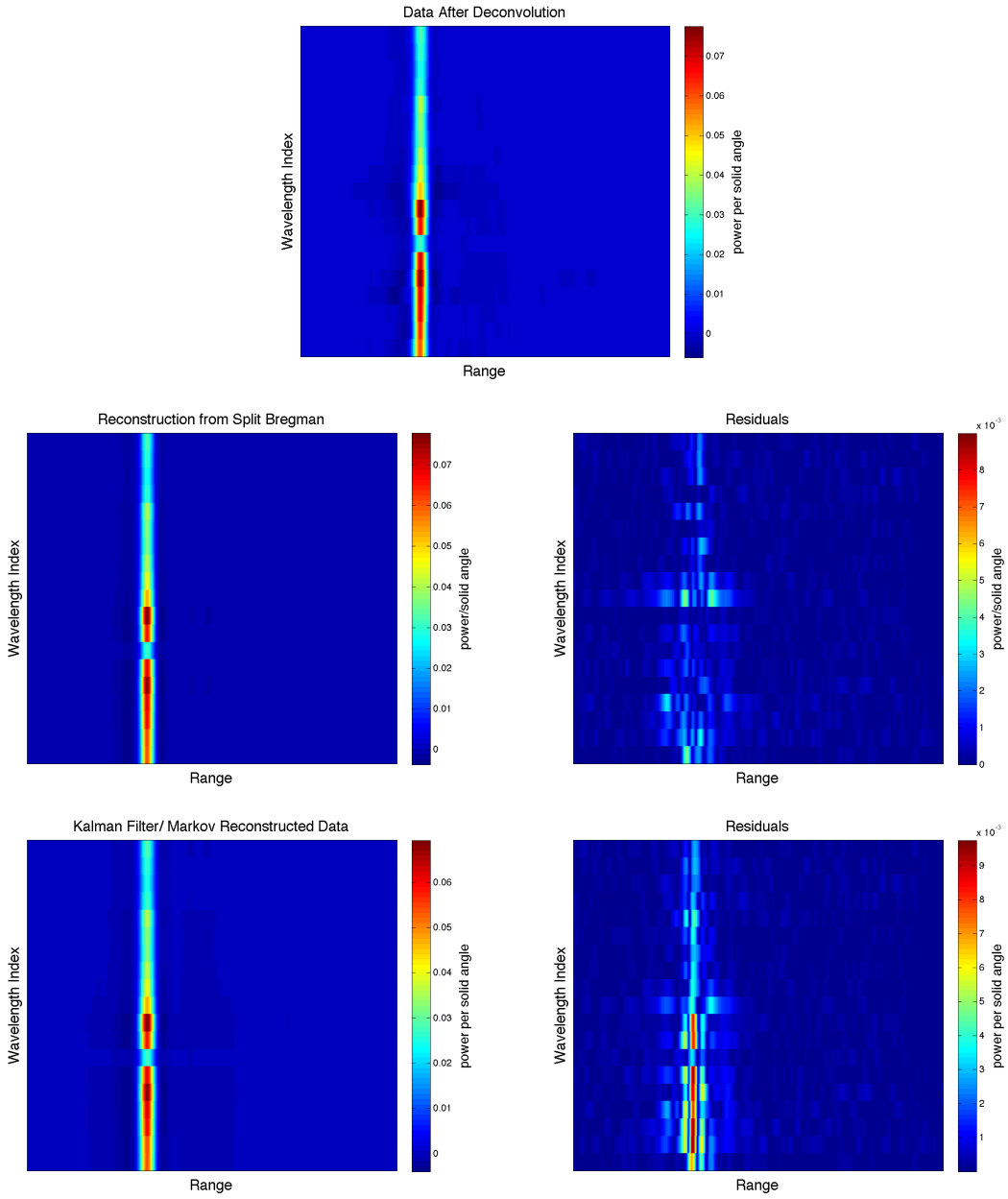


Figure 3.4.1: The top image is data from a single burst containing aerosol. Middle left image is the reconstruction of the data obtained from the output of Split Bregman while the middle right image is the residuals of the reconstruction on the left. Lower left image is the reconstruction of the data obtained from the output of the Kalman filtering processing while the lower right image is the residuals of the reconstruction on the left.

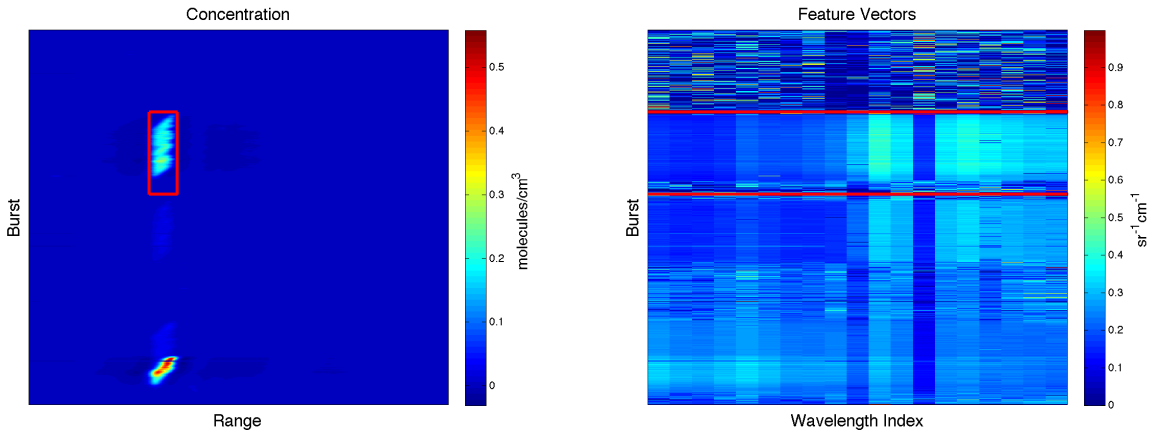


Figure 3.4.2: The image on the left contains all the concentration vectors for a single data cube with a labeled Ovalbumin release during the bursts bounded by the red box. On the right is an image of all of the feature vectors produced by Split Bregman. Burst index increases from top to bottom and range increases from left to right.

index and it also disappears before the terminating burst index. The high concentrations which show up at later bursts could be an additional unlabeled release, but this cannot be confirmed as this data cube only had one of three releases labeled. Additionally, we can see that when the concentration is nonzero over a set of bursts (an aerosol cloud is present) the feature vectors become more uniform over the corresponding group of bursts.

There were a total of four data cubes, obtained on three different days, which had labeled releases of Ovalbumin. Selecting all of the feature vectors produced at labeled bursts from the four data cubes allows us to discuss the variance of a feature vector. In Figure 3.4.3 we see the 95% confidence interval for an Ovalbumin feature vector taken from the four different data cubes. A great deal of the variance of the feature vectors can be attributed to the fact that the output of Split Bregman seems to consistently start detecting presence late and terminates too early.

Moreover, if we perform Split Bregman on a data cube with two labeled releases, results of which are shown in Figure 3.4.4, suggest that the feature vectors detected during the two different labeled releases do in fact contain differentiating information. The quantity of differentiating information will be presented in the Discussion section.

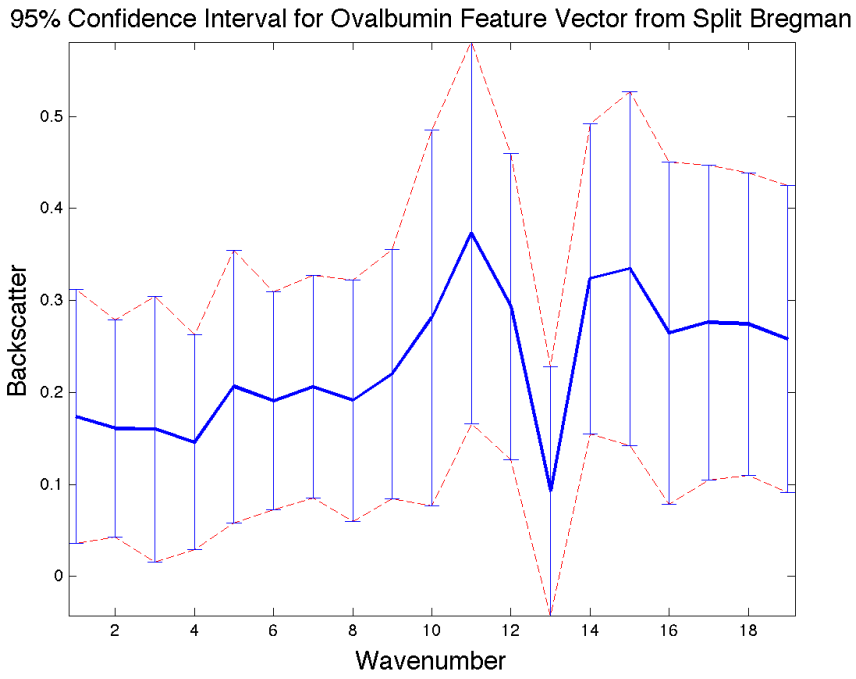


Figure 3.4.3: The 95% confidence interval for an Ovalbumin feature vector.

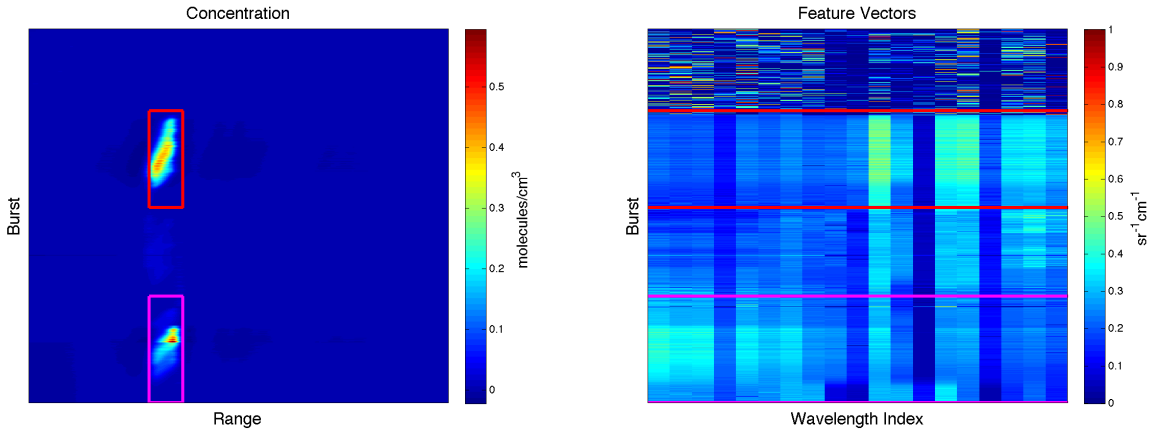


Figure 3.4.4: The image on the left contains all the concentration vectors for a single data cube with a labeled Ovalbumin release (bounded by the red box) and a labeled Smoke release (bounded by the magenta box). On the right is an image of all of the feature vectors produced by Split Bregman. Burst index increases from top to bottom and range increases from left to right.

The results of Split Bregman on the FAL unmixing problem show great potential, but is not free from its own challenges. For example, the quality of the preprocessed data will effect the performance of the algorithm. Additionally, the method is highly sensitive to parameter value selection. Due to the updates which depend on the parameters (r and c particularly), it is possible to end up with non numerical solutions. These are important challenges to keep in mind as areas with room for improvement in future work.

3.5 Discussion

3.5 Modification of Split Bregman Applied to FAL

This discussion piece primarily focuses on the choice of regularization norms being used in the optimization problem. While the phenomenology of the model suggests that concentration should be sparse with respect to range at a given burst (due to the localization of an aerosol cloud), there is no such argument to support the claim of sparsity in the backscatter feature vector. Each molecule will have some nonzero backscatter cross section for each of the wavelengths.

Solving

$$\underset{||\boldsymbol{\rho}_i||_2=1,\boldsymbol{\rho}_{ij}\geq 0,\boldsymbol{C}_{ij}\geq 0}{\text{minimize}} \quad \mu_\rho|\boldsymbol{\rho}|_1+\mu_C|\boldsymbol{C}|_1+\frac{1}{2}||\boldsymbol{G}-\boldsymbol{\rho}\boldsymbol{C}||_F^2 \tag{3.5.1}$$

using the Split Bregman algorithm for a single data cube containing one labeled release of Ovalbumin produces the results shown in Figure 3.5.1.

Since we have labeled burst indices for background, Ovalbumin, and Smoke, we can plot all of the one dimensional feature vectors from labeled bursts according to their label, shown in Figure 3.5.2. We see that the greatest variance occurs in the background feature vectors. Additionally, we see that there appears to be strong similarities in the shape of a feature vector within an aerosol class (Ovalbumin or Smoke) and across class differences. The significance of the across class differences is highlighted in Figure 3.5.3. There is enough differentiating information contained in the feature vectors to obtain separation of the classes in the space spanned by the first three

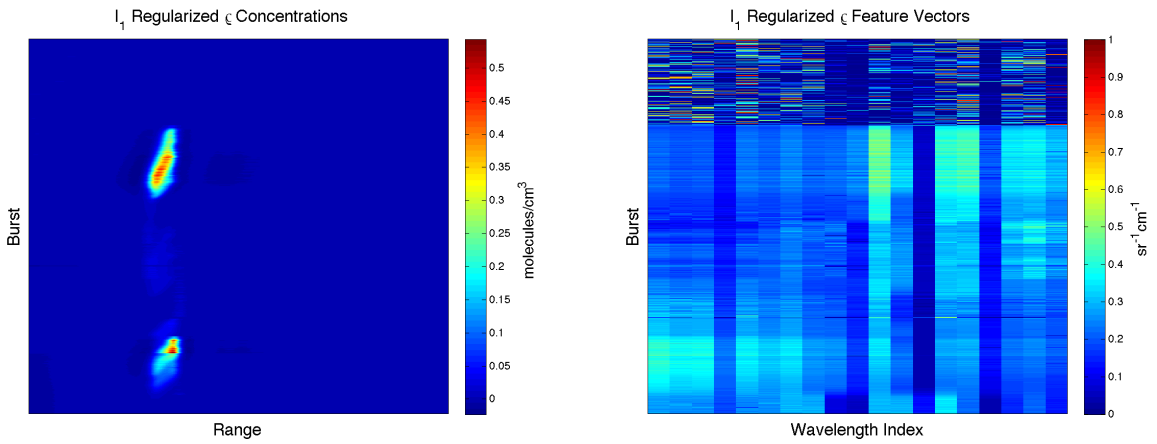


Figure 3.5.1: The image on the left contains all the concentration vectors for a single data cube with labeled Ovalbumin and Smoke releases. On the right is an image of all of the feature vectors produced by 1-norm regularized Split Bregman. Burst index increases from top to bottom and range increases from left to right.

principal components of the data cube's feature vectors. We note that with each of the aerosols you see one primary direction in which there is nontrivial variance.

Use of the 1-norm for regularization of $\boldsymbol{\rho}$ introduces the need for one auxiliary variable, which in turn puts into action a constraint and corresponding dual variable and fidelity parameter. Thus, by considering a 2-norm regularization of $\boldsymbol{\rho}$ we reduce the complexity of the model from six variables and four parameters to four variables and three parameters. The optimization problem for a 2-norm regularized $\boldsymbol{\rho}$ is given by

$$\underset{||\boldsymbol{\rho}_i||_2=1,\boldsymbol{\rho}_{ij}\geq 0,\boldsymbol{C}_{ij}\geq 0}{\text{minimize}} \quad \mu_\rho||\boldsymbol{\rho}||_2 + \mu_C|\boldsymbol{C}|_1 + \frac{1}{2}||\boldsymbol{G}-\boldsymbol{\rho}\boldsymbol{C}||_F^2. \tag{3.5.2}$$

Split Bregman can still be used to solve this optimization problem. We take this opportunity to note that 2-norm regularization on $\boldsymbol{\rho}$ can be considered redundant. Since the normalization (for well posedness) is placed on $\boldsymbol{\rho}$ via row-wise 2-norm normalization, the 2-norm regularization term in the objective function is in actuality an additive constant. With this acknowledged, the implementation and results presented were produced with this redundancy in place. This choice was made so that our initial objective function was consistent with the form of the objective function used by [39] which will be discussed as a point of comparison. Throughout the the remainder of

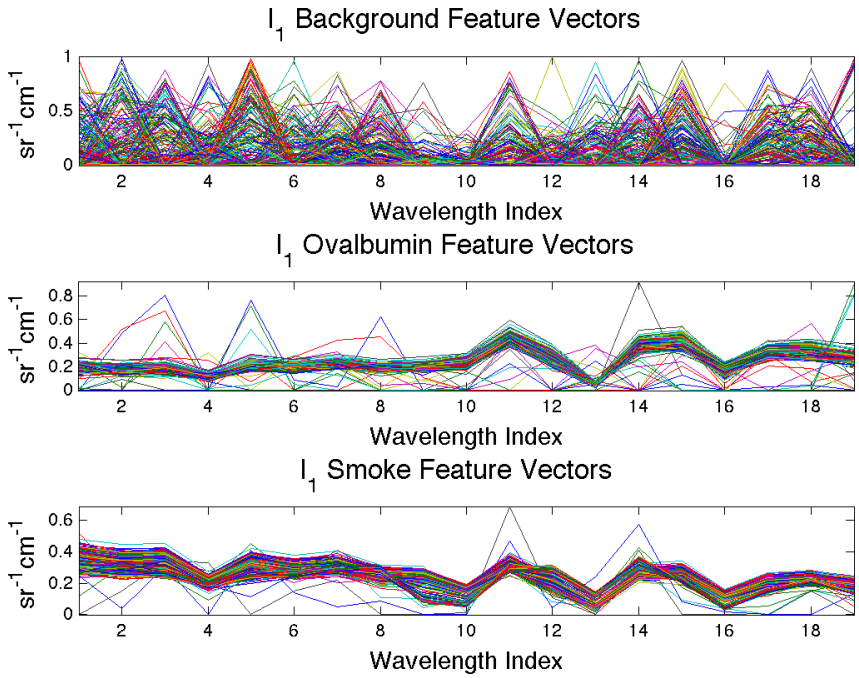


Figure 3.5.2: All the feature vectors from labeled bursts grouped according to label as produced by 1-norm regularized Split Bregman.

this chapter if l_2 is present in a figure title, or caption, it refers to results from using Split Bregman to solve Equation 3.5.2. Similarly, l_1 indicates results generated from using Split Bregman to solve Equation 3.5.1. The expressions "1-norm Split Bregman" and "l_1" are used interchangeably, as are "2-norm Split Bregman" and "l_2."

With only one non differentiable term in the objective function we only need to introduce one auxiliary variable to produce the constrained problem

$$\begin{aligned} \underset{\boldsymbol{C},c,||\boldsymbol{\rho}_i||_2^2=1,\boldsymbol{\rho}_{ij}\geq 0,c_{ij}\geq 0}{\text{minimize}} \quad & \frac{\mu_\rho}{2}||\boldsymbol{\rho}||_F^2 + \mu_C|\boldsymbol{C}|_1 + \frac{1}{2}||\boldsymbol{G}-\boldsymbol{\rho}\boldsymbol{C}||_F^2 \\ \text{subject to} \quad & c = \boldsymbol{C}, \end{aligned}$$

with corresponding augmented Lagrangian

$$L(\boldsymbol{\rho}, c, \boldsymbol{C}, b_c) = \frac{\mu_\rho}{2}||\boldsymbol{\rho}||_2^2 + \mu_C|c| + \frac{1}{2}||\boldsymbol{G}-\boldsymbol{\rho}\boldsymbol{C}||_F^2 + \frac{\lambda_C}{2}||c-\boldsymbol{C}-b_C||_F^2.$$

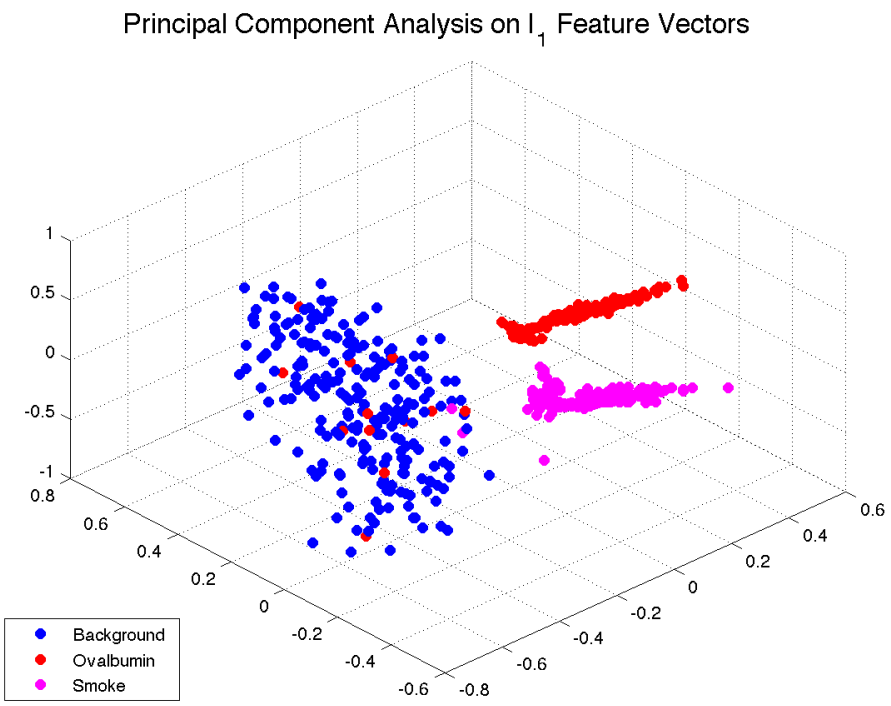


Figure 3.5.3: All labeled feature vectors projected onto the space spanned by the first three principal components of the full data cube feature vector set as produced by 1-norm regularized Split Bregman.

From the augmented Lagrangian we have the following updates

$$\begin{aligned}
\boldsymbol{C}^{n+1} &= \underset{\boldsymbol{C}}{\text{argmin}} \quad L(r^n, \boldsymbol{\rho}^n, b_r^n, c^n, \boldsymbol{C}, b_c^n), \\
c^{n+1} &= \underset{c, c_{ij} \geq 0}{\text{argmin}} \quad L(r^n, \boldsymbol{\rho}^n, b_r^n, c, \boldsymbol{C}^{n+1}, b_c^n), \\
b_C^{n+1} &= b_C^n + \boldsymbol{C}^{n+1} - c^{n+1}, \\
\boldsymbol{\rho}^{n+1} &= \underset{\boldsymbol{\rho}}{\text{argmin}} \quad L(r^n, \boldsymbol{\rho}, b_r^n, c^{n+1}, \boldsymbol{C}^{n+1}, b_c^{n+1}).
\end{aligned}$$

The updates for the first three variables are the same as for the 1-norm regularized problem, while now the update for $\boldsymbol{\rho}$ has a closed form given by

$$\boldsymbol{\rho}_i^{n+1} = ((\boldsymbol{G}(\boldsymbol{C}^{n+1})^T)(\mu_\rho I + (\boldsymbol{C}^{n+1})(\boldsymbol{C}^{n+1})^T)^{-1})_i.$$

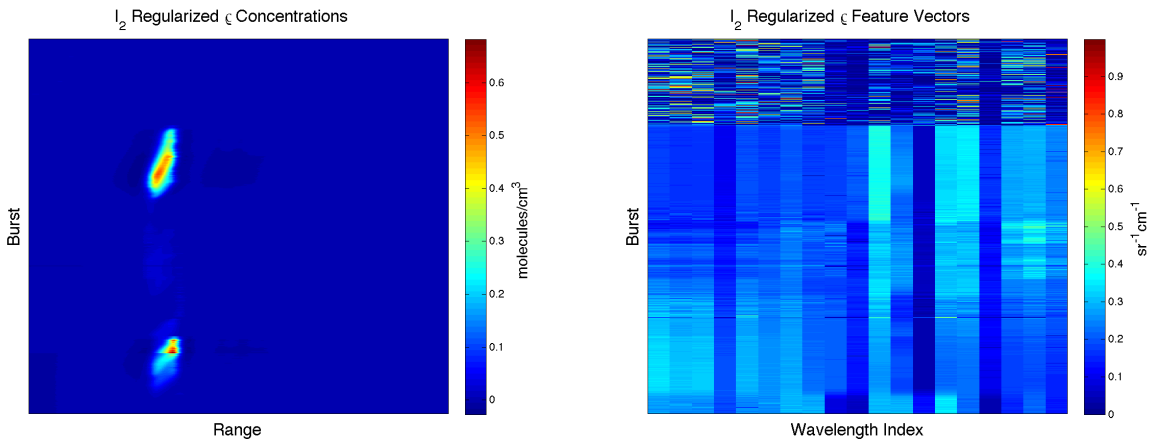


Figure 3.5.4: The image on the left contains all the concentration vectors for a single data cube with labeled Ovalbumin and Smoke releases. On the right is an image of all of the feature vectors produced by 2-norm regularized Split Bregman. Burst index increases from top to bottom and range increases from left to right.

Using the same value for the three parameters as were used in the 1-norm regularized problem, we produce the results shown in Figure 3.5.4. Figure 3.5.5 shows the magnitude of the difference between the concentration estimates and feature vectors produced the two different approaches. As an added benefit, we see a decrease in computational time: l_1 regularization takes about 12 seconds to solve for a single data cube, while l_2 regularization takes about 3 seconds.

The feature vectors produced by the 2-norm regularized algorithm plotted and grouped according to label are shown in Figure 3.5.6. Notice that there is an apparent decrease in the within class variance in Figure 3.5.6 compared to Figure 3.5.2. This reduction in variance is shown, also, in the three dimensional PCA of the 2-norm feature vectors shown in Figure 3.5.7. Although the separation of the aerosol classes from the background class is weaker in Figure 3.5.7, the separation between the aerosol classes is still clear. This begs the question whether it is appropriate to perform the detection based on the feature vector or the concentration. If you want to detect based on the feature vector, as proposed in [34], the 1-norm regularized feature vectors have stronger separability but at the expense of greater variance.

Running both the 1-norm and 2-norm regularization of feature vectors on multiple data cubes and plotting the two dimensional PCA of feature vectors suggests the superiority of the 2-norm approach. Principal components were determined based on the feature vectors of the entire data

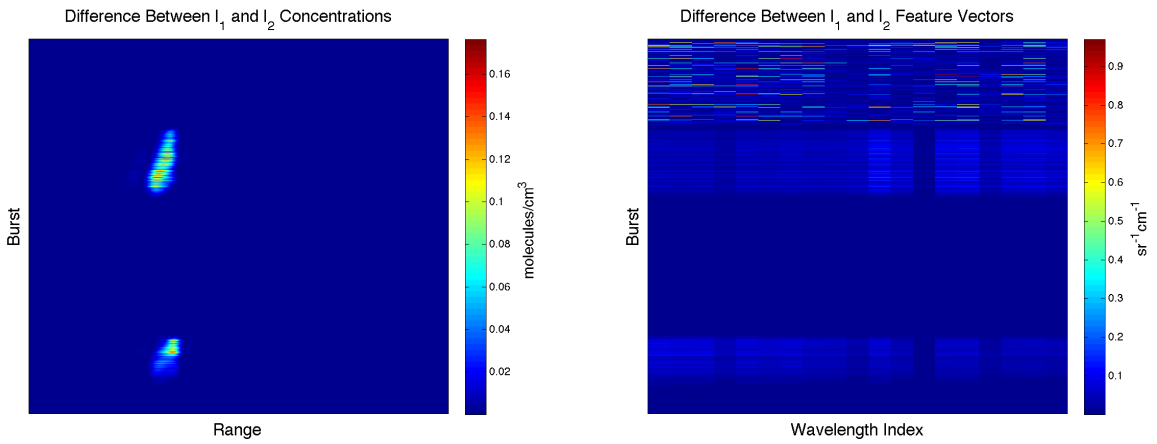


Figure 3.5.5: The image on the left contains the magnitude of the difference between the 1-norm and 2-norm concentration vectors on the same data cube with labeled Ovalbumin and Smoke releases. On the right is an image of all of the magnitude of the difference between the 1-norm and 2-norm feature vectors. Burst index increases from top to bottom and range increases from left to right.

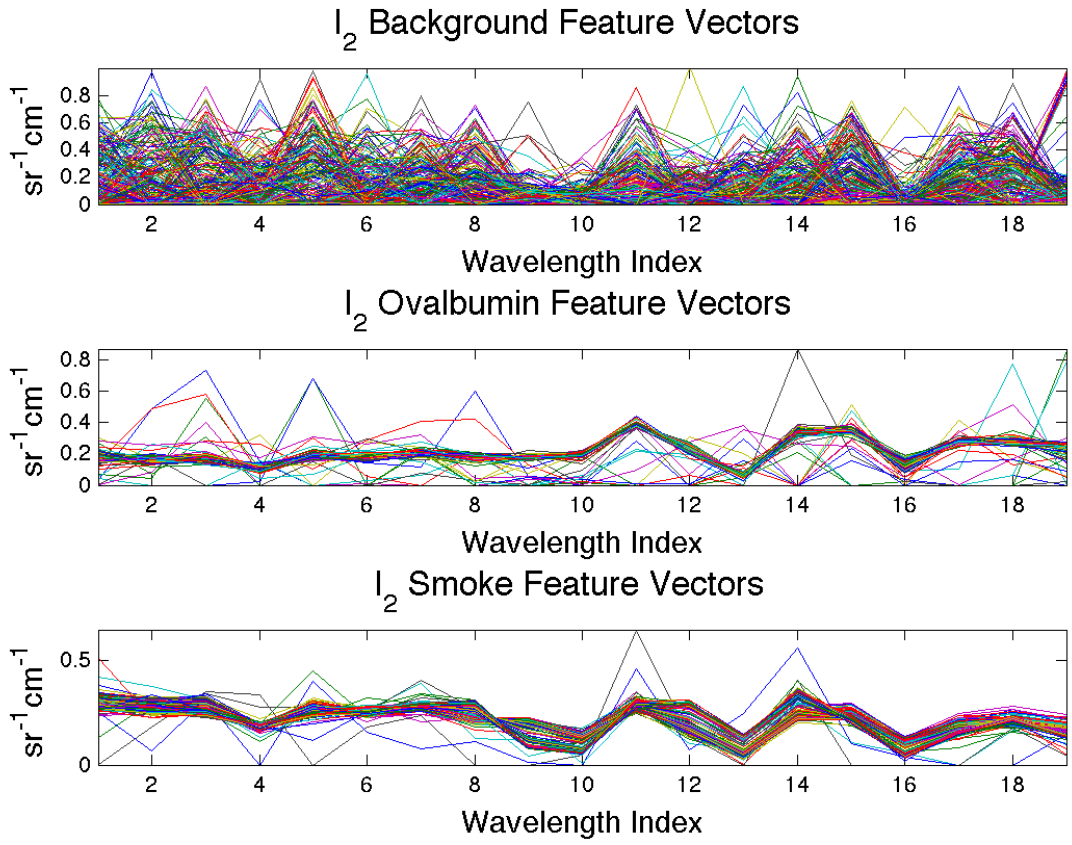


Figure 3.5.6: All the feature vectors from labeled bursts grouped according to label as produced by 2-norm regularized Split Bregman.

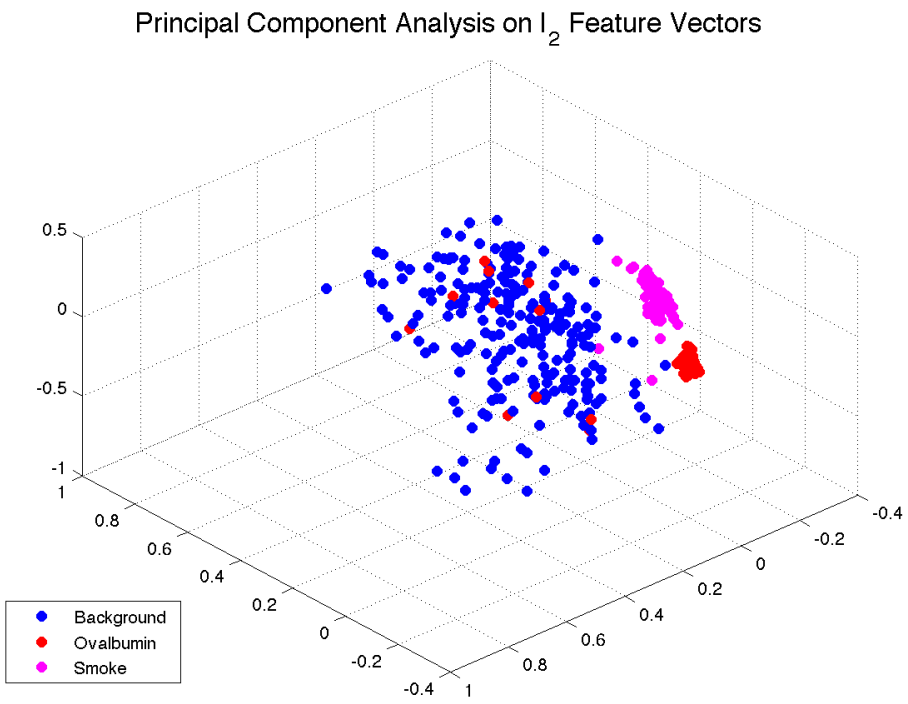


Figure 3.5.7: All labeled feature vectors projected onto the space spanned by the first three principal components of the full data cube feature vector set as produced by 2-norm regularized Split Bregman.

cube, but only feature vectors corresponding to concentrations larger than the upper bound of the 99% confidence interval of the background concentration were plotted. Figures 3.5.8,3.5.9, and 3.5.10 are examples of the 2-norm Split Bregman feature vectors reducing the variance, relative to the 1-norm feature vectors, within obvious clusters. Alternatively, Figure 3.5.11 is an example of a data cube on which the number of distinct clusters differs between the 1-norm and 2-norm feature vectors. Finally, Figure 3.5.12 shows an example of a data cube in which neither approach produced strong separation in the two dimensional case. However, it could be argued that there is a more obvious breaking point between two clusters for the feature vectors from the 2-norm approach.

Visualization of the feature vectors in two dimensions based on a concentration threshold suggests that the 2-norm regularization of feature vectors is advantageous. We do not mean to suggest that detection should be done based on concentration since determining a threshold requires previous knowledge of background burst indices and is consequently impractical for an online

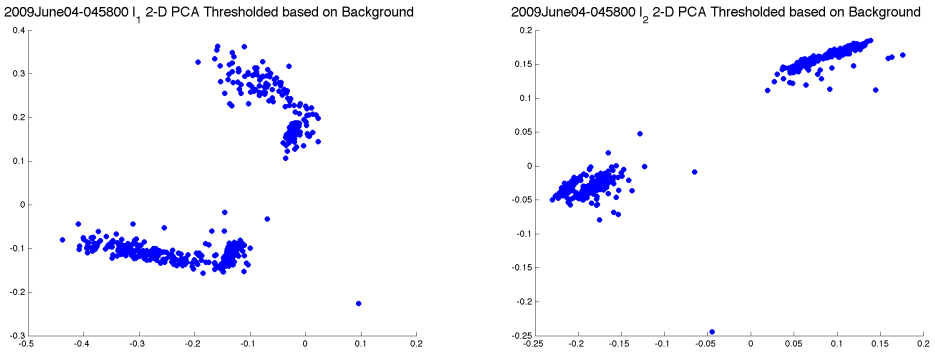


Figure 3.5.8: The left image is the 2-D PCA projection of background concentration selected feature vectors from the 1-norm Split Bregman. On the right, is the 2-D PCA projection of background concentration selected feature vectors from the 2-norm Split Bregman.

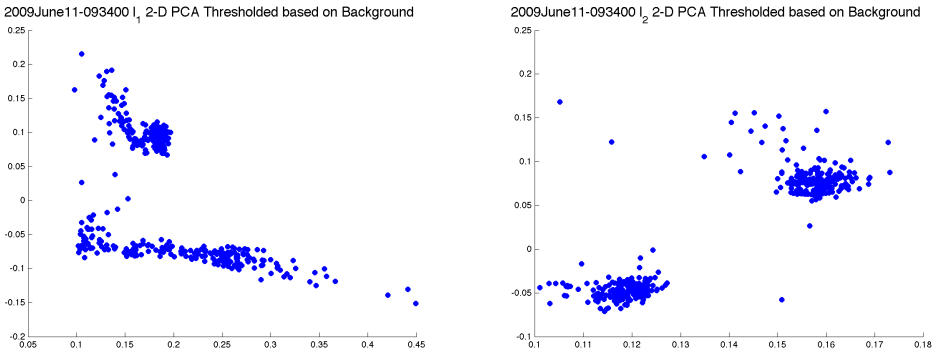


Figure 3.5.9: The left image is the 2-D PCA projection of background concentration selected feature vectors from the 1-norm Split Bregman. On the right, is the 2-D PCA projection of background concentration selected feature vectors from the 2-norm Split Bregman.

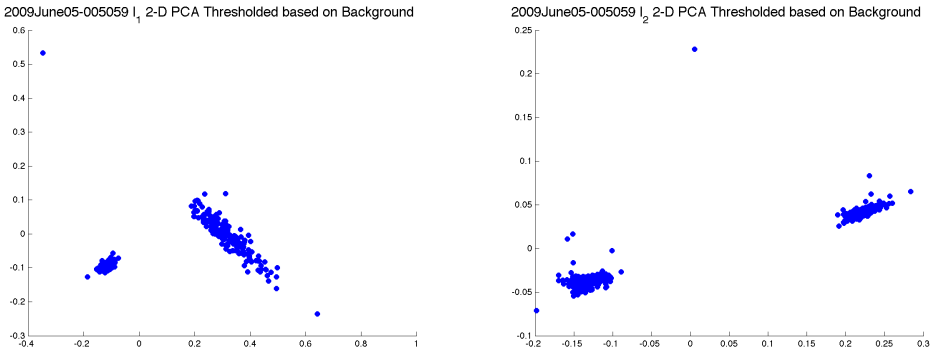


Figure 3.5.10: The left image is the 2-D PCA projection of background concentration selected feature vectors from the 1-norm Split Bregman. On the right, is the 2-D PCA projection of background concentration selected feature vectors from the 2-norm Split Bregman.

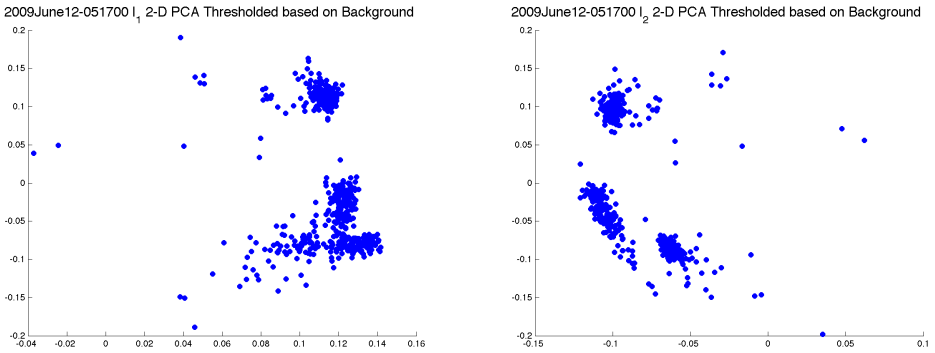


Figure 3.5.11: The left image is the 2-D PCA projection of background concentration selected feature vectors from the 1-norm Split Bregman. On the right, is the 2-D PCA projection of background concentration selected feature vectors from the 2-norm Split Bregman.

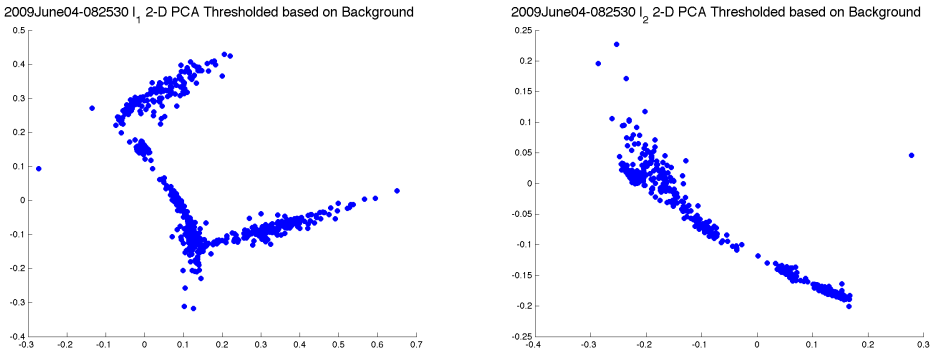


Figure 3.5.12: The left image is the 2-D PCA projection of background concentration selected feature vectors from the 1-norm Split Bregman. On the right, is the 2-D PCA projection of background concentration selected feature vectors from the 2-norm Split Bregman.

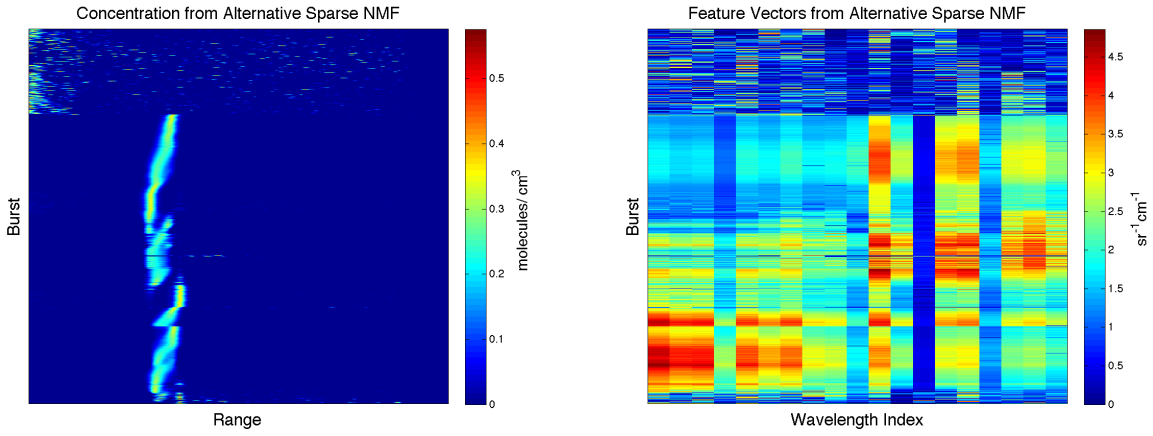


Figure 3.5.13: The image on the left contains all the concentration vectors for a single data cube with labeled Ovalbumin and Smoke releases. On the right is an image of all of the feature vectors produced by the non-negative matrix factorization method described in [39]. Burst index increases from top to bottom and range increases from left to right.

algorithm. Identification has successfully been done based on the feature vectors in [34, 40]. Although the background is not strongly separated from the other classes for all data cubes in lower dimensions, this does not imply a lack of separability in the ambient space of feature vectors.

It is important to note that Split Bregman is not the unique method for solving this problem. As we have previously stated, the problem is essentially a sparse non-negative matrix factorization (NMF) problem. The area of NMF has grown and there are many methods which could be applicable. For example, another method proposed in [39] also has one 1-norm regularized factor and one 2-norm regularized factor. Additionally, the method in [39] only requires at most two parameter values based on the desired level of sparseness. Implementing the code provided in the paper one can obtain the concentrations and feature vectors shown in Figure 3.5.13. In order to implement this method we first remove all negative values from the data matrix (which are there as an artifact of the side lobes after deconvolution). Moreover, we choose that we do not want to enforce any level of sparsity on the feature vectors while we would like a sparsity level of 0.8 for the concentration (as defined in [39]). Finally, we note a difference in the color scale between this method and the Split Bregman approach. Our Split Bregman approach regularizes the feature vector matrix, while the NMF method regularizes the concentration matrix.

The Split Bregman approach seems to be superior in several ways when the output is compared to that of the [39] method. For example, the feature vectors have lower variance and there is less concentration showing up in the beginning background bursts. However, the concentration vectors from Split Bregman show less movement of the aerosol than the concentrations shown in Figure 3.5.13. We do not claim optimality in either case, but rather wish to highlight that multiple approaches to the problem could be taken each with their own challenges and assets.

Another area for potential improvement would be to use elastic net regularization on the concentration. Concentration should be sparse over range, but it is also correlated over range within the aerosol cloud. As we have stated, 1-norm regularization tends to suppress all but one of the variables from within a correlated group of variables [37, 38]. Since aerosol clouds at a near range

pose a great threat, it is important to take necessary measures to ensure the most accurate ranges for aerosol presence.

3.5 Internal Dimension Identification via Average Vector Sparsity

The first approach leverages the nonnegativity constraint through the addition of another regularization term in the objective function. Consider a nonnegative matrix $\mathbf{A}$. If the average value of a row of $\mathbf{A}$ is zero then necessarily all the elements of the row must be zero to honor the nonnegativity constraint. As a result, if we consider the vector $\boldsymbol{\alpha}$ that contains the row averages for $\mathbf{A}$, by promoting sparsity in $\boldsymbol{\alpha}$ extraneous dimensions might be identified. This could be paired with general matrix sparsity as in the previous formulation. The constrained optimization problem becomes

$$\begin{aligned} \underset{\mathbf{A},\mathbf{B},||\mathbf{B}_i||_2^2=1}{\text{minimize}} \quad & \frac{1}{2}||\mathbf{AB}-\mathbf{F}||_F^2+\mu_A|\mathbf{A}|_1+\mu_B|\mathbf{B}|_1+\mu_\alpha|\boldsymbol{\alpha}|_1+\mu_\beta|\boldsymbol{\beta}|_1 \\ \text{subject to} \quad & \boldsymbol{\alpha}=\frac{1}{M}\mathbf{1}_{1xM}\mathbf{A} \\ & \mathbf{A}_{ij}\geq 0 \\ & \mathbf{B}_{ij}\geq 0, \\ & \boldsymbol{\alpha}_j\geq 0, \end{aligned}$$

where $\mathbf{A}$ is the M×L matrix factor, $\mathbf{B}$ is the L×N matrix factor, and $\mu_A,\ \mu_B$, and μ_α are user specified parameters. Not all regularization terms should necessarily be present in general. For example, if there is phenomenological support only for the sparsity of $\mathbf{A}$ then μ_B should be zero. Additionally, while the formulation above regularizes based on the vector of row averages of $\mathbf{A}$, this could alternatively be done for the vector of column averages for $\mathbf{B}$ if the matrix norm constraints were switched to $\mathbf{A}$. The regularization term for $\boldsymbol{\alpha}$ can be present in the absence of generalized matrix sparsity terms so long as the nonnegativity constraint is present for $\mathbf{A}$.

The premise of this approach requires knowledge of an upper bound for the internal factoring dimension. The tighter the upper bound, the better one expects this approach to perform. Results

on synthetic data will be discussed. A potential draw back to this method is that there are several user specified parameter values involved in determining the optimal solution.

Split Bregman has been successfully used to solve the sparse nonnegative matrix factorization problem with internal factoring dimension known (insert citation). It can also be employed to solve our amended version. Split Bregman is implemented as follows. First, we begin with our initial optimization problem.

$$\begin{aligned}
\text{minimize} \quad & \frac{1}{2}||\mathbf{AB}-\mathbf{F}||_F^2 + \mu_A|\mathbf{A}|_1 + \mu_B|\mathbf{B}|_1 + \mu_\alpha|\boldsymbol{\alpha}|_1 + \mu_\beta|\boldsymbol{\beta}|_1 \\
\text{subject to} \quad & \boldsymbol{\alpha} = \frac{1}{M}\mathbf{1}_{1xM}\mathbf{A} \\
& \mathbf{A}_{ij} \geq 0 \\
& \mathbf{B}_{ij} \geq 0 \\
& \boldsymbol{\alpha}_j \geq 0 \\
& ||\mathbf{B}_i||_2^2 = 1.
\end{aligned}$$

We then write the equivalent constrained optimization problem with the introduction of two auxiliary variables $\mathbf{a}$ and $\mathbf{b}$.

$$\begin{aligned}
\text{minimize} \quad & \frac{1}{2}||\mathbf{AB}-\mathbf{F}||_F^2 + \mu_A|\mathbf{a}|_1 + \mu_B|\mathbf{b}|_1 + \mu_\alpha|\boldsymbol{\alpha}|_1 + \mu_\beta|\boldsymbol{\beta}|_1 \\
\text{subject to} \quad & \boldsymbol{\alpha} = \frac{1}{M}\mathbf{1}_{1xM}\mathbf{A} \\
& \mathbf{a} = \mathbf{A}, \quad \mathbf{a}_{ij} \geq 0 \\
& \mathbf{b} = \mathbf{B}, \quad \mathbf{b}_{ij} \geq 0 \\
& \boldsymbol{\alpha}_j \geq 0 \quad \text{and} \quad ||\mathbf{b}_i||_2^2 = 1.
\end{aligned}$$

Next, we compute the scaled form augmented Lagrangian for this constrained problem as

$$L(\mathbf{a},\mathbf{b},\boldsymbol{\alpha},\mathbf{A},\mathbf{B},\mathbf{c}_A,\mathbf{c}_B,\mathbf{c}_\alpha) = \frac{1}{2}||\mathbf{AB}-\mathbf{F}||_F^2 + \mu_A|\mathbf{a}|_1 + \mu_B|\mathbf{b}|_1 + \mu_\alpha|\boldsymbol{\alpha}| + \frac{\lambda_A}{2}||\mathbf{a}-\mathbf{A}-\mathbf{c}_A||_F^2 + \frac{\lambda_B}{2}||\mathbf{b}-\mathbf{B}-\mathbf{c}_B||_F^2 + \frac{\lambda_\alpha}{2}||\boldsymbol{\alpha}-\frac{1}{M}\mathbf{1}_{1xM}\mathbf{A}-\mathbf{c}_\alpha||_F^2.$$

Continuing the implementation of the algorithm involves splitting variables and alternating between primal and dual variable updates. That is to say, we update a single variable at a time while keeping all others fixed. Consequently, we seek the following updates:

$$\begin{aligned}
\mathbf{a}^{n+1} &= \underset{\mathbf{a}}{\text{argmin}} \quad L(\mathbf{a},\mathbf{b}^n,\boldsymbol{\alpha}^n,\mathbf{A}^n,\mathbf{B}^n,\mathbf{c}_A^n,\mathbf{c}_B^n,\mathbf{c}_\alpha^n),\\
\mathbf{b}^{n+1} &= \underset{\mathbf{b},||\mathbf{b}_i||_2^2=1}{\text{argmin}} \quad L(\mathbf{a}^{n+1},\mathbf{b},\boldsymbol{\alpha}^n,\mathbf{A}^n,\mathbf{B}^n,\mathbf{c}_A^n,\mathbf{c}_B^n,\mathbf{c}_\alpha^n),\\
\boldsymbol{\alpha}^{n+1} &= \underset{\boldsymbol{\alpha}}{\text{argmin}} \quad L(\mathbf{a}^{n+1},\mathbf{b}^{n+1},\boldsymbol{\alpha},\mathbf{A}^n,\mathbf{B}^n,\mathbf{c}_A^n,\mathbf{c}_B^n,\mathbf{c}_\alpha^n),\\
\mathbf{A}^{n+1} &= \underset{\mathbf{A}}{\text{argmin}} \quad L(\mathbf{a}^{n+1},\mathbf{b}^{n+1},\boldsymbol{\alpha}^{n+1},\mathbf{A},\mathbf{B}^n,\mathbf{c}_A^n,\mathbf{c}_B^n,\mathbf{c}_\alpha^n),\\
\mathbf{B}^{n+1} &= \underset{\mathbf{B}}{\text{argmin}} \quad L(\mathbf{a}^{n+1},\mathbf{b}^{n+1},\boldsymbol{\alpha}^{n+1},\mathbf{A}^{n+1},\mathbf{B},\mathbf{c}_A^n,\mathbf{c}_B^n,\mathbf{c}_\alpha^n),\\
\mathbf{c}_A^{n+1} &= \mathbf{c}_A^n + \mathbf{A}^{n+1} - \mathbf{a}^{n+1},\\
\mathbf{c}_B^{n+1} &= \mathbf{c}_B^n + \mathbf{B}^{n+1} - \mathbf{b}^{n+1},\\
\mathbf{c}_\alpha^{n+1} &= \mathbf{c}_\alpha^n + \frac{1}{M}\mathbf{1}_{1xM}\mathbf{A}^{n+1} - \boldsymbol{\alpha}^{n+1}.
\end{aligned}$$

The variables, $\mathbf{c}_A$, $\mathbf{c}_B$, and $\mathbf{c}_\alpha$ are dual variables and are updated using gradient ascent. The variables $\mathbf{a}$, $\mathbf{b}$, and $\boldsymbol{\alpha}$ are updated using the shrinkage operator (together with a normalization step for the rows of $\mathbf{b}$) as follows:

$$\begin{aligned}
\mathbf{a}^{n+1} &= \text{shrinkage}(\mathbf{A}^n + \mathbf{c}_A^n - \frac{\mu_A}{\lambda_A}\mathbf{1}_{MxL}),\\
\boldsymbol{\alpha}^{n+1} &= \text{shrinkage}(\frac{1}{M}\mathbf{1}_{1xM}\mathbf{A}^n + \mathbf{c}_\alpha^n - \frac{\mu_\alpha}{\lambda_\alpha}\mathbf{1}_{Lx1}),\\
\mathbf{b}_i^{n+1} &= \text{shrinkage}\left(\frac{(\mathbf{B}^n + \mathbf{c}_B^n - \frac{\mu_B}{\lambda_B}\mathbf{1}_{LxN})_i}{||(\mathbf{B}^n + \mathbf{c}_B^n - \frac{\mu_B}{\lambda_B}\mathbf{1}_{LxN})_i||_2}\right).
\end{aligned}$$

Variable $\mathbf{B}$ has a closed form update equal to

$$\mathbf{B}^{n+1} = (((\mathbf{A}^{n+1})^T(\mathbf{A}^{n+1}) + \lambda_B \mathbf{I}_L)^{-1}((\mathbf{A}^{n+1})^T\mathbf{F} + \lambda_B(\mathbf{b}^{n+1} - \mathbf{c}_B^n)).$$

Finally, the update of variable $\mathbf{A}$ doesn't have a closed form because of the introduction of $\boldsymbol{\alpha}$ and is instead computed using gradient descent based on

$$\frac{dL}{d\mathbf{A}} = -\lambda_\alpha(\mathbf{a} - \mathbf{A} - \mathbf{c}_A) + \mathbf{A}\mathbf{B}\mathbf{B}^T - \mathbf{F}\mathbf{B}^T - \frac{\lambda_\alpha}{M}\mathbf{1}_{Mx1}(\boldsymbol{\alpha} - \frac{1}{M}\mathbf{1}_{1xM}\mathbf{A} - \mathbf{c}_\alpha).$$

This algorithmic approach to solving the proposed optimization problem is computationally quick and often requires a relatively small number of iterations for a fixed set of weighting parameter values. The one glaring drawback to the approach is the need for 6 user specified parameter values and consequently the need for a search of a six dimensional parameter space for optimal values when training data is available. Our second approach trades the need for user specification of parameters for a more computationally expensive solver. Finally, we note that the number of parameters is reduced by two, or four, when there is no sparsity expected in one, or both, of the individual matrix factors. This is to say, our proposed approach is still viable and logically sound in the case of general nonnegative matrix factorization.

Chapter 4
Reduced Rank Oblique Projections and Oblique Pseudo Inverses for Controlling Bias and Variance in Estimators

Presented here is a framework for dimension reduction by controlling the bias-variance trade in estimators for components of the *linear mixing model* (LMM). Optimally reduced rank matrices produce estimators of components of the LMM which belong to lower dimensional spaces than those produced via full rank variations. The LMM assumes that each observation can be written as a linear combination of signal and noise and is a prevalent tool used in signal processing and regression analysis [41, 42, 43, 44, 45, 46, 47, 48, 49]. Left multiplication of measurements of the LMM by particular full-rank matrices allows unbiased estimation of components of the LMM. We consider multiplication by reduced rank matrices resulting in a trade off between bias and variance.

The LMM has been used to consider detection problems in hyperspectral images [41, 42, 43, 44], linear block encoding [50, 51], polynomial regression [49], and modal analysis [52, 53, 54]. Polynomial regression and modal analysis are two other important examples of a LMM. In one case polynomials of known degree determine the mixing matrix, and in the other, complex exponential modes determine the mixing. Dimensionality reduction has been applied within the context of the LMM as a preprocessing step but not from the perspective of reducing the dimension of estimators. That is to say, dimensionality reduction has been applied to the high dimensional data that is modeled by the LMM but not to the dimension of components of the LMM. Unlike the application of dimensionality reduction to the data directly, reducing the dimension of component estimators produces strict, interpretable order determination rules dependent on parameter values that are naturally bounded for real data.

Considered herein is the LMM which includes two types of noise: structured interference and white noise. Corrupted receiver channel(s) is an example of structured interference. In many applications considered, the white noise comes from sensor and equipment noise while the structured interference comes from the setting of the experiment or something external to the sensing equipment.

The geometric setting of and background for the linear mixing model is described in Section 4.1. Section 4.3 presents our approach for producing reduced dimension estimators for components of the linear mixing model: namely the signal mode weight vector and signal component of the data. Focus is placed on the derivation of an objective function for determining estimators of optimal rank. The derived framework is applied to two applications. The first application involves the use of hyperspectral imaging for ground cover identification and is presented in Section 4.4. Also described in Section 4.4 is generalized modal analysis. Experimental designs and results are presented in Section **??**. Finally, we discuss the implications of our findings in Section **??**. Work presented here is an extension of the work [4] which had a narrowed scope focusing exclusively on the formulation of a reduced dimension matched subspace detector based on a specific reduced dimension estimator. The extended version of [4] is currently in preparation for submission.

4.1 Background

The linear mixing model can be written as

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{S}\boldsymbol{\phi} + \mathbf{n} \tag{4.1.1}$$

where the columns of $\mathbf{H}$ are a basis for the signal subspace, $\langle \mathbf{H} \rangle$, $\mathbf{H} \in \mathcal{C}^{b \times h}$, $\theta \in \mathcal{C}^h$ is the vector of signal mode weights, the columns of $\mathbf{S} \in \mathcal{C}^{b \times s}$, are a basis for the interference subspace $\langle \mathbf{S} \rangle$, $\phi \in \mathcal{C}^s$ is the vector of interference mode weights, and $\mathbf{n}$ is white noise with covariance matrix $\beta^2\mathbf{I}$. We assume that θ and ϕ are random vectors with known covariance matrices. Additionally, we assume that $\mathbf{H}$ and $\mathbf{S}$ have the trivial intersection. It is useful to discuss the geometry of the linear mixing model. The signal, $\mathbf{x} = \mathbf{H}\theta \in \mathcal{C}^b$ is the component of the data contained in the h dimensional signal subspace of $\mathcal{C}^b$ spanned by the columns of $\mathbf{H}$. The interference $\mathbf{w} = \mathbf{S}\phi \in \mathcal{C}^b$ is the data contained in the s dimensional interference subspace of $\mathcal{C}^b$ spanned by the columns of $\mathbf{S}$. Consequently, the sum $\mathbf{x} + \mathbf{w} = \mathbf{H}\theta + \mathbf{S}\phi \in \mathcal{C}^b$ is a vector in the $h + s$ dimensional subspace of $\mathcal{C}^b$ spanned by the combined column spaces of $\mathbf{H}$ and $\mathbf{S}$. The direct sum of the two spaces will be

denoted as $\langle \mathbf{H} \rangle \oplus \langle \mathbf{S} \rangle$. The orthogonal projection of $\mathbf{y}$ onto the space $\langle \mathbf{H} \rangle \oplus \langle \mathbf{S} \rangle$ is denoted $\mathbf{P}_{HS}$ where

$$\mathbf{P}_{HS} = [\mathbf{H}, \mathbf{S}][[\mathbf{H}, \mathbf{S}]^H [\mathbf{H}, \mathbf{S}]]^{-1} [\mathbf{H}, \mathbf{S}]^H$$

and $[\mathbf{H}, \mathbf{S}]$ indicates a block matrix composed of the signal and interference basis matrices. Since we are assuming that $\mathbf{x} + \mathbf{w} \in \langle \mathbf{H} \rangle \oplus \langle \mathbf{S} \rangle$ we have that $\mathbf{P}_{HS}(\mathbf{x} + \mathbf{w}) = \mathbf{x} + \mathbf{w}$.

The orthogonal projection $\mathbf{P}_{HS}$ can be decomposed as

$$\mathbf{P}_{HS} = \mathbf{P}_S^{\perp} + \mathbf{P}_G$$

where $\mathbf{P}_S^{\perp}$ is the orthogonal projection onto the orthogonal complement space of $\langle \mathbf{S} \rangle$ and $\mathbf{P}_G$ is the orthogonal projection onto $\langle \mathbf{G} \rangle$ where $\mathbf{G} = \mathbf{P}_S^{\perp} \mathbf{H}$ is the subspace of $\langle \mathbf{H} \rangle$ contained in the orthogonal complement space of $\langle \mathbf{S} \rangle$ (henceforth to be referred to as $\langle \mathbf{S}^{\perp} \rangle$). Alternatively, we can resolve the orthogonal projection $\mathbf{P}_{HS}$ as

$$\mathbf{P}_{HS} = \mathbf{E}_{H,S} + \mathbf{E}_{S,H}$$

where $\mathbf{E}_{H,S}$ is the *oblique projection* onto $\langle \mathbf{H} \rangle$ in the direction of $\langle \mathbf{S} \rangle$ and $\mathbf{E}_{SH}$ is the oblique projection onto $\langle \mathbf{S} \rangle$ in the direction of $\langle \mathbf{H} \rangle$. A diagram of the decomposition of ambient data space, as well as the connections between the different projections of an observation, is given in Figure 4.1.1.

4.2 Related Work

While understanding the geometry of the space of the LMM is informative, in practice one is likely to be interested in estimating specific components of the LMM. For example, the relative weights of the different signal dimensions (θ) are often of interest. Under the LMM, the expected coherence matrix of the observed data $\mathbf{y}$ for the assumptions $\mathbb{E}[\theta\theta^H] = \mathbf{R}_{\theta\theta}$, $\mathbb{E}[\phi\phi^H] = \mathbf{R}_{\phi\phi}$, and $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \beta^2 \mathbf{I}$ is

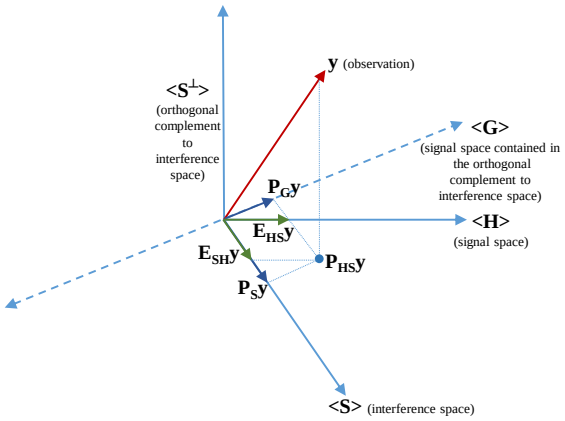


Figure 4.1.1: A diagram of the decomposition of the ambient space of data including their geometric relationships.

$$\mathbf{R}_{yy} = \mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{H}\mathbf{R}_{\theta\theta}\mathbf{H}^H + \mathbf{S}\mathbf{R}_{\phi\phi}\mathbf{S}^H + \beta^2\mathbf{I}$$

where $\mathbb{E}[*]$ indicates expectation. In general we will use the notation $\mathbf{R}_{ab} = \mathbb{E}[\boldsymbol{a}\boldsymbol{b}^H]$.

We define $\boldsymbol{X}$ to be the component of the LMM that we seek to estimate, and use $\hat{\boldsymbol{x}}$ to be an estimator of $\boldsymbol{x}$. We assume that $\hat{\boldsymbol{x}}$ can be produced by multiplying the observed data $\mathbf{y}$ by some matrix $\boldsymbol{Q}$. That is, $\hat{\boldsymbol{x}} = \boldsymbol{Q}\mathbf{y}$. In this setting we call $\boldsymbol{Q}$ a *linear estimator*. Furthermore, we define the error of this estimator to be $\hat{\boldsymbol{n}} = \boldsymbol{x} - \hat{\boldsymbol{x}}$. When one seeks to minimize the mean squared error over all possible linear estimators, the matrix $\boldsymbol{Q}$ which achieves that minimum error is called the Linear Minimum Mean-Squared Error (LMMSE) estimator. It has been shown that the LMMSE estimator for $\boldsymbol{x}$ is produced when $\boldsymbol{Q} = \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}$ [55].

For the case where $\boldsymbol{x} = \theta$ then the LMMSE estimator is $\boldsymbol{Q} = \mathbf{R}_{\theta\theta}\mathbf{H}^H\mathbf{R}_{yy}^{-1}$. The estimator for θ resulting by from left multiplication by this operator ($\hat{\theta} = \mathbf{R}_{\theta\theta}\mathbf{H}^H\mathbf{R}_{yy}^{-1}\mathbf{y}$) is an unconditionally unbiased estimator of θ. However, despite minimizing the mean squared error, the LMMSE estimator is conditionally biased. That is to say, given θ_0, $\mathbb{E}[\hat{\theta}|\theta_0] = \mathbf{R}_{\theta\theta}\mathbf{H}^H\mathbf{R}_{yy}^{-1}\mathbf{H}\theta_0 \neq \theta_0$. This may be an undesirable statistical property. An alternative estimator for θ results from left multiplication by $\mathbf{L} = \mathbf{L}_{HS} = (\mathbf{H}^H\mathbf{P}_S^{\perp}\mathbf{H})^{-1}\mathbf{H}^H\mathbf{P}_S^{\perp}$. In the literature this matrix is called the *oblique pseudo inverse*. Multiplication by $\mathbf{L}$ produces an estimator that is both unconditionally and conditionally unbiased.

This alternative estimator will produce a larger mean squared error (except in the case where the sensor noise tends to zero), but may have more desirable statistical properties depending on user preferences. Analogously, a conditionally and unconditionally unbiased estimator for $\mathbf{H}\theta$ can be produced by left multiplication by the *oblique projection* onto the signal space in the direction of the interference space given by $\mathbf{E} = \mathbf{E}_{HS} = \mathbf{H}(\mathbf{H}^H\mathbf{P}_S^{\perp}\mathbf{H})^{-1}\mathbf{H}^H\mathbf{P}_S^{\perp}$. Again, this matrix differs from the LMMSE operator but does not suffer from be conditionally biased. Additional details can be found in [48]. The framework presented in the next section uses reduced rank versions of the matrices $\mathbf{L}$ and $\mathbf{E}$ produce biased estimators in the case where the mean squared error of the estimator can be reduced by systematically introducing bias in exchange for lower variance.

Reduced rank matrices have been considered in the literature as a means to produce reduced dimension estimators. In the existing literature the question posed is of the form " for a given r how can we produce the best estimator of dimension r?" Equivalently, " for a given r how do we construct the best rank r version of our linear estimator?" To answer these questions authors try control some aspect of the error covariance matrix,

$$\mathbb{E}[\hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^H] = \mathbb{E}[(\boldsymbol{x} - \hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}})^H] = \mathbb{E}[(\boldsymbol{x} - \boldsymbol{Q}\mathbf{y})(\boldsymbol{x} - \boldsymbol{Q}\mathbf{y})^H] = \boldsymbol{Z}.$$

When a rank r, where $r < \text{rank}(\boldsymbol{Q})$, version of $\boldsymbol{Q}$ is used to produce an estimator the error covariance matrix is altered. Let the r dimensional estimator of $\boldsymbol{x}$ produced by the rank r version of $\boldsymbol{Q}$ be denoted $\hat{\boldsymbol{x}}_r$ with $\hat{\boldsymbol{x}}_r = \boldsymbol{Q}_r\mathbf{y}$. The error covariance matrix for the dimension r estimator is

$$\mathbb{E}[(\boldsymbol{x} - \boldsymbol{Q}_r\mathbf{y})(\boldsymbol{x} - \boldsymbol{Q}_r\mathbf{y})^H] = \boldsymbol{Z} + (\boldsymbol{Q} - \boldsymbol{Q}_r)\mathbf{R}_{yy}(\boldsymbol{Q} - \boldsymbol{Q}_r)^H = \boldsymbol{Z} + \boldsymbol{K}_r.$$

The approaches in [55] and [56] yield rules for producing rank r variations of $\boldsymbol{Q}$ that minimize $\text{tr}(\boldsymbol{K}_r)$ or $\det(\boldsymbol{K}_r)$, respectively. This is done by minimizing over all rank r versions of $\boldsymbol{Q}$ and in both cases the optimal rank r version is constructed by truncating the singular value decomposition of a particular matrix.

Our proposed framework differs from the existing literature in a few keys ways. First, in the works discussed they consider rank r versions of the LMMSE estimator for the component of interest. As we previously metioned, there are other linear estimators that yield larger mean squared errors but have more desirable statistical properties. In our approach we consider reduced rank versions of the oblique pseudo-inverse and oblique projection, which differ from the LMMSE estimators, to yield reduced dimension versions of θ and $\mathbf{H}\theta$. Additionally, we seek to minimize the mean squared error of the estimator as a function of the rank explicitly. Instead of finding the best rank r version of our estimator, we minimize the mean squared error over all possible ranks. This allows us to simultaneously identify the optimal rank of our estimator as well as produce order determination rules for the construction of the optimal rank matrix. Finally, unlike the exisiting methods, the reduced rank versions of our estimators can achieve lower mean squared errors than their full rank counterparts. This is not the case when you consider a reduced rank variation of the LMMSE estimator.

4.3 Reduced Dimension Estimators

Suppose that you have multi-dimensional signal and interference subspaces. Multiplying the measurement, $\mathbf{y}$, by a specific matrix allows one to estimate the signal mode weight vector θ or to estimate the signal component $\mathbf{H}\theta$, as described in the previous section. Structured rank reduction on the multiplying matrix allows one to produce a reduced dimension estimator for θ, or $\mathbf{H}\theta$, that takes into consideration the statistical properties of the estimator. The objective functions are based on minimization of the mean squared error of the reduced rank estimator. Minimization of the mean square error of the estimator yields a conditionally and unconditionally unbiased estimator whose variance can be controlled with dimension reduction by systematically introducing bias. Following are the derivations of the objective functions for optimal rank determination as well as the associated order determination rule for each component of interest.

4.3 Optimal Dimension Determination for Signal Mode Weight Vector Estimation

Begin by assuming that the observed data can be modeled by the linear mixing model of Equation 4.1.1. Consider the matrix $\mathbf{L} = \mathbf{L}_{HS} = (\mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H})^{-1} \mathbf{H}^H \mathbf{P}_S^{\perp} \in \mathcal{C}^{h \times b}$. We refer to this matrix as the *oblique pseudoinverse*. It has two useful properties which will be exploited. First, $\mathbf{LH} = \mathbf{I}$ and secondly that $\mathbf{LS} = \mathbf{0}$.

Multiplying on both sides of the linear mixing model by $\mathbf{L}$ we obtain

$$\hat{\theta} = \mathbf{Ly} = \mathbf{LH}\theta + \mathbf{LS}\phi + \mathbf{Ln} = \theta + \mathbf{Ln},$$

where $\mathbf{Ln}$ is now colored noise. Consequently, we see that $\hat{\theta}$ is an unbiased estimator of θ. The conditional error covariance matrix for this unbiased estimator is

$$\mathbf{V} = \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\boldsymbol{H}} | \boldsymbol{\theta}] = \boldsymbol{\beta^2 \mathbf{L}\mathbf{L}^H} = \boldsymbol{\beta^2} (\mathbf{H}^{\boldsymbol{H}} \mathbf{P}_{\boldsymbol{S}}^{\perp} \mathbf{H})^{-1}$$

when $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \beta^2 \mathbf{I}$. The corresponding variance is

$$V = \text{tr}[\mathbf{V}] = \beta^2 \text{tr}[\mathbf{L}\mathbf{L}^H] = \beta^2 \text{tr}[(\mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H})^{-1}]$$

for this unbiased estimator. Recall that for an unbiased estimator the mean squared error is equal to the variance. The matrix $\mathbf{L}$ may be given a thin Singular Value Decomposition (SVD), $\mathbf{L} = \mathbf{U}\Sigma\mathbf{V}^H$, $\Sigma = \text{diag}[\sigma_1, \sigma_2, ..., \sigma_h]$, $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_h$. Consequently, the corresponding unbiased estimator error covariance matrix has eigenvalue decomposition $\beta^2 \mathbf{L}\mathbf{L}^H = \beta^2 \mathbf{U}\Sigma^2 \mathbf{U}^H$, and the σ_i determine a concentration ellipse. The matrix $\mathbf{L}\mathbf{L}^H = (\mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H})^{-1}$, so the eigenvalues of $\mathbf{L}\mathbf{L}^H$ are the inverses of the eigenvalues of $(\mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H})$, which we write as

$$\begin{aligned} \sigma_i^2 &= \frac{1}{\text{ev}_{h-i+1}(\mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H})} \\ &= \frac{1}{\text{ev}_{h-i+1}(\mathbf{G}^H \mathbf{G})} \end{aligned}$$

where $\mathbf{G} = \mathbf{P}_S^{\perp}\mathbf{H}$ and $\text{ev}_k(*)$ indicates the k-th largest eigenvalue of the matrix. When the subspaces $\langle\mathbf{H}\rangle$ and $\langle\mathbf{S}\rangle$ are close, then some eigenvalues of the Grammian $\mathbf{G}^H\mathbf{G}$ will be small, some eigenvalues of $\mathbf{L}\mathbf{L}^H$ will be large, some principal axes of $\mathbf{L}\mathbf{L}^H$ will be lengthy, and the mean squared error of the unbiased estimator (the variance) will be large. Perhaps it can be reduced by introducing bias. This concept is often referred to as the bias-variance trade. It is in this problem that we find the motivation for dimension reduction. Let $\mathbf{L}_r$ be a reduced rank variant of $\mathbf{L}$. Multiplying the measurement by $\mathbf{L}_r$, instead of $\mathbf{L}$ we produce a reduced dimension estimator, $\hat{\theta}_r$, defined as

$$\hat{\theta}_r = \mathbf{L}_r\mathbf{y} = \mathbf{L}_r\mathbf{H}\theta + \mathbf{L}_r\mathbf{S}\phi + \mathbf{L}_r\mathbf{n}.$$

The estimator $\hat{\theta}_r$ is no longer unbiased, and the variance of $\hat{\theta}_r$ is no longer V. This motivates the question, how do we determine an optimal r?

We propose to determine an optimal r by minimizing the mean square error (MSE) of the estimator $\hat{\theta}_r$. Recall that $MSE(\hat{\theta}_r) = \mathbb{E}[||\hat{\theta}_r - \theta||_2^2]$ where $\mathbb{E}[*]$ is expectation. First, define the mean square error as a function of our reduced dimension estimator by $g(r) = MSE(\hat{\theta}_r)$:

$$\begin{aligned} g(r) = tr(\mathbb{E}[&((\mathbf{L}_r - \mathbf{L})\mathbf{H}\theta + \mathbf{L}_r\mathbf{S}\phi + \mathbf{L}_r\mathbf{n}) \\ &\times((\mathbf{L}_r - \mathbf{L})\mathbf{H}\theta + \mathbf{L}_r\mathbf{S}\phi + \mathbf{L}_r\mathbf{n})^H]). \end{aligned}$$

Using the linearity of expectation, and making the assumption that the sensor noise is independent of both the signal and the interference (causing all terms containing a product of the sensor noise and signal or interference to be zero), we can write

$$\begin{aligned} g(r) = \text{tr}(\mathbf{L}_r - \mathbf{L})\mathbf{H}\mathbb{E}[\theta\theta^H]\mathbf{H}^H(\mathbf{L}_r - \mathbf{L})^H + \\ (\mathbf{L}_r - \mathbf{L})\mathbf{H}\mathbb{E}[\theta\phi^H]\mathbf{S}^H\mathbf{L}_r^H + \mathbf{L}_r\mathbf{S}\mathbb{E}[\phi\theta^H]\mathbf{H}^H(\mathbf{L}_r - \mathbf{L})^H \\ + \mathbf{L}_r\mathbf{S}\mathbb{E}[\phi\phi^H]\mathbf{S}^H\mathbf{L}_r^H + \mathbb{E}[\mathbf{L}_r\mathbf{n}\mathbf{n}^H\mathbf{L}_r^H]). \end{aligned}$$

The random variables are θ, ϕ, and $\mathbf{n}$. Furthermore, assume that $\mathbb{E}[\theta\theta^H] = \lambda_\theta^2\mathbf{I}$, $\mathbb{E}[\phi\phi^H] = \lambda_\phi^2\mathbf{I}$, $\mathbb{E}[\phi\theta^H] = \mathbf{0}$, and $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \beta^2\mathbf{I}$. Let $\mathbf{L} = \mathbf{U}\Sigma\mathbf{V}^H = \sum_{1\leq i\leq p}\mathbf{u}_i\sigma_i\mathbf{v}_i^H$ be the thin singular value decomposition of $\mathbf{L}$, which is a symmetric square root of $\mathbf{L}\mathbf{L}^H$. Our reduced rank version of $\mathbf{L}$ is based on a specific subset of the singular values. Let $\mathbf{I}_r$ be the indexing set for the r singular values we will keep, $\bar{\mathbf{I}}_r$ be the indexing set for the $h-r$ (where h is the rank of $\mathbf{L}$) singular values we discard. This allows us to write $\mathbf{L}_r = \mathbf{U}\Sigma_r\mathbf{V}^H = \sum_{\mathbf{I}_r}\mathbf{u}_i\sigma_i\mathbf{v}_i^H$. Furthermore, $\mathbf{L}_r - \mathbf{L} = -\mathbf{U}\Sigma_{\bar{r}}\mathbf{V}^H = -\sum_{\bar{\mathbf{I}}_r}\mathbf{u}_i\sigma_i\mathbf{v}_i^H$. Rewriting $g(r)$, we obtain

$$\begin{aligned}
g(r) &= \text{tr}(\lambda_\theta^2(\mathbf{L}_r-\mathbf{L})\mathbf{H}\mathbf{H}^H(\mathbf{L}_r-\mathbf{L})^H + \lambda_\phi^2\mathbf{L}_r\mathbf{S}\mathbf{S}^H\mathbf{L}_r^H + \beta^2\mathbf{L}_r\mathbf{L}_r^H) \\
&= \lambda_\theta^2\text{tr}(\mathbf{U}\Sigma_{\bar{r}}\mathbf{V}^H\mathbf{H}\mathbf{H}^H\mathbf{V}\Sigma_{\bar{r}}\mathbf{U}^H) + \lambda_\phi^2\text{tr}(\mathbf{U}\Sigma_r\mathbf{V}^H\mathbf{S}\mathbf{S}^H\mathbf{V}\Sigma_r\mathbf{U}^H) \\
&\quad + \beta^2\text{tr}(\mathbf{U}\Sigma_r\mathbf{V}^H\mathbf{V}\Sigma_r\mathbf{U}^H) \\
&= \lambda_\theta^2\text{tr}(\Sigma_{\bar{r}}\mathbf{V}^H\mathbf{H}\mathbf{H}^H\mathbf{V}\Sigma_{\bar{r}}) + \lambda_\phi^2\text{tr}(\Sigma_r\mathbf{V}^H\mathbf{S}\mathbf{S}^H\mathbf{V}\Sigma_r) + \beta^2\text{tr}(\Sigma_r\Sigma_r).
\end{aligned}$$

In terms of the indexing sets we have

$$\begin{aligned}
g(r) = \text{tr}(&\sum_{\bar{\mathbf{I}}_r}\sum_{\bar{\mathbf{I}}_r}\mathbf{u}_i\sigma_i\mathbf{v}_i^H\mathbf{H}\lambda_\theta^2\mathbf{I}\mathbf{H}^H\mathbf{v}_j\sigma_j\mathbf{u}_j^H \\
&+ \sum_{\mathbf{I}_r}\sum_{\mathbf{I}_r}\mathbf{u}_i\sigma_i\mathbf{v}_i^H\mathbf{S}\lambda_\phi^2\mathbf{I}\mathbf{S}^H\mathbf{v}_j\sigma_j\mathbf{u}_j^H \\
&+ \beta^2\sum_{\mathbf{I}_r}\sum_{\mathbf{I}_r}\mathbf{u}_i\sigma_i\mathbf{v}_i^H\mathbf{v}_j^H\sigma_j\mathbf{u}_j^H).
\end{aligned}$$

Equivalently, using the linearity and invariance to cyclic permutations of the trace operation together with the orthonormality of $\mathbf{U}$ and $\mathbf{V}$ (allowing us to write in terms of single sums not double sums),

$$g(r) = \lambda_\theta^2\sum_{\bar{\mathbf{I}}_r}\sigma_i^2\mathbf{v}_i^H\mathbf{H}\mathbf{H}^H\mathbf{v}_i \tag{4.3.1}$$

$$+ \lambda_\phi^2\sum_{\mathbf{I}_r}\sigma_i^2\mathbf{v}_i^H\mathbf{S}\mathbf{S}^H\mathbf{v}_i + \beta^2\sum_{\mathbf{I}_r}\sigma_i^2, \tag{4.3.2}$$

where v_i denotes the i^{th} column of $\mathbf{V}$.

In the current form, the objective function appears to be a function of the three parameters λ_θ^2, λ_ϕ^2, and β^2. However, it can be written as a function with a single parameter once some observations have been made. Firstly, we recall that $\mathbf{LH} = \mathbf{I}$. Using the SVD of $\mathbf{L}$ (as we have previously defined it) we have

$$\begin{aligned}\mathbf{I}_h =& \mathbf{LHH}^H\mathbf{L}^H = \mathbf{U}\Sigma\mathbf{V}^H\mathbf{HH}^H\mathbf{V}\Sigma^H\mathbf{U}^H \\ \implies & \sigma_i^2\mathbf{v}_i^H\mathbf{HH}^H\mathbf{v}_i^H = 1.\end{aligned}$$

Similarly, using $\mathbf{LS} = \mathbf{0}$ and the SVD of $\mathbf{L}$ we have

$$\begin{aligned}\mathbf{0}_h =& \mathbf{LSS}^H\mathbf{L}^H = \mathbf{U}\Sigma\mathbf{V}^H\mathbf{SS}^H\mathbf{V}\Sigma^H\mathbf{U}^H \\ \implies & \sigma_i^2\mathbf{v}_i^H\mathbf{SS}^H\mathbf{v}_i^H = 0.\end{aligned}$$

Consequently, we can rewrite Equation 4.3.1 as

$$g(r) = \lambda_\theta^2 \sum_{\bar{\mathbf{I}}_r} 1 + \beta^2 \sum_{\mathbf{I}_r} \sigma_i^2 = \lambda_\theta^2 |\bar{\mathbf{I}}_r| + \beta^2 \sum_{\mathbf{I}_r} \sigma_i^2 \tag{4.3.3}$$

where σ_i^2 is the ith component of noise gain an $|\bar{\mathbf{I}}_r|$ is the cardinality of the indexing set for the dimensions to be discarded. Here the first term in the objective function is the only remaining term measuring the bias of the reduced dimension estimator while the second measures the variance of the estimator. The λ_θ^2 in Equation 4.3.3 came from the assumption that $\mathbb{E}[\theta\theta^H] = \lambda_\theta^2\mathbf{I}$. If this assumption is changed to $\mathbb{E}[\theta\theta^H] = \mathbf{\Lambda} = \text{diag}[\Lambda_1, ..., \Lambda_h]$, then the objective function is instead

$$g(r) = \sum_{\bar{\mathbf{I}}_r} \Lambda_i^2 + \beta^2 \sum_{\mathbf{I}_r} \sigma_i^2, \tag{4.3.4}$$

where Λ_i^2 is the signal power in the ith signal mode. To exclude mode i is to sustain bias squared Λ_i, but to retain it is to pay with variance $\beta^2\sigma_i$. So a mode is retained when the signal power

is greater than the noise power. For further intuition, notice again that the eigenvalues of $\mathbf{L}\mathbf{L}^H$ correspond to the major axes of the error covariance matrix of the full rank matrix, so, the order fitting rule is to retain dimension i when

$$\Lambda_i^2 > \frac{\beta^2}{ev_{h-i+1}(\mathbf{H}^H\mathbf{P}_S^{\perp}\mathbf{H})}.$$

This inequality results directly from the indexing sets for the two terms in $g(r)$. An index is kept when $\beta^2\sigma_i^2$ increases $g(r)$ less than the corresponding Λ_i^2.

Using the form of the objective function in Equation 4.3.3 a counter intuitive ordering rule is revealed: the 'optimal' rank one solution (when λ_θ^2 is fixed) involves keeping only the last singular vector of the oblique pseudo inverse, i.e. the direction capturing minimal, not maximal, variance. This fact was first observed empirically based on solving the combinatorial optimization problem and seeing an inherited structure between optimal indexing sets at each rank. Additionally, this is intuitively supported when interpreting this as keeping the error direction of minimal magnitude.

4.3 Optimal Dimension Determination for Signal Component Estimation

We begin by making assumptions consistent with the previous section. However, we now consider the matrix

$$\mathbf{E} = \mathbf{E}_{HS} = \mathbf{H}(\mathbf{H}^H\mathbf{P}_S^{\perp}\mathbf{H})^{-1}\mathbf{H}^H\mathbf{P}_S^{\perp},$$

the oblique projection onto $\langle\mathbf{H}\rangle$ in the direction of $\langle\mathbf{S}\rangle$. Two useful properties of $\mathbf{E}$ are that $\mathbf{E}\mathbf{H} = \mathbf{H}$ and $\mathbf{E}\mathbf{S} = \mathbf{0}$.

Multiplying on both sides of the linear mixing model by $\mathbf{E}$ we obtain

$$\begin{aligned}\widehat{\mathbf{H}\theta} = \mathbf{E}\mathbf{y} &= \mathbf{E}\mathbf{H}\theta + \mathbf{E}\mathbf{S}\phi + \mathbf{E}\mathbf{n} \\ &= \mathbf{H}\theta + \mathbf{E}\mathbf{n},\end{aligned}$$

where $\mathbf{En}$ is now colored noise. Consequently, we see that $\widehat{\mathbf{H}\theta}$ is an unbiased estimator of $\mathbf{H}\theta$, the signal component of the data. The conditional mean-squared error of this estimator is the variance $V = \beta^2 \text{tr}[\mathbf{E}\mathbf{E}^H]$. As in the case of estimating the signal mode weight vector, we seek to balance the bias and variance to determine an optimal reduced dimension estimate. Let $\mathbf{E}_r$ be a reduced rank variant of $\mathbf{E}$. Multiplying the linear mixing model by $\mathbf{E}_r$, instead of $\mathbf{E}$ we produce a reduced dimension estimator, $\widehat{\mathbf{H}\theta}_r$, defined as

$$\widehat{\mathbf{H}\theta}_r = \mathbf{E}_r\mathbf{y} = \mathbf{E}_r\mathbf{H}\theta + \mathbf{E}_r\mathbf{S}\phi + \mathbf{E}_r\mathbf{n}.$$

The MSE of the estimator $\widehat{\mathbf{H}\theta}_r$, that we seek to minimize, is written as $z(r) = MSE(\widehat{\mathbf{H}\theta}_r) = \mathbb{E}[||\widehat{\mathbf{H}\theta}_r - \mathbf{H}\theta||_2^2]$. Expanding, we write

$$\begin{aligned} z(r) = \text{tr}(\mathbb{E}[((\mathbf{E}_r - \mathbf{E})\mathbf{H}\theta + \mathbf{E}_r\mathbf{S}\phi + \mathbf{E}_r\mathbf{n}) \\ \times ((\mathbf{E}_r - \mathbf{E})\mathbf{H}\theta + \mathbf{E}_r\mathbf{S}\phi + \mathbf{E}_r\mathbf{n})^H]). \end{aligned}$$

As in Section 4.3.1 we expand using the distribution properties of the Hermitian transpose, linearity of expectation, and make the assumption that the sensor noise is independent of both the signal and the interference. Now consider the thin singular value decomposition of $\mathbf{E}$. In order to simplify notation we use $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}$ again for the singular value decomposition. For the remainder of this derivation we write $\mathbf{E} = \mathbf{U}\Sigma\mathbf{V}^H$. Similar to Section 4.3.1, the reduced rank version of $\mathbf{E}$ is based on a specific subset of the singular values/vectors. This allows us to write $\mathbf{E}_r = \mathbf{U}\Sigma_r\mathbf{V}^H$. Again we have, $\mathbf{E}_r - \mathbf{E} = -\mathbf{U}\Sigma_{\bar{r}}\mathbf{V}^H$ where the $\mathbf{I}_r$ and $\bar{\mathbf{I}}_r$ indicate the dimensions being retained and discarded, respectively.

Using again the linearity and invariance to cyclic permutations of the trace operation together with the orthonormality of $\mathbf{U}$ and $\mathbf{V}$, we arrive at

$$z(r) = \lambda_\theta^2 \sum_{\bar{\mathbf{I}}_r} \sigma_i^2 \mathbf{v}_i^H \mathbf{H}\mathbf{H}^H \mathbf{v}_i + \lambda_\phi^2 \sum_{\mathbf{I}_r} \sigma_i^2 \mathbf{v}_i^H \mathbf{S}\mathbf{S}^H \mathbf{v}_i + \beta^2 \sum_{\mathbf{I}_r} \sigma_i^2, \quad (4.3.5)$$

where $\mathbf{I}_r$ is the indexing set for the r singular values we will keep, $\bar{\mathbf{I}}_r$ is the indexing set for the $h - r$ (where h is the rank of $\mathbf{E}$) singular values we discard, and v_i denotes the i^{th} column of $\mathbf{V}$.

Further simplifications can be made using two properties of $\mathbf{E}$. First, using $\mathbf{ES} = \mathbf{0}$ and the SVD of $\mathbf{E}$ we have

$$\begin{aligned} \mathbf{0}_b &= \mathbf{LSS}^H\mathbf{L}^H = \mathbf{U}\Sigma\mathbf{V}^H\mathbf{SS}^H\mathbf{V}\Sigma^H\mathbf{U}^H \\ &\implies \sigma_i^2\mathbf{v}_i^H\mathbf{SS}^H\mathbf{v}_i^H = 0. \end{aligned}$$

Second, recall that $\mathbf{EH} = \mathbf{H}$. Using the SVD of $\mathbf{E}$ we also have

$$\begin{aligned} \mathbf{HH}^H =& \mathbf{EHH}^H\mathbf{E}^H = \mathbf{U}\Sigma\mathbf{V}^H\mathbf{HH}^H\mathbf{V}\Sigma^H\mathbf{U}^H \\ \mathbf{U}^H\mathbf{HH}^H\mathbf{U} =& \Sigma\mathbf{V}^H\mathbf{HH}^H\mathbf{V}\Sigma^H \end{aligned}$$

Consequently, we can rewrite Equation 4.3.5 as

$$z(r) = \lambda_\theta^2 \sum_{\bar{\mathbf{I}}_r} \mathbf{u}_i^H\mathbf{HH}^H\mathbf{u}_i + \beta^2 \sum_{\mathbf{I}_r} \sigma_i^2. \tag{4.3.6}$$

Relaxing the assumption that $\mathbb{E}[\theta\theta^H] = \lambda_\theta^2\mathbf{I}$ to $\mathbb{E}[\theta\theta^H] = \mathbf{\Lambda}$, with $\mathbf{\Lambda}$ a diagonal matrix, we have

$$z(r) = \sum_{\bar{\mathbf{I}}_r} \Lambda_i^2\mathbf{u}_i^H\mathbf{HH}^H\mathbf{u}_i + \beta^2 \sum_{\mathbf{I}_r} \sigma_i^2, \tag{4.3.7}$$

where Λ_i^2 is the signal power in the ith channel. Therefore, the order determination rule is that a dimension is retained when

$$\Lambda_i\mathbf{u}_i^H\mathbf{HH}^H\mathbf{u}_i > \beta^2\sigma_i^2.$$

When the power of the signal in the i-th mode is larger than the noise gain, a dimension is kept. An interesting observation can further be made about this order determination rule. Consider the principal angles β between the subspaces $\langle\mathbf{H}\rangle$ and $\langle\mathbf{S}\rangle$. It was shown in [48] that the singular values

of $\mathbf{E}$ are equal to the the inverse of the sine of the principal angles;

$$\sigma_i = \frac{1}{\sin \beta_i}$$

where β_i is the i-th principal angle and σ_i is the i-th singular value of $\mathbf{E}$. Consequently, the order determination rule can be equivalently written as

$$\Lambda_i \mathbf{u}_i^H \mathbf{H}\mathbf{H}^H \mathbf{u}_i > \frac{\beta^2}{\sin^2 \beta_i}. \tag{4.3.8}$$

When the β_i tends towards $\pi/2$ this amounts to a comparison between the signal power in the i-th left singular vector and the white noise power only, not the noise gain.

4.4 Applications

We consider two applications: generalized modal analysis and matched subspace detection in hyperspectral imaging. The application in generalized modal analysis describes a fully synthetic data set in which all needed parameters are known and consequently the benefit of performing dimension reduction can be truly quantitatively evaluated. Ground cover identification from hyperspectral imaging provides a real world problem in which the true parameter values needed for optimization are unknown and are instead chosen in an ad hoc fashion. The results of applying the presented framework to both applications yields marked improvment.

4.4 Generalized Modal Analysis

In order to quantitatively demonstrate the benefit of our framework we explore the application of the LMM to modal analysis. One can consider the signal space discussed in previous sections as a set of basis vectors/ modes of a decomposition. There are many ways in which one can decompose observed data. Differences between these decompositions largely depends on the assumptions placed on $\mathbf{H}$ in the LMM. For example, if the columns of $\mathbf{H}$ are generated by rationally related fre-

quencies (harmonics) we are in the setting of a discrete Fourier decomposition. Assuming that the rows of $\mathbf{H}$ correspond to monomials (up to a fixed degree) corresponds to polynomial regression.

For experiments presented, we assume the signal mixing matrix is known and is a Vandermonde matrix structured from a set of random complex generators chosen from the complex unit circle. Explicitly, $\mathbf{H} = [\mathbf{h}_1, ..., \mathbf{h}_M] \in \mathcal{C}^{N \times M}$ where $\mathbf{h}_m = [1\, e^{i\theta_m}\, e^{(2)i\theta_m} \ldots e^{(N-1)i\theta_m}]$ and θ_m is the mth mode frequency for mode $\mathbf{h}_m$. Additionally, we consider an interference space that is "close" to the signal space. Often when the signal and interference spaces are close (small principal angles) the performance of many algorithms falls off. A basis for the interference is assumed to be known while the amplitudes of the interference modes are unknown. These assumptions allow us to create a synthetic data set with the ability to exactly construct $\mathbf{E}$ and $\mathbf{L}$ and evaluate the benefit of a reduced rank approach. We further note that the presented framework for dimension reduction in this generalized modal analysis framework differs from standard techniques in that it accounts for the presence of both structured interference and white noise in the data we seek to decompose.

4.4 Matched Subspace Detection: Hyperspectral Ground Cover Detection

The linear mixing model is a widely used tool in applied hyperspectral signal detection. It assumes that each pixel can be written as a linear combination of signal and noise [41, 42, 43, 44]. A standard model for the noise contribution consists of a combination of structured noise/interference coming from characteristics of the imaging environment in addition to unstructured/random noise attributable to the imaging equipment and system. We address a geometrically motivated variation of the linear mixing model for the detection of chemicals in hyperspectral images using an approximate oblique pseudo-inverse technique.

We propose a framework for dimension reduction in matched subspace detection [45] based on reduced dimension estimators of the signal mode weight vector presented above. Consider the signal detection problem in hyperspectral imaging. In this setting, each data point corresponds to a single pixel consisting of values from b hyperspectral bands, i.e., $\mathbf{y} \in \mathcal{C}^b$. The goal is to determine an appropriate statistical model of the observed data. The null hypothesis models the data as a

linear combination of structured noise and white noise. The alternative hypothesis models the data as a linear combination of signal, structured noise, and noise. Detection of the signal component in an individual pixel is based on a hypothesis test of these null and alternative hypotheses. For the derivation of the generalized likelihood ratio test comparing these two hypotheses, we refer the reader to [46, 47, 57].

To better understand the form of the generalized likelihood ratio, it is useful to again recall the geometry of the linear mixing model. We assume $\mathbf{x} = \mathbf{H}\theta \in \mathcal{C}^b$ is the portion of the data contained in the h-dimensional subspace of $\mathcal{C}^b$ spanned by the columns of $\mathbf{H}$. We will denote this subspace by $\langle \mathbf{H} \rangle$. Similarly, $\mathbf{w} = \mathbf{S}\phi \in \mathcal{C}^b$ is the portion of our data contained in the s-dimensional subspace of $\mathcal{C}^b$ spanned by the columns of $\mathbf{S}$ denoted by $\langle \mathbf{S} \rangle$. Consequently, the sum $\mathbf{x} + \mathbf{w} = \mathbf{H}\theta + \mathbf{S}\phi \in \mathcal{C}^b$ is a vector in the $h + s$ dimensional subspace of $\mathcal{C}^b$ spanned by the combined column spaces of $\mathbf{H}$ and $\mathbf{S}$. This direct sum of the two spaces will be denoted as $\langle \mathbf{HS} \rangle$. The orthogonal projection of $\mathbf{y}$ onto the subspace $\langle \mathbf{H}, \mathbf{S} \rangle$ is denoted $\mathbf{P}_{HS}$ where

$$\mathbf{P}_{HS} = [\mathbf{H}, \mathbf{S}][[\mathbf{H}, \mathbf{S}]^H [\mathbf{H}, \mathbf{S}]]^{-1} [\mathbf{H}, \mathbf{S}]^H.$$

Since we are assuming that $\mathbf{x} + \mathbf{w} \in \langle \mathbf{HS} \rangle$ we have that $\mathbf{P}_{HS}(\mathbf{x} + \mathbf{w}) = \mathbf{x} + \mathbf{w}$.

The orthogonal projection $\mathbf{P}_{HS}$ can be decomposed as

$$\mathbf{P}_{HS} = \mathbf{P}_S^{\perp} + \mathbf{P}_G$$

where $\mathbf{P}_S^{\perp}$ is the orthogonal projection onto the orthogonal complement space of $\langle \mathbf{S} \rangle$ and $\mathbf{P}_G$ is the orthogonal projection onto $\langle \mathbf{G} \rangle$ where $\mathbf{G} = \mathbf{P}_S^{\perp}\mathbf{H}$ is the subspace of $\langle \mathbf{H} \rangle$ contained in the orthogonal complement space of $\langle \mathbf{S} \rangle$ (henceforth to be referred to as $\langle \mathbf{S}^{\perp} \rangle$).

Using this decomposition, and other projection identities found in [45, 47], the likelihood ratio can be written as

$$L(\mathbf{y}) = \frac{\mathbf{y}^H \mathbf{P}_S^{\perp} \mathbf{P}_G \mathbf{P}_S^{\perp} \mathbf{y}}{\mathbf{y}^H \mathbf{P}_S^{\perp} \mathbf{y}} = \frac{\mathbf{y}^H \mathbf{L}^H (\mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H}) \mathbf{L} \mathbf{y}}{\mathbf{y}^H \mathbf{P}_S^{\perp} \mathbf{y}} \tag{4.4.1}$$

$$= \frac{\hat{\theta}^H \mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H} \hat{\theta}}{\mathbf{y}^H \mathbf{P}_S^{\perp} \mathbf{y}} \tag{4.4.2}$$

where $\mathbf{L} = \mathbf{L}_{HS} = (\mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H})^{-1} \mathbf{H}^H \mathbf{P}_S^{\perp}$. and $\hat{\theta} = \mathbf{L}\mathbf{y}$. This form of likelihood measure will be manipulated in the course of the research presented here. The value of this likelihood ratio is used as a detection score as shown in [57] and measures the relative coherence of the observed data to a space orthogonal to the interference. In the scenario where noise properties are adaptively determined from the data, the detector is commonly referred to as ACE (Adaptive Coherence Estimator) and has been shown to have many desirable properties [46, 47].

Recall the score of Equation 4.4.1:

$$\begin{aligned} L_r(\mathbf{y}) &= \frac{\mathbf{y}^H \mathbf{P}_S^{\perp} \mathbf{P}_G \mathbf{P}_S^{\perp} \mathbf{y}}{\mathbf{y}^H \mathbf{P}_S^{\perp} \mathbf{y}} \\ &= \frac{\mathbf{y}^H \mathbf{P}_S^{\perp} (\mathbf{P}_S^{\perp} \mathbf{H} (\mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{P}_S^{\perp} \mathbf{H})^{-1} \mathbf{H}^H \mathbf{P}_S^{\perp}) \mathbf{P}_S^{\perp} \mathbf{y}}{\mathbf{y}^H \mathbf{P}_S^{\perp} \mathbf{y}} \\ &= \frac{\mathbf{y}^H \mathbf{L}^H (\mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H}) \mathbf{L} \mathbf{y}}{\mathbf{y}^H \mathbf{P}_S^{\perp} \mathbf{y}} \\ &= \frac{\hat{\theta}^H (\mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H}) \hat{\theta}}{\mathbf{y}^H \mathbf{P}_S^{\perp} \mathbf{y}}. \end{aligned}$$

This series of equations shows that our original detection score can be written in terms of the estimated signal mode weights $\hat{\theta}$. Since it is this value that we are estimating by multiplying through by the reduced rank $\mathbf{L}$, it is logical to make the score depend on our reduced dimension estimate by replacing $\hat{\theta}$ with $\hat{\theta}_r$. The resulting score given by

$$L_r(\mathbf{y}) = \frac{\hat{\theta}_r^H \mathbf{H}^H \mathbf{P}_S^{\perp} \mathbf{H} \hat{\theta}_r}{\mathbf{y}^H \mathbf{P}_S^{\perp} \mathbf{y}} \tag{4.4.3}$$

will be referred to as ACE-Theta.

The final piece needed to compute the detection score is the matrix $\mathbf{P}_S^{\perp}$. If it is assumed that the sampled covariance matrix, $\mathbf{W}$, is full rank and can be written as

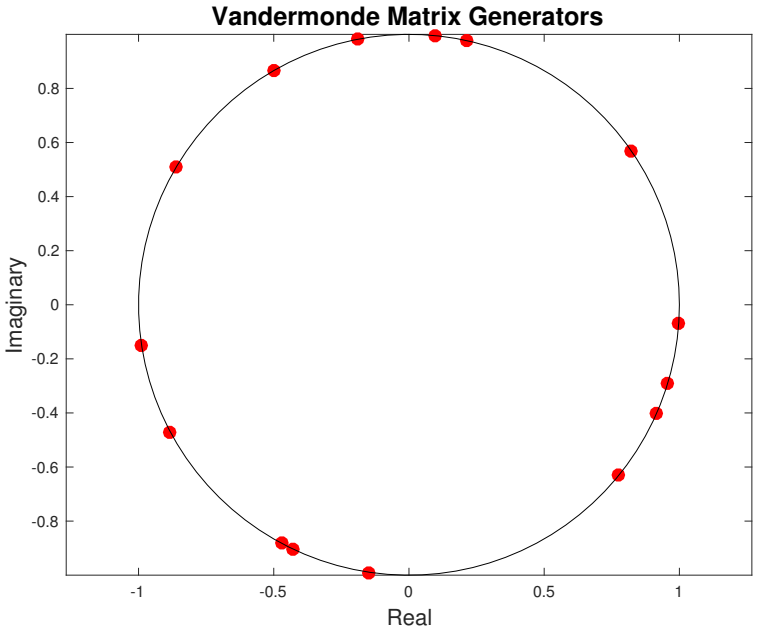


Figure 4.5.1: This figure shows the mode frequencies for the modes of the signal space (i.e. generators for the Vandermonde signal mixing matrix).

$$\mathbf{W} = \mathbf{S}\alpha_{\phi}^{2}\mathbf{I}\mathbf{S}^{H} + \beta^{2}\mathbf{I},$$

then the Woodbury matrix identity [58] provides a rationale for using the inverse of the covariance matrix as an estimate for $\mathbf{P}_{S}^{\perp}$ [48, 59].

4.5 Experiments and Results

4.5 Generalized Modal Analysis

The signal basis matrix is modeled as a Vandermonde matrix formed by randomly chosen complex generators all on the unit circle. We provide a single, but representative, exemplar of a signal space which we pair with a structured interference space that is constructed such that the largest principal angle between $\mathbf{H}$ and $\mathbf{S}$ is $\pi/10$. The fifteen generators for the exemplar signal subspace are shown in Figure 4.5.1. We have chosen the ambient space to be a $\mathcal{C}^{40}$ resulting in $\mathbf{H} \in \mathcal{C}^{40\times 15}$. For the structured interference which is close to the signal space, we have made the interference space 10 dimensional ($\mathbf{S} \in \mathcal{C}^{40\times 10}$).

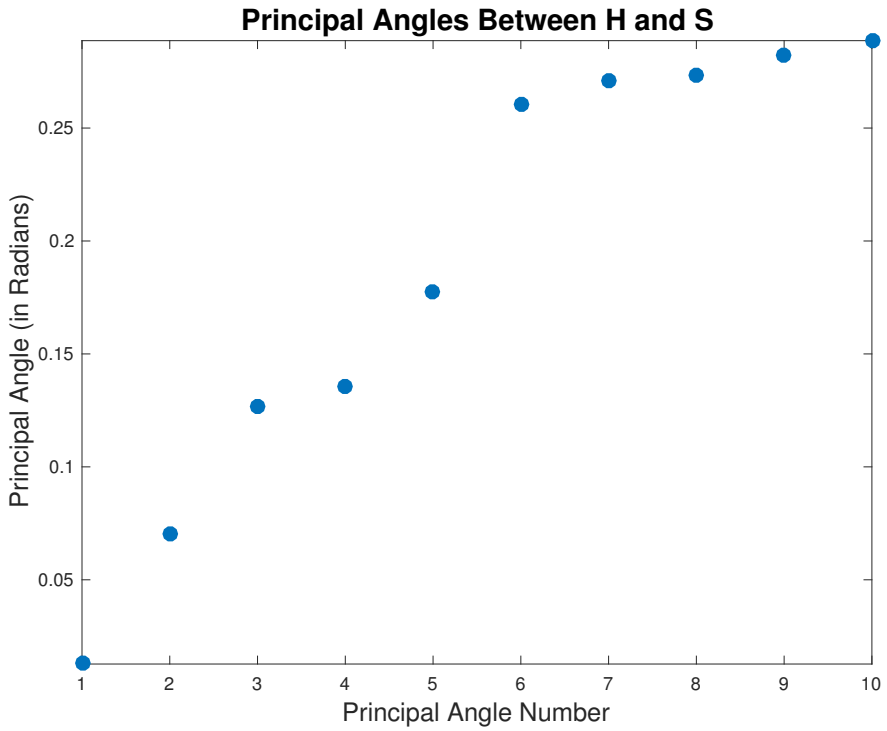


Figure 4.5.2: This figure shows the principal angles (in radians) between the signal space $\langle \mathbf{H} \rangle$ and the interference space $\langle \mathbf{S} \rangle$.

For the provided exemplar pair ($\mathbf{H}$ and $\mathbf{S}$), one thousand random realizations of the signal mode weights ($\theta \in \mathcal{C}^{15\times 1}$) and the interference amplitudes ($\phi \in \mathcal{C}^{10\times 1}$) are generated for each noise level. Thus, an ensemble of realizations is produced for each level of white noise (the noise vectors are randomly chosen once and then scaled by different values to produce different input signal to noise ratios (SNRs)). Reconstruction error, measured as the normalized mean squared error, is averaged over the 1000 realizations comprising the ensemble at a each noise level. Normalized mean-squared error is computed as

$$MSE(\widehat{\mathbf{H}\theta}) = \frac{||\widehat{\mathbf{H}\theta} - \mathbf{H}\theta||_2^2}{||\mathbf{H}\theta||_2^2}$$

and are plotted against $10\log_{10} SNR$. Similarly, we consider the normalized mean-squared error for the signal mode weight vector θ. The principal angles between the signal and interference spaces for the experiment are shown in Figure 4.5.2. In the experiment $\lambda_\theta^2 = 1$ and β^2 (in

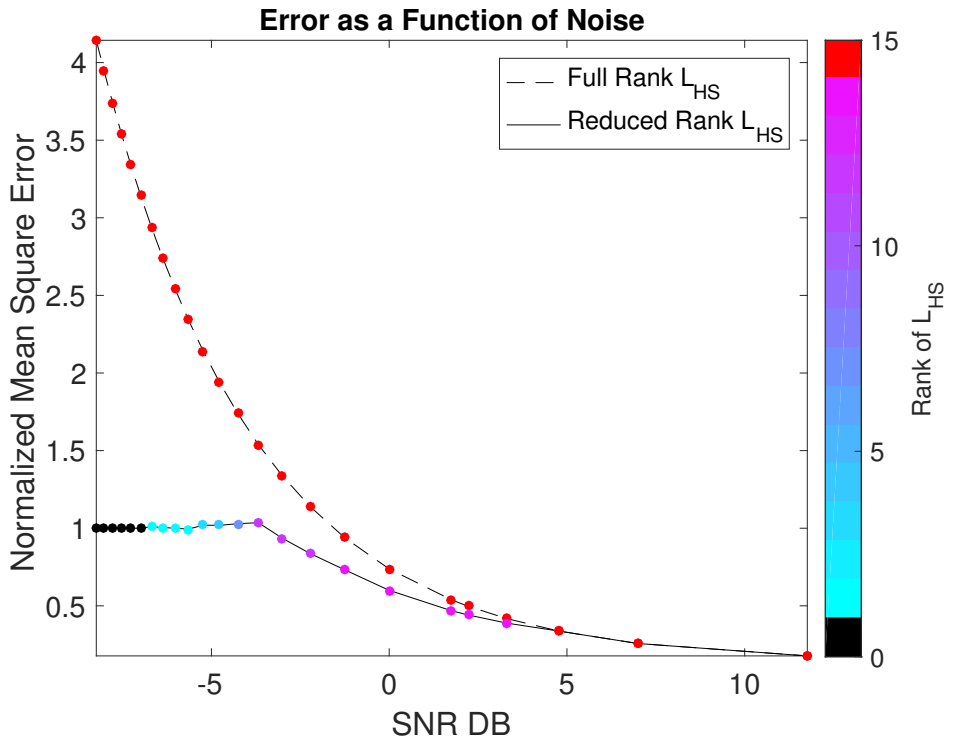


Figure 4.5.3: This figure shows the reconstruction error for the signal mode weight vector as a function of the input signal to noise ratio (SNR) in DB.

Equations 4.3.4 and 4.3.7) are known and these values are used together with our presented order determination rules shown in Section 4.3.1.

Notice that for estimation of both the signal mode weights and the signal component there are input SNRs where interesting rank reductions are chosen. Below a certain SNR the optimal matrix is the rank zero matrix populated with all zeros and above a certain SNR full rank becomes optimal. When considering Figure 4.5.4 it is useful to think about Equation 4.3.8. Recall, the order determination rule for the signal component compares the signal power in a given mode to the white noise variance scaled by a function of the principal angles that increase as the angles get smaller. Thus, when we see that dimensions get thrown out for lower SNRs it is due to the fact that the signal power is lower than the scaled white noise variance.

As the white noise variance gets smaller (larger SNR), and the principal angles are near orthogonal, the condition for a dimension to be retained is that the signal power exceeds the white noise variance is more easily satisfied than when the principal angles are small.

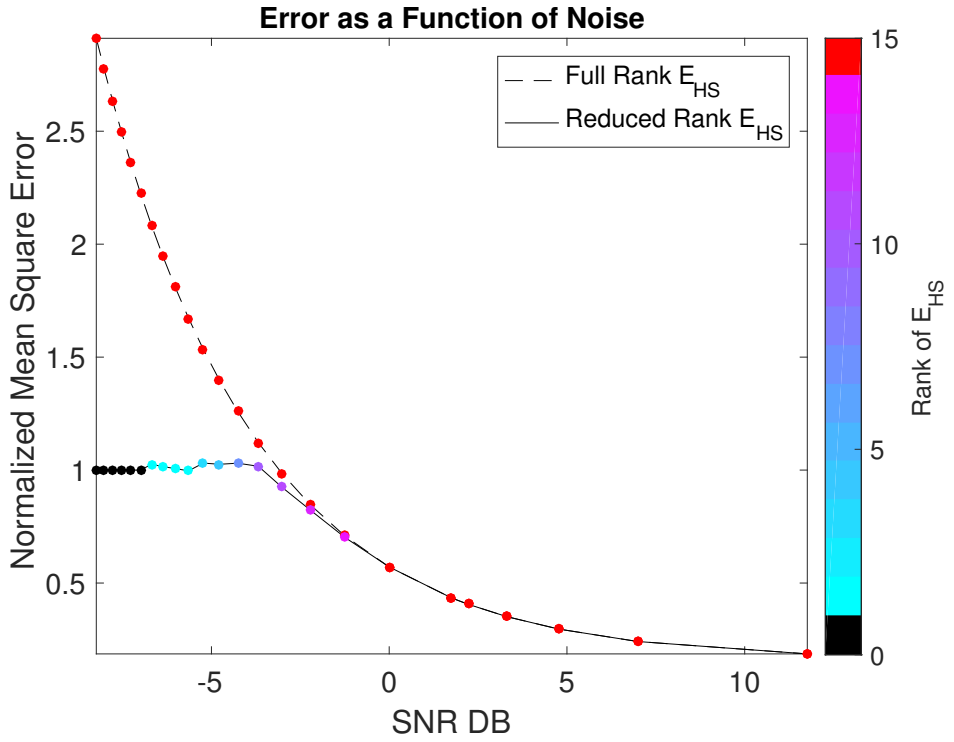


Figure 4.5.4: This figure shows the reconstruction error for the signal component as a function of the signal to noise ratio (SNR).

4.5 Matched Subspace Detection: Hyperspectral Ground Cover Detection

In order to test our approach, we consider two benchmark hyperspectral data sets: Indian Pines and Salinas [60, 61]. Two experiments were performed for each detection task. In the first experiment $\lambda^2 = 1$, in Equation 4.3.4 and a single Principal Component (PC) from each subclass was used to generate a basis for the signal space. For the second experiment $\lambda = 10^5$ in Equation 4.3.4 and two PCs were taken from each subclass to form the basis of our signal space. For example, the class of "corn" contains the three subclasses "corn," "corn-notill," and "corn-midtill." For a "corn" detector, we obtain a signal basis by taking the first PC (or first two PCs) of each subclass to create a three (or six) dimensional signal space for "corn". The subclasses of each broader class are all listed in Figure 4.5.5. Figure 4.5.5 shows the score images for both ACE and ACE-Theta for the second experiment.

Table 4.1 shows the performance of both standard ACE and ACE-Theta as measured by the Area Under the Curve (AUC) as the metric of performance. Additionally, the optimal rank, as

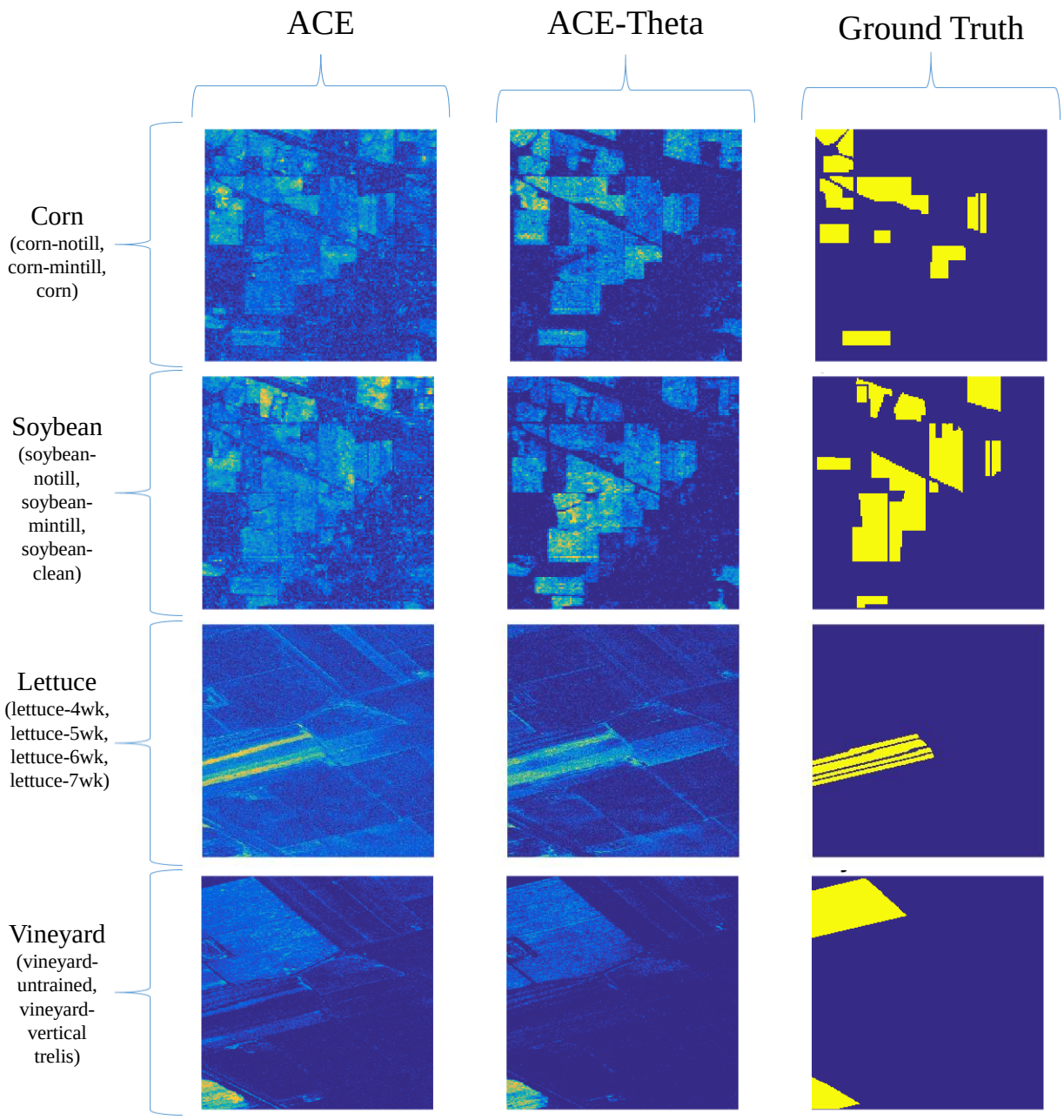


Figure 4.5.5: Images of ACE and ACE-Theta detection scores for each classification task.

Table 4.1: Performance of ACE and ACE-Theta on each classification task where the AUC is reported. For rows 2-5, $\lambda^2 = 1$ and a single PC from each subclass is used to form a signal space basis. Rows 6-9 $\lambda^2 = 10^5$ and two PCs from each subclass generate the signal space basis.

	ACE-Theta	ACE	Opt. Rank
Corn Detector	**0.9432**	0.9212	1
Soybean Detector	**0.9451**	0.9124	1
Lettuce Detector	0.9611	**0.9767**	1
Vineyard Detector	0.9628	**0.9631**	1
Corn Detector	**0.9437**	0.8983	4
Soybean Detector	**0.9221**	0.8691	4
Lettuce Detector	**0.9595**	0.9578	5
Vineyard Detector	**0.9583**	0.9389	3

determined by minimizing Equation 4.3.4, is reported. The highest AUC for each experiment is in bold. We can see that for all of the detection tasks in the second experiment ACE-Theta achieves a higher AUC while in the first experiment ACE-Theta is superior in two of the detection tasks and nearly comparable in the remaining two tasks.

Dimensionality reduction on the signal mode weight vector based on minimization of the bias squared plus variance over all possible ranks yields an order determination rule relating the noise gain to the signal to noise ratio in different signal modes. Specifically, the order determination rule captures the subspace dimensions in which the noise gain is lower than the signal to noise ratio in the corresponding signal mode.

4.6 Discussion

Our motivation for developing these order determination rules was to improve the performance of matched subspace detection like ACE [47] for hyperspectral imaging. In most hyperspectral applications observed data does not come with ground-truth signal and interference mixing matrices. Consequently, to comprehensively evaluate the benefits of our framework we also present results on a synthetic data set in which all parameter values for optimization are known.

Many applications of the LMM suffer when the signal and interference spaces are "close" together. The synthetic data set, representing generalized modal analysis, was generated to evaluate the performance of our framework was constructed to satisfy this property: all the principal angles between our interference and signal spaces are less than $\pi/10$. In the experiments performed all components of the linear mixing model are known, as are the critical values for implementing the order determination rules presented.

Reduced dimension estimators for components of the linear mixing model based on minimization of the bias-squared plus variance over all possible ranks yields order determination rules relating the noise gain to the signal-to-noise ratio in different signal modes. These order determination rules are dependent on parameter values that have an intuitive interpretation and are bounded in the context of real data. An experiment illustrative of modal analysis highlights the improvement in reconstruction when the order determination rules are employed. Reconstruction errors for both signal component and signal mode weight estimation are reduced in the majority of trials over a range of input signal-to-noise ratios that could realistically arise in data acquisition.

To illustrate the methodology on a real world problem we consider detection of four different ground covers: corn, soybean, lettuce, and vineyard. We observe that the detection based on the reduced dimension estimator can achieve superior results relative to the full dimension detector. Strong performance differences between standard matched subspace detector and a dimension reduced matched subspace detector occurred in the experiment with higher dimensional signal spaces suggesting validity to structured dimension reduction in this setting.

Future work, and other real world applications, will require development of techniques for estimating these parameters and components in the absence of prior knowledge.

Chapter 5
Towards Nearly Isometric Maps Between Grassmannian Manifolds

Dimension reduction can be appealing for a wide variety of reasons that have been previously discussed. While there are many reasons a user might desire to reduce the dimension of their data, it should not be done without careful considerations. An important consideration is whether or not information is being lost as the dimension is changed. One way that information is lost is through collapsing. Data collapses when two (or more) points are mapped to the same point in the reduced dimensional space. Consider, for example, two patients represented by vectors of the form [hair color, eye color, age, gender]. The first patient is represented by [red, blue, 28, female] and the second patient by [red, blue, 28, male]. When all coordinates are considered the two patients are distinct points but if we project the patients onto the first three coordinates they are mapped to the same point, [red, blue, 28].

Another way to think about data collapse is in terms of the pairwise distances between points under a mapping. Two distinct points, $\boldsymbol{x}$ and $\boldsymbol{y}$, in a space will have a distance between them that is positive, $d(\boldsymbol{x}, \boldsymbol{y}) > 0$. If those two distinct points are mapped to the same point, $\boldsymbol{z}$, in a reduced dimension space by the function f, then $d(f(\boldsymbol{x}), f(\boldsymbol{y})) = d(z, z) = 0$. Specifically, $d(\boldsymbol{x}, \boldsymbol{y}) \neq d(f(\boldsymbol{x}), f(\boldsymbol{y}))$, i.e. the pairwise distance is not preserved under the mapping. Mappings that exactly preserve all pairwise distances are called *isometric mappings.* Alternatively, maps which nearly preserve all pairwise distances are called *nearly isometric mappings.* Nearly isometric mappings have been extensively studied for maps between Euclidean spaces. In the final project of this dissertation we consider nearly isometric mappings of data existing on a Grassmannian manifold.

The Grassmannian $Gr(N, K)$ is the manifold parameterization where each point corresponds to a $K-$dimensional linear subspace of $\mathbb{R}^N$. There are many applications that lend themselves to problem formulations on a Grassmannian manifold. One example we hope to explore in future work is that of mental task identification via ElectroEncephaloGram (EEG). In EEG voltage is

recorded over time at multiple locations in the brain. The number of time steps at which voltage is recorded could correspond to N while the number of locations where voltage is being recorded could determine K. The data acquired while performing one mental task would yield a single point on the Grassmannian. Clustering could then be performed on the manifold using distances computed in that space. The result of being able to map these points to a lower dimensional manifold ($Gr(N,k)$ with $k < K$, $Gr(n,K)$ with $n < N$, or $Gr(N,K)$ with $n < N$ and $k < K$) would suggest which time increments are important or which brain nodes are more significant. This is just one of many applications that could be well suited to a Grassmannian framework.

Neccessary background and definitions related to Grassmannian manifolds and probability theory can be found in Section 5.1. Section 5.2 states and recreates a statistically based proof of a well known result called the Johnson-Lindenstrauss Lemma which pertains to nearly isometric mappings between Euclidean spaces. In Section 5.3 results related to isometric mappings of data existing on a Grassmannian into Euclidean space are presented and discussed in terms of when they can be considered dimension reduction. The existence of nearly isometric, dimension reducing maps between Grassmannians characterized by different dimensions is conjectured in Section 5.4. Also in this section two potential approaches for proving the conjecture are presented. The first approach is based on creating analagous statistical statements for Grassmannians as are used in the statistical proof of the Johnson-Lindenstrauss Lemma that is presented. The second approach, which we find to be more mathematically compelling and more promising connects nearly isometric mappings to a well known packing problem. Finally, the chapter concludes with a discussion of potential applications and insights in Section 5.5.

5.1 Background

The proof of the Johnson-Lindenstrauss lemma that is reproduced in this chapter relies heavily on statistics and linear algebra. Our proposed approach for creating an analogous lemma and proof for isometric, dimension reducing maps between Grassmannians also rely on the tools. This background section contains some general definitions and proofs that the reader is assumed to be

comfortable with in the later sections of this chapter. Theorems and proofs that are more complex and specifically required for the larger proofs discussed are contained in sections they are first relevant to. An informed reader can refer back to this section for notation as needed and not need to read it in its entirety. Definitions were taken predominantly from [62, 63, 64, 65].

5.1 Grassmannian Definitions and Theorems

Definition 5.1.1 (Group)**.** A *group* is a set $\mathcal{X}$ together with an operation (often called the *group action*) $\star$, denoted $(\mathcal{X}, \star)$, such that

(a) for all $x, y \in \mathcal{X}$, $x \star y \in \mathcal{X}$,

(b) there exists an element $1 \in \mathcal{X}$ such that $1 \star x = x \star 1 = x$ for all $x \in \mathcal{X}$,

(c) and for all $x \in \mathcal{X}$ there exists an element $x^{-1} \in \mathcal{X}$ such that $x \star x^{-1} = x^{-1} \star x = 1$.

Definition 5.1.2 (Real Lie Group)**.** A *real Lie group* is a group that is also a finite-dimensional, real smooth manifold in which the group operations of multiplication and inversion are smooth maps.

Definition 5.1.3 (Topological Space)**.** A *topological space* is a pair denoted $(\mathcal{X}, \tau)$ where $\mathcal{X}$ is a set and τ is a collection of subsets of $\mathcal{X}$ such that

(a) $\emptyset, \mathcal{X} \in \tau$ (where $\emptyset$ indicates the empty set),

(b) arbitrary (finite or infinite) unions of elements in τ are also elements of τ,

(c) and finite intersections of elements in τ are also elements of τ.

The elements of τ are called *open sets* and the collection τ of open sets is called the *topology*. An open set that contains $x \in \mathcal{X}$ is called an *open neighborhood* of x.

Definition 5.1.4 (Continuous Function with Respect to Topology)**.** A function f between two topological spaces X and Y, $f : X \to Y$, is *continuous* if for evey open set $V \subseteq Y$ the inverse image of V ($f^{-1}(V) = \{x \in X \, s.t. \, f(x) \in V\}$) is open in X ($f^{-1}(V) \subseteq X$).

Definition 5.1.5 (Topological Group)**.** A *topological group* is a group together with a topology on the group who's group action is continuous with respect to the topology.

Definition 5.1.6 (Compact Topological Space)**.** A *compact topological space* is a topological space such that every open cover of the topological space has a finite subcover.

Definition 5.1.7 (Metric)**.** Let $\mathcal{M}$ be a set. A function $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is called a *metric* if for any $x, y, z \in \mathcal{M}$ it satisfies

(a) $d(x, y) \geq 0$,

(b) $d(x, y) = 0 \Leftrightarrow x = y$,

(c) $d(x, y) = d(y, x)$, and

(d) $d(x, y) \leq d(x, z) + d(z, y)$.

Definition 5.1.8 (Hausdorff Space)**.** A *Hausdorff space* is a topological space $(\mathcal{X}, \tau)$ satisying that for distint points $x, y \in \mathcal{X}$ there exist open neighborhoods of x and y whose intersection is empty.

Definition 5.1.9 (Locally Compact Group)**.** A *locally compact group* is topological group where the underlying topology is locally compact and Hausdorff.

Definition 5.1.10 (Grassmannian Manifold $Gr(N, K)$)**.** The real *Grassmannian manifold* $Gr(N, K)$ is a space which parameterizes linear subspaces of dimensioin K of $\mathbb{R}^N$. Each point $\mathbf{V}$ on $Gr(N, K)$ corresponds to a linear $K-$dimensional subspace of $\mathbb{R}^N$, $\mathbf{V} \in \mathbb{R}^{N \times K}$. $Gr(N, K)$ is a smooth manifold of dimension $K(N - K)$. Frequently $Gr(N, K)$ is defined as a quotient space by

$$Gr(N, K) = \frac{O(N)}{O(K) \times O(N - K)}.$$

Moreover, $Gr(N, K)$ is a metric space when endowed with the distance function $d(\mathbf{V}, \mathbf{U}) = ||\mathbf{P}_V - \mathbf{P}_U||$ for points $\mathbf{V}, \mathbf{U} \in Gr(N, K)$ with corresponding projection matrices $\mathbf{P}_V$ and $\mathbf{P}_U$, respectively.

Definition 5.1.11 (Principal Angles)**.** Let $\mathbf{U}$ and $\mathbf{V}$ be two linear subspaces of $\mathbb{R}^N$ of dimensions K_U and K_V, respectively. Assume, without loss of generality that $K_U \leq K_V$. There exists a sequence of K_U angles $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_{K_U} \leq \frac{\pi}{2}$ called the *principal angles* between $\mathbf{U}$ and $\mathbf{V}$. The first principal angle is defined by

$$\theta_1 = \min \left\{ \arccos\left(\frac{|\boldsymbol{u}^T \boldsymbol{v}|}{||\boldsymbol{u}||_2 ||\boldsymbol{v}||_2} \right) \quad \text{such that} \quad \boldsymbol{u} \in \mathbf{U}, \boldsymbol{v} \in \mathbf{V} \right\}.$$

The remaining principal angles are then defined recursively by

$$\theta_i = \min \left\{ \arccos\left(\frac{|\boldsymbol{u}^T \boldsymbol{v}|}{||\boldsymbol{u}||_2 ||\boldsymbol{v}||_2} \right) \text{ s.t. } \boldsymbol{u} \in \mathbf{U}, \boldsymbol{v} \in \mathbf{V}, \boldsymbol{u} \perp \boldsymbol{u}_j, \boldsymbol{v} \perp \boldsymbol{v}_j \, \forall j \in \{1, \ldots, i-1\} \right\}.$$

Theorem 5.1.1 (SVD to Compute Principal Angles)**.** *Let $\boldsymbol{A} \in \mathbb{R}^{N \times a}$ and $\boldsymbol{B} \in \mathbb{R}^{N \times b}$ be real matrices whose column spaces form orthonormal bases for $\langle \boldsymbol{A} \rangle$ and $\langle \boldsymbol{B} \rangle$, respectively. Assume, without loss of generality that $a \leq b$. Let the SVD of $\boldsymbol{A}^T \boldsymbol{B}$ be $\mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T$ where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and $\boldsymbol{\Sigma}$ is an $a \times b$ diagonal matrix with $diag(\boldsymbol{\Sigma}) = [\sigma_1, \ldots, \sigma_a]^T$. Then $\cos(\theta_i) = \sigma_i$ where θ_i is the i^{th} principal angle between $\langle \boldsymbol{A} \rangle$ and $\langle \boldsymbol{B} \rangle$.*

Definition 5.1.12 (Chordal Distance on the Grassmannian)**.** Let $\mathbf{U}, \mathbf{V} \in Gr(N, K)$. The *chordal distance* between $\mathbf{U}$ and $\mathbf{V}$ is defined as

$$d_c(\mathbf{U}, \mathbf{V}) = \sqrt{\sin^2(\theta_1) + \ldots \sin^2(\theta_K)}$$

where θ_i is the i^{th} principal angle between $\mathbf{U}$ and $\mathbf{V}$.

Theorem 5.1.2 (Chordal Distance from Frobenius Norm of Difference in Projection Matrices)**.** *Let $\mathbf{U}, \mathbf{V} \in Gr(N, K)$ and let $\{\theta_1, \ldots, \theta_K\}$ be the K principal angles between $\mathbf{U}$ and $\mathbf{V}$. Without loss of generality, assume that the coolumns of $\mathbf{U}$ and $\mathbf{V}$ form orthonormal bases for their respective spaces. Then the following two statements hold:*

1. *$d_c^2(\mathbf{U}, \mathbf{V}) = \frac{1}{2} ||\mathbf{P}_U - \mathbf{P}_V||_F^2$ where $\mathbf{P}_U$ and $\mathbf{P}_V$ are the orthogonal projection matrices onto $\mathbf{U}$ and $\mathbf{V}$, respectively,*

2. *and* $d_c^2(\mathbf{U}, \mathbf{V}) = K - ||\mathbf{U}^T\mathbf{V}||_F^2$.

5.1 Probability Definitions and Theorems

Definition 5.1.13 ($\sigma-$algebra)**.** Let $\mathcal{X}$ be a set and let $P(\mathcal{X})$ denote its power set (the set of all possible subsets of $\mathcal{X}$). A subset $\Sigma \subset P(\mathcal{X})$ is called a $\sigma-$*algebra over* $\mathcal{X}$ if it satisfies

(a) $\mathcal{X} \in \Sigma$ (where $\emptyset$ indicates the empty set),

(b) Σ is closed under complements (i.e. if $\sigma \in \Sigma$ then $\mathcal{X}/\sigma \in \Sigma$),

(c) and Σ is closed under finite unions.

Elements of a $\sigma-$algebra are called *measurable sets*. An odered pair $(\mathcal{X}, \Sigma)$ where $\mathcal{X}$ and Σ is a $\sigma-$algebra over $\mathcal{X}$ is called a *measurable space*.

Definition 5.1.14 (Borel Set)**.** Let $(\mathcal{X}, \tau)$ be a topological space. Any set in the topological space that can be formed from open sets (elements of τ) via the operations of (1) countable union, (2) countable intersection, and (3) relative complements is called a *Borel set.*

Definition 5.1.15 (Borel Algebra)**.** The *Borel algebra* is the collection of all the Borel sets of $(\mathcal{X}, \tau)$. The Borel algebra is a $\sigma-$algebra and is the smallest $\sigma-$algebra containing all the elements of τ (containing all the open sets as defined by the topology).

Definition 5.1.16 (Measure)**.** Let $(\mathcal{X}, \Sigma)$ be a measurable space. Let $\mu : \Sigma \to (0, \infty)$. We say μ is *measure* on $(\mathcal{X}, \Sigma)$ if

(a) for all $\sigma \in \Sigma$ we have $\mu(\sigma) \geq 0$,

(b) $\mu(\emptyset) = 0$,

(c) for all countable collections $\{\sigma_i\}_{i=1}^{\infty}$ of pairwise disjoint measurable sets, $\sigma_i \cap \sigma_j = \emptyset \, \forall i, j$,

$$\mu\left(\bigcup_{i=1}^{\infty} \sigma_i\right) = \sum_{i=1}^{\infty} \mu(\sigma_i).$$

Remark. The Borel algebra is the smallest $\sigma-$algebra that makes all open sets measureable.

Definition 5.1.17 (Left(Right)-Translation-Invariant)**.** Let $(\mathcal{X}, \tau, \star)$ be a locally compact Hausdorff topological group. Let Σ be the Borel algebra of $(\mathcal{X}, \tau)$. If $x \in \mathcal{X}$ and $\sigma \in \Sigma$ we define the *left (right) translate of* σ as

$$\text{Left Translate} = x\sigma = \{x \star s \text{ s.t. } s \in \sigma\}$$

$$\text{Right Translate} = \sigma x = \{s \star x \text{ s.t. } s \in \sigma\}$$

A measure μ on Σ is called *left(right)-translation-invariant* if for all $\sigma \in \Sigma$ and all $x \in \mathcal{X}$ then $\mu(\sigma) = \mu(x\sigma)$ ($\mu(\sigma) = \mu(\sigma x)$).

Theorem 5.1.3 (Haar's Theorem)**.** *There is a unique (up to a multiplicative constant), nontrivial measure μ on the Borel subsets of $(\mathcal{X}, \tau)$ satisfying*

(a) μ is left-translation-invariant,

(b) the measure μ is finite on every compact set,

(c) $\mu(\sigma) = \inf\{\mu(\tilde{\sigma})$ s.t. $\sigma \subset \tilde{\sigma}$, $\tilde{\sigma}$ open$\}$,

(d) and $\mu(\sigma) = \sup\{\mu(\tilde{\sigma})$ s.t. $\tilde{\sigma} \subset \sigma$, $\tilde{\sigma}$ compact$\}$.

Such a measure is called a left Haar measure*. The definition for the* right Haar measure *is similar and it is important to note that the left and right Haar measures need not coincide.*

Definition 5.1.18 (Probability Space)**.** A *probability space* is a triplet denoted $(\mathcal{X}, \tau, \mu)$. We refer to $\mathcal{X}$ as the *sample space* and is the set of all possible outcomes. Σ is a $\sigma-$algebra of $\mathcal{X}$ and μ is a *probability measure* (a measure whose integral over Σ equals to one).

Definition 5.1.19 (Uniform Distribution with Respect to Haar Measure)**.** Let $(\mathcal{X}, \Sigma, \mu)$ be a probability space where Σ is the Borel algebra over $\mathcal{X}$ and μ is the left Haar probability measure on Σ. The *uniform distribution* with respect to Haar measure assigns equal probability to each Borel subset. Every element in Σ is equally likely under the Haar probability measure.

Remark. Although in some senses a uniform distribution is somewhat intuitive when we simplify it to "every outcome is equally likely," methods for generating pseudo random events according to a uniform distribution can be far from trivial to prove. Another useful interpretation of a uniform distribution is a distribution such that the chance of an event is unchanged by left transformation (whatever appropriate invariant transformation).

Theorem 5.1.4 (Uniform Random Points on $Gr(N, K)$ (Theorem 2.2.2 in [65]))**.** *The following statements are in regard to uniform random points on* $Gr(N, K)$:

(a) If $\mathbf{U}$ *is uniformly distributed on* $Gr(N, K)$ *so is* $\mathbf{HUH}^T$ *for any* $\mathbf{H} \in O(N)$ *which is independent of* $\mathbf{U}$*, and hence we have that* $\mathbb{E}[\mathbf{U}] = \frac{K}{N}\mathbf{I}_N$.

(b) Let $\mathbf{H}$ *be uniformly distributed on* $O(N)$ *and let* $\mathbf{U}_0$ *be a point on* $Gr(N, K)$ *(represented by its projection matrix* $\mathbf{P}_{U_0}$*), constant or independent of* $\mathbf{H}$*. Then the random matrix* $\mathbf{HP}_{U_0}\mathbf{H}^T$ *is uniformly distributed on* $O(N)$.

(c) $\mathbf{P}_U = \mathbf{UU}^T$ *is uniform on* $Gr(N, K)$ *if and only if* $\mathbf{U} = \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1/2}$ *for* $\mathbf{V} \in \mathbb{R}^{N\times K}$ *with all of the elements of* $\mathbf{V}$ *independent and identically distributed according to the standard normal* $N(0, 1)$.

Remark. In [65], as well as in many other papers and texts, points on $Gr(N, K)$ are frequently represented by the orthogonal projection onto the space spanned by the point. This allows particular properties of projection matrices to be leveraged or exploited. The three components of the above Theorem will be leveraged heavily in our contributions towards proving the existence of dimension reducing maps between Grassmannians.

5.2 Johnson-Lindenstrauss Lemma

The statement of the *Johnson-Lindenstrauss Lemma* (JLL) as it applies to mappings over $\mathbb{R}$ is as follows:

Lemma 5.2.1 (Johnson-Lindenstrauss). *For any $0 < \epsilon < 1$ and any integer m, let n be a positive integer such that*

$$n \geq 4(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})^{-1} \ln(m).$$

Then, for any set V of m points in $\mathbb{R}^N$, there is a map $f : \mathbb{R}^N \to \mathbb{R}^n$ such that for all $u, v \in V$

$$(1 - \epsilon)||u - v||_2^2 \leq ||f(u) - f(v)||_2^2 \leq (1 + \epsilon)||u - v||_2^2.$$

This lemma falls out of the original paper by Johnson and Lindenstrauss on extensions of Lipschitz-continuous maps into Hilbert space [6]. The proof recreated here follows that of Dasgupta and Gupta [7] but contains more complete algebraic manipulations and commentary on the theory behind many of the algebraic statements.

5.2 Proofs of Theorems and Lemmas Used in Statistical Proof of JLL

The following subsections will prove several key facts used in the proof of the JLL. These facts include the expected length of the projection of a unit length random vector from an ambient space into a lower dimensional space, Markov's Inequality, a pertinent lemma. Key ideas from statistics will be addressed as they are needed.

5.2 Expected Length of a Random Vector Under Projection

The JLL, as stated, boils down to wanting to predict the length of the image of a unit vector in the domain, $\mathbb{R}^N$, in the new space $\mathbb{R}^n$. In [7], they state the challenge as "Estimating the length of a unit vector in $\mathbb{R}^n$ when it is projected onto a random $K-$dimensional subspace." Primarily, the tools involved in this proof of the JLL come from statistics, and in particular the chi-squared distribution. To make use of this particular distribution requires the use of the Euclidean norm in the statement of the problem. In order to attack this problem, the authors of [7] state that distributions of the length of a fixed unit vector in $\mathbb{R}^N$ after being projected onto a random $K-$dimensional sub-

space, and the length of a random vector in $\mathbb{R}^N$ projected onto a fixed $K-$dimensional subspace, are equivalent, and therefore it is sufficient to consider the latter distribution.

The statement that the length of the projection of a fixed unit vector (without loss of generality assumed to be the first standard basis vector) onto a random $K-$dimensional subspace is equivalent to the length of the projection of a random unit vector onto a fixed $K-$dimensional subspace (without loss of generality assumed to be the subspace spanned by the first K standard basis vectors) is made in every reproduction of the proof given in [6].

We take the opportunity to formally prove the statement.

Theorem 5.2.2. *The following distributions are equivalent:*

(a) Length of projection of a fixed unit length vector in $\mathbb{R}^N$, $\mathbf{y} \in \mathbb{R}^N$, onto a uniform random (with respect to Haar measure) $K-$dimensional subspace, $\mathbf{V} \in \mathbb{R}^{N\times K}$.

(b) Length of the projection of a uniform random unit vector (with respect to Haar measure) in $\mathbb{R}^N$, $\mathbf{y} \in \mathbb{R}^N$ such that $||\mathbf{y}||_2 = 1$, onto a fixed $K-$dimensional subspace, $\mathbf{V} \in \mathbb{R}^{N\times K}$.

Proof. Consider a fixed unit vector in $\mathbb{R}^N$. Without loss of generality, let the fixed unit vector under consideration be the first standard basis vector of of $\mathbb{R}^N$, $\mathbf{y} = \boldsymbol{e}_1 = [1, 0, \ldots, 0]^T$. Let $\mathbf{V}$ be a basis for a fixed $K-$dimensional subspace. Without loss of generality, we let $\mathbf{V}$ be the subspace spanned by the first K standard basis vectors in $\mathbb{R}^N$,

$$\begin{aligned}\mathbf{V} &= \begin{bmatrix}\boldsymbol{e}_1 & \boldsymbol{e}_2 & \ldots & \boldsymbol{e}_K\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{I}_K \\ \mathbf{0}_{N-K,K}\end{bmatrix}.\end{aligned}$$

Let $\mathbf{U}$ be a random element of the orthogonal group, $\mathbf{U} \in O(N)$. A uniform random (with respect to the Haar measure) unit vector in $\mathbb{R}^N$ can be generated as $\mathbf{y}_0 = \mathbf{U}^T\mathbf{y}$. Similarly, a uniform random (with respect to the Haar measure) $K-$dimensional subspace can be generated as $\mathbf{V}_0 = \mathbf{UV}$. Recall, the projection onto the column space of $\boldsymbol{A}$ can be written as $\mathbf{P}_A = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$.

Consider the length of the fixed vector $\mathbf{y}$ projected onto the random $K-$dimensional subspace $\mathbf{V}_0$. We yield the following sequence of equalities

$$\begin{aligned}
||\mathbf{P}_{V_0}\mathbf{y}||_2^2 &= (\mathbf{P}_{V_0}\mathbf{y})^T(\mathbf{P}_{V_0}\mathbf{y}) \\
&= \mathbf{y}^T\mathbf{P}_{V_0}^T\mathbf{P}_{V_0}\mathbf{y} = \mathbf{y}^T\mathbf{P}_{V_0}\mathbf{y} = \mathbf{y}^T(\mathbf{V}_0(\mathbf{V}_0^T\mathbf{V}_0)^{-1}\mathbf{V}_0^T)\mathbf{y} \\
&= \mathbf{y}^T(\mathbf{U}\mathbf{V}((\mathbf{U}\mathbf{V})^T(\mathbf{U}\mathbf{V}))^{-1}(\mathbf{U}\mathbf{V})^T)\mathbf{y} \\
&= \mathbf{y}^T(\mathbf{U}\mathbf{V}(\mathbf{V}^T\mathbf{U}^T\mathbf{U}\mathbf{V})^{-1}\mathbf{V}^T\mathbf{U}^T\mathbf{y}.
\end{aligned}$$

From the assumption that $\mathbf{U} \in O(N)$ we have that $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_N$. Using this fact we have

$$\begin{aligned}
||\mathbf{P}_{V_0}\mathbf{y}||_2^2 &= \mathbf{y}^T(\mathbf{U}\mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T\mathbf{U}^T)\mathbf{y} \\
&= \mathbf{y}^T\mathbf{U}\mathbf{P}_V\mathbf{U}^T\mathbf{y} = \mathbf{y}^T\mathbf{U}\mathbf{P}_V^T\mathbf{P}_V\mathbf{U}^T\mathbf{y} = (\mathbf{U}^T\mathbf{y})^T\mathbf{P}_V^T\mathbf{P}_V(\mathbf{U}^T\mathbf{y}) \\
&= \mathbf{y}_0^T\mathbf{P}_V^T\mathbf{P}_V\mathbf{y}_0 \\
&= (\mathbf{P}_V\mathbf{y}_0)^T(\mathbf{P}_V\mathbf{y}_0) \\
&= ||\mathbf{P}_V\mathbf{y}_0||_2^2.
\end{aligned}$$

The value $||\mathbf{P}_V\mathbf{y}_0||_2^2$ is the length of projection of the uniform random unit vector $\mathbf{y}_0$ onto the fixed $K-$dimensional subspace spanned by $\mathbf{V}$. Thus, we have that the distributions of the lengths are equivalent. □

Once the equivalence of these two statements is proved, it is sufficient to consider the expected value of the length of the projection of the uniform random unit vector onto its first K coordinates. In the proof of the previous theorem we used the fact that a uniform random unit vector in $\mathbb{R}^N$ can be generated by left multiplication of a fixed unit vector by a random element of $O(N)$. An alternative way of producing a uniform random unit vector in $\mathbb{R}^N$ is by populating a vector with random entries drawn from the standard normal distribution, $N(0,1)$, and normalizing it to have unit length. For the proof of the JLL in [7] the uniform random unit vector is generated using the second approach. Generation of the random unit vector in this fashion allows us to make and prove

the following theorem about the expected length of a random vector after projection onto its first K coordinates.

Theorem 5.2.3 (Expected Length of a Random Unit Vector After Projection)**.** *Let* $\mathbf{y}$ *be a uniform random unit vector in* $\mathbb{R}^N$*. If* $\boldsymbol{L}$ *is the length of the projection of* $\mathbf{y}$ *onto its first* K *coordinates, then* $\mathbb{E}[L] = K/N$.

Proof. Let $\mathbf{x} = [x_1, x_2, ..., x_N]^T$ be a random vector in $\mathbb{R}^N$, where each x_i is drawn from $N(0, 1)$, the standard normal distribution with mean zero and deviation 1. Now, let $\mathbf{y} = \frac{1}{||\boldsymbol{x}||_2}\boldsymbol{x}$. Hence, $\mathbf{y}$ can also be considered a uniform random unit vector in $\mathbb{R}^N$. A uniform random unit vector can also be thought of as a random point on the surface of the $N-1$ dimensional sphere. Recalling that each x_i is drawn from $N(0, 1)$ implies that x_i^2 is drawn from a Chi-Squared distribution with one degree of freedom, $x_i^2 \sim \chi^2(1)$. From this we have that $\mathbb{E}[x_i^2] = 1$ for all i. Using the linearity of expectation we have that

$$\begin{aligned}\mathbb{E}[||\boldsymbol{x}||_2^2] &= \mathbb{E}[x_1^2 + x_2^2 + ... + x_N^2] = \mathbb{E}[x_1^2] + \mathbb{E}[x_2^2] + \cdots + \mathbb{E}[x_N^2] \\ &= \sum_{i=1}^{N} \mathbb{E}[x_i^2] = N\mathbb{E}[x_i^2] = N\end{aligned}$$

Now consider $\mathbf{z}$ to be the projection of $\mathbf{y}$ onto its first K coordinates,

$$\begin{aligned}\boldsymbol{z} &= [y_1, y_2, \ldots, y_K]^T \\ &= [\frac{x_1}{||\boldsymbol{x}||_2}, \frac{x_2}{||\boldsymbol{x}||_2}, \ldots, \frac{x_K}{||\boldsymbol{x}||_2}]^T.\end{aligned}$$

Let $\boldsymbol{L}$ be the squared length of $\boldsymbol{z}$, $\boldsymbol{L} = ||\boldsymbol{z}||_2^2$. Then we have

$$\begin{aligned}\mathbb{E}[\mathbf{L}] = \mathbb{E}[||\mathbf{Z}||_2^2] &= \mathbb{E}[\frac{x_1^2 + x_2^2 + \cdots + x_K^2}{||\boldsymbol{x}||_2^2}] \\ &= \mathbb{E}[\frac{x_1^2 + x_2^2 + \cdots + x_K^2}{x_1^2 + x_2^2 + \cdots + x_N^2}]\end{aligned}$$

It can be shown that if x_i are independently identically distributed, $\boldsymbol{s}_k = \sum_{i=1}^{k} x_i$, and $m < n$, then $\mathbb{E}[\boldsymbol{s}_m/\boldsymbol{s}_n] = m/n$. Consequently,

$$\mathbb{E}[\mathbf{L}] = \mathbb{E}[\frac{x_1^2 + x_2^2 + \cdots + x_K^2}{x_1^2 + x_2^2 + \cdots + x_N^2}] = \frac{K}{N}.$$

□

5.2 Markov's Inequality

A key component of the proof in [7] is Markov's Inequality (MI). This inequality relates the probability of a random variable exceeding a value to the expected value of the random variable. The statement of the the inequality as well as its proof follow.

Theorem 5.2.4 (Markov's Inequality)**.** *Let X be a random variable belonging to some probablility distribution, and $X, a \geq 0$. Then, $P[X \geq a] \leq \frac{1}{a}\mathbb{E}[X]$, where $\mathbb{E}[X]$ is the expected value of the variable X.*

Proof. In order to prove MI we first define the indicator function. For any event E, let I_E be the indicator function of E defined as

$$I_E = \begin{cases} 1 & \text{if} \quad \text{E occurs} \\ 0 & \text{if} \quad \text{E does not occur} \end{cases}.$$

Consequently, we can define

$$I_{x \geq a} = \begin{cases} 1 & \text{if} \quad x \geq a \\ 0 & \text{if} \quad x < a \end{cases}.$$

One can show that $aI_{x \geq a} \leq x$. First, if $I_{x \geq a} = 0$, then we have that $aI_{x \geq a} = a \cdot 0 = 0 \leq x$. Second, $I_{x \geq a} = 1$, then we have that $aI_{x \geq a} = a \cdot 1 = a \leq x$, since in order for $I_{x \geq a} = 1$ we have that $a \leq x$. Additionally, we recall that expectation is a linear operator. Consequently, from $aI_{x \geq a} \leq x$ we have $\mathbb{E}[aI_{x \geq a}] \leq \mathbb{E}[x]$. Furthermore,

$$\begin{aligned}\mathbb{E}[aI_{x\geq a}] &= a\mathbb{E}[I_{x\geq a}] \\ &= a(1 \cdot P[x \geq a] + 0 \cdot P[x < a]) \\ &= aP[x \geq a]\end{aligned}$$

Thus, $\mathbb{E}[aI_{x\geq a}] \leq \mathbb{E}[x]$ implies that $aP[x \geq a] \leq \mathbb{E}[x]$ and finally that

$$P[x \geq a] \leq \frac{\mathbb{E}[x]}{a}.$$

□

5.2 Bounding Probability Using the Moment Generating Function

The proof of JLL in [7] heavily relies on a lemma presented in the paper. Primarily, this lemma bounds the probability of a Chi-Squared distribution exceeding a particular value by leveraging MI and the closed form of the moment generating function of a Chi-Squared distribution. The lemma is stated and proved as two pieces. The complete statement is provided and a proof of one piece is shown in detail and the second piece (which is proved using similar construction) is left to the reader.

Lemma 5.2.5 (Lemma 1 [7])**.** *Let $K < N$ and let* **L** *be as as previously defined (i.e. the length of a uniform random unit vector in $\mathbb{R}^N$ after projection onto its first K coordinates).*

(a) If $\beta < 1$, then

$$\begin{aligned}P[\mathbf{L} \leq \beta(K/N)] &\leq \beta^{\frac{K}{2}}(1 + \frac{(1-\beta)K}{N-K})^{\frac{N-K}{2}} \\ &\leq exp(\frac{K}{2}(1 - \beta + \ln(\beta))).\end{aligned}$$

(b) If $\beta > 1$, then

$$P[\mathbf{L} \geq \beta(K/N)] \leq \beta^{\frac{K}{2}}(1 + \frac{(1-\beta)K}{N-K})^{\frac{N-K}{2}}$$
$$\leq exp(\frac{K}{2}(1 - \beta + \ln(\beta))).$$

Proof. We will proceed by proving Lemma 1(a). The proof of Lemma 1(b) is very similar. Let $\beta < 1$. We begin by rewriting $P[\mathbf{L} \leq \beta(K/N)]$ as follows

$$P[\mathbf{L} \geq \beta(K/N)] = P[N(x_1^2 + x_2^2 + ... + x_k^2) \leq \beta K(x_1^2 + x_2^2 + ... + x_N^2)]$$
$$= P[\beta K(x_1^2 + x_2^2 + ... + x_N^2) - N(x_1^2 + x_2^2 + ... + x_K^2) \geq 0].$$

We now use a property from probability theory which allows us to write

$$= P[e^{t(\beta K(x_1^2+x_2^2+...+x_N^2)-N(x_1^2+x_2^2+...+x_K^2))} \geq e^{t\cdot 0}]$$
$$= P[e^{t(\beta K(x_1^2+x_2^2+...+x_N^2)-N(x_1^2+x_2^2+...+x_K^2))} \geq 1],$$

for all $t > 0$.

Using MI we have

$$P[e^{t(\beta K(x_1^2+x_2^2+...+x_N^2)-N(x_1^2+x_2^2+...+x_K^2))} \geq 1] \leq \mathbb{E}[e^{t(\beta K(x_1^2+x_2^2+...+x_N^2)-N(x_1^2+x_2^2+...+x_K^2))}].$$

Next, recalling that x_i are all drawn from the same normal distribution and rules of exponents, we have that

$$P[e^{t(\beta K(x_1^2+x_2^2+...+x_N^2)-N(x_1^2+x_2^2+...+x_K^2))} \geq 1] \leq \mathbb{E}[e^{t(\beta K)X^2}]^{N-K}\mathbb{E}[e^{t(\beta K-N)X^2}]^K,$$

where X is a single random normal variable, and X^2 is equivalent to a chi-squared distribution with one degree of freedom. Further, the expected value of a distribution is equivalent to the first moment where the moment generating function for a random variable X is defined as

$$M_X = \mathbb{E}[e^{tX}],$$

for $t < \frac{1}{2}$. The moment for a chi-squared distribution with k degrees of freedom is given by

$$M_{\chi^2_K} = (1-2t)^{-\frac{K}{2}}.$$

Consequently, we have

$$\begin{aligned}\mathbb{E}[e^{t(\beta K)X^2}]^{N-K} &= ((1-2tK\beta)^{-\frac{1}{2}})^{N-K} \\ \mathbb{E}[e^{t(\beta K-N)X^2}]^{K} &= ((1-2t(K\beta-N))^{-\frac{1}{2}})^{K},\end{aligned}$$

whenever $(1-2tK\beta) < \frac{1}{2}$ and $(1-2t(K\beta-N)) < \frac{1}{2}$. We note that the first bound implies the second and so we only consider the first constraint. Combining all of these concepts we have that

$$P[\mathbf{L} \leq \beta(K/N)] \leq (1-2tK\beta)^{-\frac{N-K}{2}}(1-2t(K\beta-N))^{-\frac{K}{2}}.$$

Let

$$g(t) = (1-2tK\beta)^{-\frac{N-K}{2}}(1-2t(K\beta-N))^{-\frac{K}{2}}.$$

We want to bound $P[\mathbf{L} \leq \beta(K/N)]$ tightly, so we consider the following optimization problem

$$\begin{aligned}&\text{minimize} \quad g(t) \\ &\text{subject to} \quad 0 < t < (1-2tK\beta).\end{aligned}$$

Minimizing $g(t)$ is equivalent to maximizing

$$f(t) = (1-2tK\beta)^{N-K}(1-2t(K\beta-N))^{K}$$

since $g(t) = 1/\sqrt{f(t)}$. To find a maximum we first compute the derivative. We find that $f'(t) = A \cdot B$ when we define

$$A = -2K(1 - 2t(K\beta - N))^{K-1}(1 - 2tK\beta)^{N-K-1}$$

$$B = -N + \beta N - 2tK\beta^2 N + 2t\beta N^2.$$

Solving $f'(t) = 0$ amounts to solving $B = 0$ which occurs when

$$t^* = \frac{1-\beta}{2\beta(N - \beta K)}.$$

Thus, the maximal value of $f(t)$ is, after much algebraic simplification,

$$f(t^*) = (\frac{N-K}{N-\beta K})^{N-K}(\frac{1}{\beta})^K.$$

Resultantly, the minimum value of $g(t)$ is

$$g(t^*) = \frac{1}{\sqrt{f(t^*)}} = \beta^{K/2}(1 + \frac{K(1-\beta)}{N-K})^{\frac{N-K}{2}}.$$

Furthermore, we recall that

$$\lim_{n\to+\infty}(1 + \frac{x}{n})^n = e^{(x)}.$$

Using this fact, we obtain that

$$\begin{aligned}\beta^{K/2}(1 + \frac{K(1-\beta)}{N-K})^{\frac{N-K}{2}} &\leq \beta^{\frac{K}{2}} exp(\frac{K}{2}(1-\beta)) \\ &= exp(\frac{K}{2}(1 - \beta + \ln(\beta))).\end{aligned}$$

Thus, we have shown that

$$P[\mathbf{L} \geq \beta(K/N)] \leq exp(\frac{K}{2}(1 - \beta + \ln(\beta))),$$

which is Lemma 5.2.5(a). □

5.2 Proof of the Johnson-Lindenstrauss Lemma

We are now prepared to prove the JLL.

Lemma 5.2.6 (Johnson-Lindenstrauss)**.** *For any $0 < \epsilon < 1$ and any integer m, let n be a positive integer such that*

$$n \geq 4(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})^{-1} \ln(m).$$

Then, for any set V of m points in $\mathbb{R}^N$, there is a map $f : \mathbb{R}^N \to \mathbb{R}^n$ such that for all $u, v \in V$

$$(1 - \epsilon)||u - v||_2^2 \leq ||f(u) - f(v)||_2^2 \leq (1 + \epsilon)||u - v||_2^2.$$

Proof. We consider the non-trivial case where $K < N$. Let ϵ be a value such that $0 < \epsilon < 1$. Let m be an integer, and let K be a positive integer such that

$$K \geq 4(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})^{-1} \ln(m).$$

Further, let V be a set of m points in $\mathbb{R}^N$. Let $\mathbb{S}$ be a random $K-$dimensional subspace. We let v_i' be the projection of $v_i \in V$ into $\mathbb{S}$. Set $\mathbf{L} = ||v_i' - v_j'||_2^2$ and $\mu = \frac{K}{N}||v_i' - v_j'||_2^2$. Then, from Lemma 5.2.5(a) we have that

$$P[\mathbf{L} \leq (1 - \epsilon)\mu] \leq exp(\frac{K}{2}(1 - (1 - \epsilon) + \ln(1 - \epsilon))).$$

We now recall the Taylor series expansion for $\ln(1 - \epsilon)$ centered at zero, to obtain

$$\begin{aligned} \ln(1 - \epsilon) &\approx \ln(1) - 1(\epsilon) - \frac{\epsilon^2}{2} + \dots \\ &\leq -\epsilon - \frac{\epsilon^2}{2}. \end{aligned}$$

Consequently, we obtain the following set of inequalities

$$
\begin{aligned}
P[\mathbf{L} \leq (1-\epsilon)\mu] &\leq exp(\frac{K}{2}(1-(1-\epsilon)+\ln(1-\epsilon))) \\
&\leq exp(\frac{K}{2}(1-(1-\epsilon)+(-\epsilon-\frac{\epsilon^2}{2}))) \\
&= exp(-\frac{K\epsilon^2}{4}) \\
&\leq exp(-2\ln(m)) \leq \frac{1}{m^2}.
\end{aligned}
$$

Similarly, we have from Lemma 5.2.5(b) that

$$
P[\mathbf{L} \geq (1+\epsilon)\mu] \leq exp(\frac{K}{2}(1-(1+\epsilon)+\ln(1+\epsilon))).
$$

Recalling the Taylor series expansion of $\ln(1+\epsilon)$ centered at zero we have that

$$
\begin{aligned}
\ln(1+\epsilon) &\approx \ln(1) + 1(\epsilon) - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{3}... \\
&\leq \epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{3}.
\end{aligned}
$$

This affects the following the chain of inequalities

$$
\begin{aligned}
P[\mathbf{L} \geq (1+\epsilon)\mu] &\leq exp(\frac{K}{2}(1-(1+\epsilon)+\ln(1+\epsilon))) \\
&\leq exp(\frac{K}{2}(1-(1+\epsilon)+(+\epsilon-\frac{\epsilon^2}{2}+\frac{\epsilon^3}{3}))) \\
&= exp(\frac{K}{2}(-\frac{\epsilon^2}{2}+\frac{\epsilon^3}{3})) \\
&\leq exp(-2\ln(m)) \leq \frac{1}{m^2}.
\end{aligned}
$$

By the above calculations, if we choose some mapping f such that $f(v_i) = \sqrt{\frac{N}{K}}(v_i')$ we have that for some pair i, j the chance that

$$
\frac{||f(v_i) - f(v_j)||_2^2}{||v_i - v_j||_2^2}
$$

does not fall in the range $[1-\epsilon, 1+\epsilon]$ is at most $2/m^2$. There are $\binom{m}{2}$ possible pairs from m points. Thus, by the trivial union bound, the chance that a single pair is stretched or shrunken by a factor

not in $[1-\epsilon, 1+\epsilon]$ is at most

$$\binom{m}{2} \times \frac{2}{m^2} = (1 - \frac{1}{m}).$$

Hence, f has the desired property with probability at least $1/m$, and so by picking a random projection $O(m)$ times we can boost the success to any desired level. □

5.2 Subspace Johnson-Lindenstrauss

There has been some research into a variation of Johnson-Lindenstrauss that can be considered to be a subspace variation of the JLL [66, 67]. In these variations the set of m points in $\mathbb{R}^N$ are assumed to be linearly independent and are considered as an $m-$dimensional linear subspace of $\mathbb{R}^N$. These papers assume the existence of a distance preserving map on the initial set of points. Probability statements are then made about the chance that all pairwise distances are preserved when an additional point, contained in the $m-$dimensional subspace, is added and the map is applied to the new point. Moreover, bounds for selecting reasonable isometry constants are produced and defined in terms of other parameters.

This is a fundamentally different challenge than the one we present and are interested in. Of interest to us is the existence of a mapping that preserves the distances between different subspaces not the just the points contained inside of a single subspace. Moreover, this problem performs all comptations in real space while we seek to leverage the structure of the Grassmannians and preserve distances as they would be computed in that space.

5.3 Isometric and Nearly Isometric Mappings of $Gr(N, K)$ into $\mathbb{R}^D$

Isometric, and nearly isometric, mappings of Grassmannians into Euclidean spaces have been considered over the years. The most cited paper discussing the mapping of Grassmannians into Euclidean spaces isometrically in [68]. It culminates in a beautiful theoretical result (Theorem 5.3.1) and is accompanied by many computational results. Another approach has been taken by others which leverages the popular Multi-Dimensional Scaling algorithm of [69]. One such example is

found in the work in [70] in which hyperspectral images are mapped first to a Grassmannian and then isometrically into Euclidean space.

While there is certainly compelling mathematics in these approaches, we are interested in maps that we will define as *dimensionally reductive*.

Definition 5.3.1 (Dimensionally Reductive Map)**.** Let $f : X \to Y$ be a mapping of data existing in the space X into the space Y. We say f is *dimensionally reductive* if $dim(X) < dim(Y)$.

Explicitly, if $f : Gr(N, K) \to Y$ where Y is a new manifold or space, the map f will be said to be dimensionally reductive if $dim(Y) \leq dim(Gr(N, K)) = K(N - K)$. Furthermore, an isometric, dimensionally reductive map is one satisfying that it is dimensionally reductive and exactly preserves all pairwise distances.

Theorem 5.3.1 (Theorem 2 [68])**.** *The representation of $K-$planes in $\mathbb{R}^N$ (let $\mathbf{V} \in Gr(N, K)$) by their projection matrices ($\mathbf{P}_V$) gives an isometric embedding (with respect to chordal distance) of $Gr(N, K)$ into a sphere of radius $\sqrt{K(N - K)/N}$ in $\mathbb{R}^D$, $D = \binom{N+1}{2} - 1$.*

In order to consider this isometric embedding to be dimension reductive we would need

$$D = \binom{N+1}{2} - 1 = \frac{N^2 + N}{2} - 1 < K(N - K).$$

Since $K < N$ we have that D will never be less than $K(N - K)$ and consequently the embedding will never be isometric and dimensionally reductive by our definition. However, it is important to notice that this mapping does not rely on the number of points we start with.

One might consider mapping into $\mathbb{R}^D$ where D is defined as in Theorem 5.3.1 and then applying the JLL to produce a nearly isometric, dimensionally reductive map. In order for this composition of maps to be considered dimensionally reductive the dimension resulting from the JLL would need to be less than the dimension of the original $Gr(N, K)$, i.e.

$$4(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})^{-1} \ln(M) < K(N - K) < D = \binom{N+1}{2} - 1$$

for distortion factor ϵ and M data points on $Gr(N, K)$.

From the neccessity that the above properties hold to be dimensionally reductive an expression for the number of points that could yield a dimensionally reductive map for a fixed distortion factor ϵ can be found. Requiring that the above property hold induces the following series of inequalities

$$\begin{aligned} & 4(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})^{-1} \ln(M) < K(N-K) \\ \implies & \ln(M) < \frac{K(N-K)}{4}(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}) \\ \implies & M < \exp(\frac{K(N-K)}{4}(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})). \end{aligned}$$

Moreover, $M > 2$ for all problems of interest so we have

$$\begin{aligned} & ln(2) < \ln(M) < \frac{K(N-K)}{4}(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}) \\ \implies & \ln(2) < \exp(\frac{K(N-K)}{4}(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})) \\ \implies & \frac{4\ln(2)}{K(N-K)} < \frac{\epsilon^2}{2} - \frac{\epsilon^3}{3} \\ \implies & \frac{24\ln(2)}{K(N-K)} < \epsilon^2 - \epsilon^3. \end{aligned}$$

Recalling that $0 < \epsilon < 1$ we have $\epsilon^2 - \epsilon^3 < \epsilon$ and we can produce a loose lower bound on the distortion factor needed to obtain a nearly isometric, dimensionally reductive map by

$$\frac{24\ln(2)}{K(N-K)} < \epsilon.$$

Frequently, $1/2$ is considered a threshold for a reasonable distortion factor. Taking this into consideration, together with the fact that $24\ln(2) \approx 16$ we find that dimension reduction is reasonably feasible if $\frac{16}{K(N-K)} < 1/2$. Finally, we find that for more than 2 points from $Gr(N, K)$ and an isometry constant less than 1/2 we need, approximately, $8 < K(N-K)$.

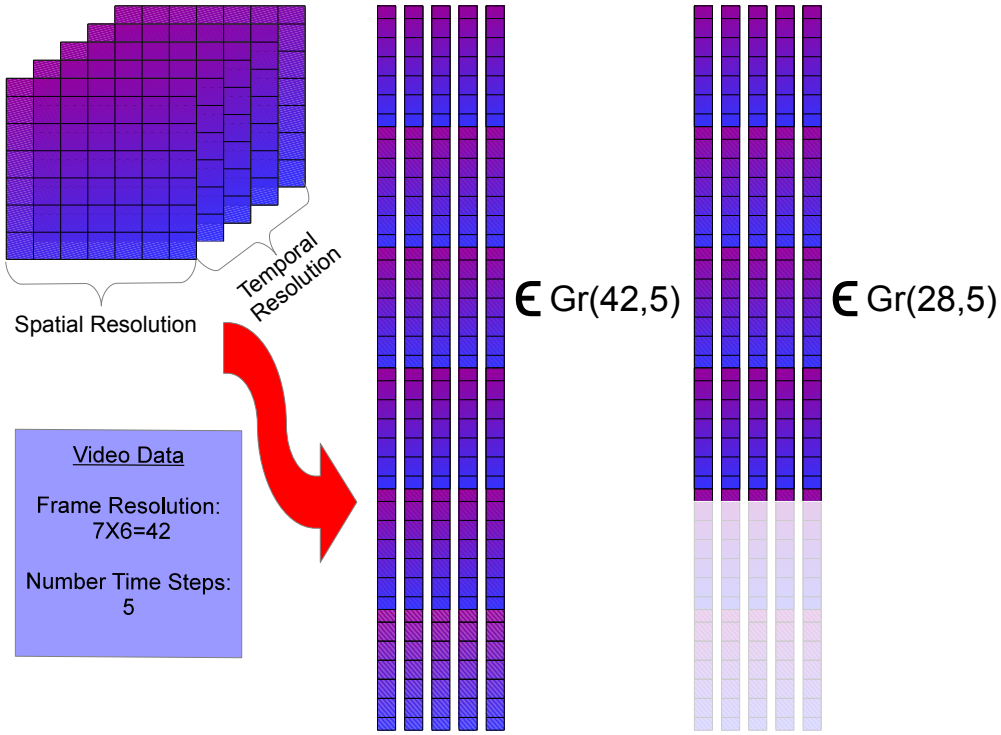


Figure 5.4.1: Schematic illustrating the effect of a dimension reducing map between Grassmannians where N is being reduced and K stays constant, $f : Gr(N, K) \to Gr(n, K)$, $n < N$.

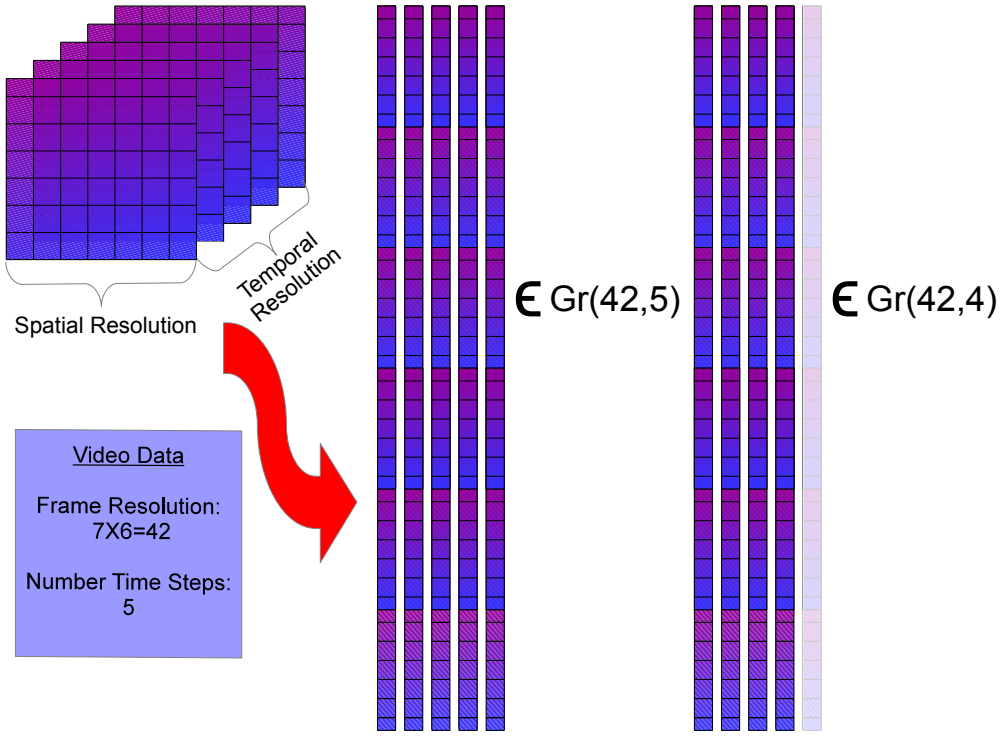


Figure 5.4.2: Schematic illustrating the effect of a dimension reducing map between Grassmannians where K is being reduced and N stays constant, $f : Gr(N, K) \to Gr(N, k)$, $k < K$.

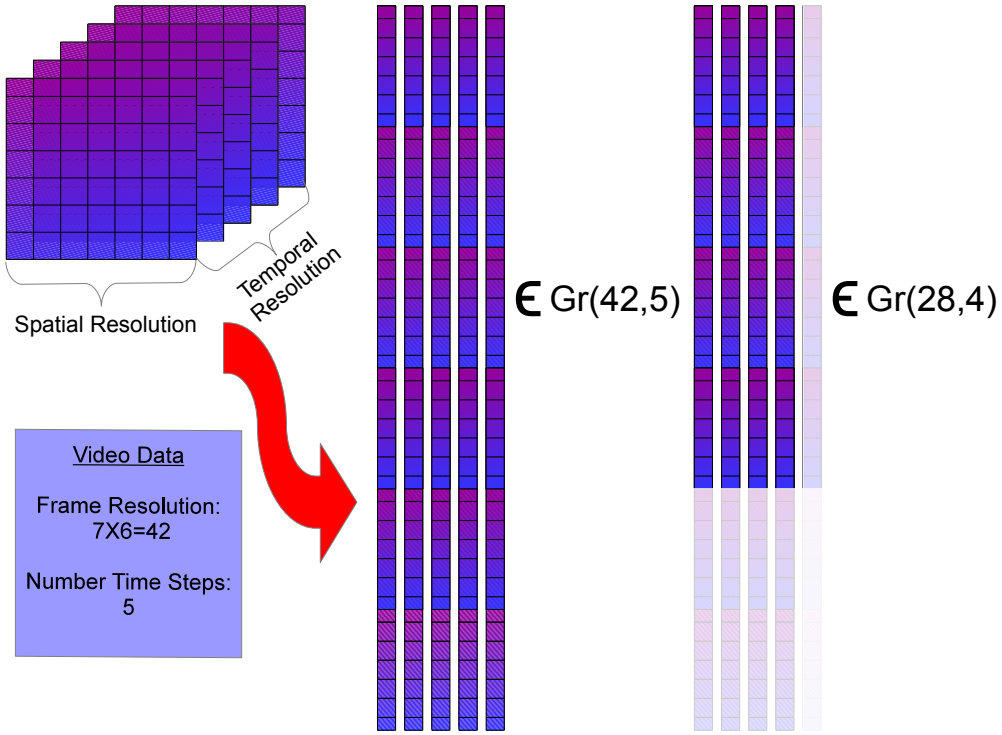


Figure 5.4.3: Schematic illustrating the effect of a dimension reducing map between Grassmannians where both N and K are reduced, $f : Gr(N, K) \to Gr(n, k)$, $n < N$ and $k < K$.

5.4 Johnson-Lindenstrauss for Grassmannians

Considered here are nearly isometric mappings between Grassmannians characterized by different dimensions. Unlike the previous section these mappings map directly from $Gr(N, K)$ into $Gr(N, k)$, $Gr(n, K)$, or $Gr(n, k)$ with $n < N$ and $k < K$. One way to interpret the two characterizing dimensions of a Grassmannian manifold is as two different resolutions. For example, N could be considered a spatial resolution and K could be a temporal resolution. Successfully mapping to a Grassmannian where at least one of N and K has been reduced while preserving distances would suggest that particular spatial (or temporal) dimensions are extraneous. Figures 5.4.1, 5.4.2, and 5.4.3 provide sample schematics for the case where spatial resolution, temporal resolution, or both resolutions are reduced, respectively. Although the schematics truncate the last dimensions of each type, this is just an example and the extraneous dimensions may be distributed throughout.

5.4 Statement of the Conjecture

Conjecture 5.4.1 (Grassmannian Johnson-Lindenstrauss)**.** Let $0 < \epsilon < 1$ and $\mathcal{V}$ be a set of m points on $Gr(N, K)$, $\mathcal{V} = \mathbf{V}_1, \ldots, \mathbf{V}_m$ such that $\mathbf{V}_i \in Gr(N, K)$. Assume that $\langle \mathbf{V}_i \rangle \cap \langle \mathbf{V}_j \rangle = \emptyset$ for all i and j. Furthermore, we denote the chordal distance between two points $\mathbf{V}_i$ and $\mathbf{V}_j$ by $d_c(\mathbf{V}_i, \mathbf{V}_j)$.

(a) There exists a map $f : Gr(N, K) \to Gr(n, K)$ with $n < N$ such that

$$(1 - \epsilon) d_c(\mathbf{V}_i, \mathbf{V}_j) \leq d_c(f(\mathbf{V}_i), f(\mathbf{V}_j)) \leq d_c(\mathbf{V}_i, \mathbf{V}_j)(1 + \epsilon)$$

for all i, j whenever $n > s(N, K, \epsilon, m)$ for some function s.

(b) There exists a map $g : Gr(N, K) \to Gr(N, k)$ with $k < K$ such that

$$(1 - \epsilon) d_c(\mathbf{V}_i, \mathbf{V}_j) \leq d_c(g(\mathbf{V}_i), g(\mathbf{V}_j)) \leq d_c(\mathbf{V}_i, \mathbf{V}_j)(1 + \epsilon)$$

for all i, j whenever $k > t(N, K, \epsilon, m)$ for some function t.

(c) There exists a map $h : Gr(N, K) \to Gr(n, k)$ with $n < N$ and $k < K$ such that

$$(1 - \epsilon) d_c(\mathbf{V}_i, \mathbf{V}_j) \leq d_c(h(\mathbf{V}_i), h(\mathbf{V}_j)) \leq d_c(\mathbf{V}_i, \mathbf{V}_j)(1 + \epsilon)$$

for all i, j whenever $n > w(N, K, \epsilon, m)$ and $k > z(N, K, \epsilon, m)$ for some pair of functions w, z.

5.4 Towards Existence: Statistical Approach

Our first course of action was to recreate the statistically based proof for the standard Johnson-Lindenstrauss Lemma. In order to take this approach a few key statistical and probabilistic statements/tools used in the approach would need to be made in terms of the Grassmannian as the domain of the distributions. This section is comprised of many of these statements. Some are

proved and many are presented as conjectures with or without ideas about how they might be proved or substeps that would be needed. Also presented are some of the challenges that arise as a consequence of the domain space no longer being Euclidean.

Theorem 5.4.1. *The following distributions are equivalent:*

(a) The distribution of the Frobenius norm of a fixed $K-$plane in $\mathbb{R}^N$ projected onto a uniform random (with respect to Haar measure) $n-$dimensional subspace of $\mathbb{R}^N$.

(b) The distribution of the Frobenius norm of a uniform random (with respect to Haar measure) $K-$plane in $\mathbb{R}^N$ projected onto a fixed $n-$dimensional subspace of $\mathbb{R}^N$.

Proof. Let the fixed $K-$dimensional subspace of $\mathbb{R}^N$ having an orthonormal basis making up the columns of the matrix $\mathbf{U}$. Without loss of generality, let the fixed subspace under consideration be the first K standard basis vector of of $\mathbb{R}^N$,

$$\begin{aligned}\mathbf{U} &= \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \dots & \boldsymbol{e}_K \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_K \\ \mathbf{0}_{N-K,K} \end{bmatrix}.\end{aligned}$$

Let $\mathbf{V}$ be a basis for a fixed $n-$dimensional subspace of $\mathbb{R}^N$. Without loss of generality, we let $\mathbf{V}$ be the subspace spanned by the first n standard basis vectors in $\mathbb{R}^N$,

$$\begin{aligned}\mathbf{V} &= \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \dots & \boldsymbol{e}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_n \\ \mathbf{0}_{N-n,n} \end{bmatrix}.\end{aligned}$$

Let $\mathbf{A}$ be a random element of the orthogonal group, $\mathbf{A} \in O(N)$. A uniform random (with respect to the Haar measure) $K-$dimensional subspace of $\mathbb{R}^N$ can be generated as $\mathbf{U}_0 = \mathbf{A}^T\mathbf{U}$. Similarly, a uniform random (with respect to the Haar measure) $n-$dimensional subspace can be generated

as $\mathbf{V}_0 = \mathbf{AV}$. Recall, the projection onto the column space of a matrix $\boldsymbol{B}$ can be written as $\mathbf{P}_B = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T$. Consider the Frobenius norm of a fixed $K-$ dimensional subspace projected onto the random $n-$dimensional subspace $\mathbf{V}_0$. We yield the following sequence of equalities

$$\begin{aligned}
||\mathbf{P}_{V_0}\mathbf{U}||_F^2 &= tr((\mathbf{P}_{V_0}\mathbf{U})^T(\mathbf{P}_{V_0}\mathbf{U})) \\
&= tr(\mathbf{U}^T\mathbf{P}_{V_0}^T\mathbf{P}_{V_0}\mathbf{U}) = tr(\mathbf{U}^T\mathbf{P}_{V_0}\mathbf{U}) = tr(\mathbf{U}^T(\mathbf{V}_0(\mathbf{V}_0^T\mathbf{V}_0)^{-1}\mathbf{V}_0^T)\mathbf{U}) \\
&= tr(\mathbf{U}^T(\mathbf{AV}((\mathbf{AV})^T(\mathbf{AV}))^{-1}(\mathbf{AV})^T)\mathbf{U}) \\
&= tr(\mathbf{U}^T(\mathbf{AV}(\mathbf{V}^T\mathbf{A}^T\mathbf{AV})^{-1}\mathbf{V}^T\mathbf{A}^T\mathbf{U}).
\end{aligned}$$

From the assumption that $\mathbf{A} \in O(N)$ we have that $\mathbf{AA}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}_N$. Using this fact we have

$$\begin{aligned}
||\mathbf{P}_{V_0}\mathbf{U}||_F^2 &= tr(\mathbf{U}^T(\mathbf{AV}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T\mathbf{A}^T)\mathbf{U}) \\
&= tr(\mathbf{U}^T\mathbf{AP}_V\mathbf{A}^T\mathbf{U}) = tr(\mathbf{U}^T\mathbf{AP}_V^T\mathbf{P}_V\mathbf{A}^T\mathbf{U}) = tr((\mathbf{A}^T\mathbf{U})^T\mathbf{P}_V^T\mathbf{P}_V(\mathbf{A}^T\mathbf{U}))) \\
&= tr(\mathbf{U}_0^T\mathbf{P}_V^T\mathbf{P}_V\mathbf{U}_0) \\
&= tr((\mathbf{P}_V\mathbf{U}_0)^T(\mathbf{P}_V\mathbf{U}_0)) \\
&= ||\mathbf{P}_V\mathbf{U}_0||_F^2.
\end{aligned}$$

The value $||\mathbf{P}_V\mathbf{U}_0||_F^2$ is the Frobenius norm of the projection of the uniform random $K-$dimensional subspace $\mathbf{U}_0$ onto the fixed $n-$dimensional subspace spanned by $\mathbf{V}$. Thus, we have that the distributions of the Frobenius norms are equivalent. □

Conjecture 5.4.2. The following distributions are equivalent:

(a) The distribution of the Frobenius norm of a fixed $K-$plane in $\mathbb{R}^N$ projected onto a uniform random (with respect to Haar measure) $k-$dimensional subspace of $\mathbb{R}^K$.

(b) The distribution of the Frobenius norm of a uniform random (with respect to Haar measure) $K-$plane in $\mathbb{R}^N$ projected onto a fixed $k-$dimensional subspace of $\mathbb{R}^K$.

Proof. Would this become a statement about frames or random sets? Some pieces that I think would need to be shown are stated in the following conjecture and theorem. □

Theorem 5.4.2 (Block Projection Matrix)**.** *Let M be an integer. If $\mathbf{P}_U$ is an $m \times m$ projection matrix, then the matrix*

$$\begin{bmatrix} \mathbf{P}_U & \mathbf{0}_{m,M-m} \\ \mathbf{0}_{M-m,m} & \mathbf{0}_{M-m} \end{bmatrix}$$

is also a projection matrix.

Proof. Recall that in order for the matrix $\boldsymbol{M}$ to be a projection matrix it must satisfy $\boldsymbol{MM} = \boldsymbol{M}$. Let

$$\boldsymbol{M} = \begin{bmatrix} \mathbf{P}_U & \mathbf{0}_{m,M-m} \\ \mathbf{0}_{M-m,m} & \mathbf{0}_{M-m} \end{bmatrix}.$$

We then have

$$\begin{aligned} \boldsymbol{MM} &= \begin{bmatrix} \mathbf{P}_U & \mathbf{0}_{m,M-m} \\ \mathbf{0}_{M-m,m} & \mathbf{0}_{M-m} \end{bmatrix} \begin{bmatrix} \mathbf{P}_U & \mathbf{0}_{m,M-m} \\ \mathbf{0}_{M-m,m} & \mathbf{0}_{M-m} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{P}_U\mathbf{P}_U & \mathbf{0}_{m,M-m} \\ \mathbf{0}_{M-m,m} & \mathbf{0}_{M-m} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{P}_U & \mathbf{0}_{m,M-m} \\ \mathbf{0}_{M-m,m} & \mathbf{0}_{M-m} \end{bmatrix} \\ &= \boldsymbol{M}. \end{aligned}$$

Thus, the defined block matrix is also a projection matrix. □

Conjecture 5.4.3. Let $\boldsymbol{W}$ be a uniform random (with respect to Haar measure) $k-$dimensional subspace of $\mathbb{R}^K$ and let $\mathbf{P}_W \in \mathbb{R}^{K\times K}$ be the projection onto the space spanned by $\langle \boldsymbol{W} \rangle$. Furthermore, let $\mathbf{V}$ be an $N \times K$ matrix. Right multiplication of $\mathbf{V}$ by

$$\tilde{\mathbf{P}}_W = \begin{bmatrix} \mathbf{P}_W & \mathbf{0}_{K,N-K} \\ \mathbf{0}_{N-K,K} & \mathbf{0}_{N-K} \end{bmatrix}$$

is equivalent to a projection onto a random subset of its columns.

In the JLL proof in [7] they are able to just consider the length of the projection of a fixed vector in $\mathbb{R}^N$ because the difference between any two points in $\mathbb{R}^N$ can be represented by a single vector corresponding to their difference vector. Also, because the mapping f is linear (a projection) they have that if $\boldsymbol{w} = \boldsymbol{v} - \boldsymbol{u}$ then $f(\boldsymbol{w}) = f(\boldsymbol{v} - \boldsymbol{u}) = f(\boldsymbol{v}) - f(\boldsymbol{u})$. The analagous statement does not hold when we represent two fixed points on $Gr(N, K)$ by their projection matrices. Explicitly, let $\mathbf{V}$ and $\mathbf{U}$ be fixed points on $Gr(N, K)$, $\mathbf{V},\ \mathbf{U} \in \mathbb{R}^{N \times K}$. Clearly we have that $\boldsymbol{W} = \mathbf{V} - \mathbf{U} \in \mathbb{R}^{N \times K}$ and that $\boldsymbol{W} \in Gr(N, K)$. However, consider the projection matrices $\mathbf{P}_V = \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T$ and $\mathbf{P}_U = \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T$. Now consider the projection matrix onto $\boldsymbol{W}$. By definition we have that

$$\begin{aligned}
\mathbf{P}_W &= \boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{W}^T \\
&= (\mathbf{V} - \mathbf{U})((\mathbf{V} - \mathbf{U})^T(\mathbf{V} - \mathbf{U}))^{-1}(\mathbf{V} - \mathbf{U})^T \\
&= (\mathbf{V} - \mathbf{U})((\mathbf{V}^T - \mathbf{U}^T)(\mathbf{V} - \mathbf{U}))^{-1}(\mathbf{V}^T - \mathbf{U}^T) \\
&= (\mathbf{V} - \mathbf{U})(\mathbf{V}^T\mathbf{V} - \mathbf{V}^T\mathbf{U} - \mathbf{U}^T\mathbf{V} + \mathbf{U}^T\mathbf{U})^{-1}(\mathbf{V}^T - \mathbf{U}^T) \\
&\neq \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T - \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T \\
\implies \mathbf{P}_W &\neq \mathbf{P}_V - \mathbf{P}_U.
\end{aligned}$$

Consequently, we cannot recreate the Euclidean JLL proof in [7] in exactly the same manner.

Need to remember that if $\mathbf{U}$ is a random subspace and $\mathbf{P}_U$ is the projection onto the subspace spanned by $\langle \mathbf{U} \rangle$ and $\mathbf{V}_i, \mathbf{V}_j$ are points on $Gr(N, K)$, then $\mathbf{P}_U(\mathbf{V}_i - \mathbf{V}_j) = \mathbf{P}_U\mathbf{V}_i - \mathbf{P}_U\mathbf{V}_j$. If we want to use the fact that we can compute the chordal distance as the Frobenius norm of the difference of projection matrix representations, then we are looking for a bound relating $||\mathbf{P}_{V_i} - \mathbf{P}_{V_j}||_F^2$ and $||\mathbf{P}_{\tilde{V}_i} - \mathbf{P}_{\tilde{V}_j}||_F^2$ with $\tilde{\mathbf{V}}_i = \mathbf{P}_U\mathbf{V}_i$ and $\tilde{\mathbf{V}}_j = \mathbf{P}_U\mathbf{V}_j$.

Another challenge is that in general $||\mathbf{P}_{V_i} - \mathbf{P}_{V_j}||_F^2 \neq ||\mathbf{P}_{U_i} - \mathbf{P}_{U_j}||_F^2$ where $\mathbf{U}_i$ is the result of projecting $\mathbf{V}_i$ onto some space. Given the challenges related to handling the projection representation of points on $Gr(N, K)$, we are led to consider an alternative approach.

5.4 Towards Existence: A Packing Problem

There are two common ways of representing a configuration of points on the Grassmannian. Let $\mathcal{V} = \{\mathbf{V}_1, \mathbf{V}_2, ...\mathbf{V}_m\}$ be a set of m points on $Gr(N, K)$, $\mathbf{V}_i \in Gr(N, K)$ for all i. The convention used by [68] and [65], as well as others, is to associate each point with its orthogonal projection matrix $\mathbf{P}_{V_i}$. As discussed in the previous section, this is often done because $d_c^2(\mathbf{V}_i, \mathbf{V}_j) = ||\mathbf{P}_{V_i} - \mathbf{P}_{V_j}||_F^2$ resulting in easy distance computations. Alternatively, in [8] a configuration can be represented by concatenating the subspaces $\bar{\mathbf{V}} = [\mathbf{V}_1\, \mathbf{V}_2\, \ldots\, \mathbf{V}_m]^T$. Looking at the Grammian $\boldsymbol{G} = \bar{\mathbf{V}}\bar{\mathbf{V}}^T$ produces a block matrix from which the off diagonal elements can readily produce the principal angles, and consequently pairwise distances, between the subspaces.

As a result of the previously mentioned challenges of handling the projection representations of the points, an alternative approach would be to look for properties of the SVD of the off diagonal block matrices. This is similar to the approach taken in [8]. In [8] the problem of interest is that of packing subspaces on a Grassmann manifold. Consider the set $\mathcal{V} = \{\mathbf{V}_1, \mathbf{V}_2, ...\mathbf{V}_m\}$ of m points on $Gr(N, K)$, $\mathbf{V}_i \in Gr(N, K)$ for all i. The packing diameter of a configuration of the points is given by $\text{pack}_{\text{chord}}(\mathcal{V}) = \min_{i \neq j} d_C(\mathbf{V}_i, \mathbf{V}_j)$. The generic packing problem on the Grassmannian seeks to identify a configuration of m points that maximizes the packing diameter of the set of points. Determining an optimal solution to a max-min problem can be far from trivial. An alternative approach is to identify if there is a feasible configuration where the packing diameter is greater than or equal to a dummy variable and then increment the dummy variable.

It is reasonable to assume that the scenario in which the minimum pairwise distance is being maximized is highly connected to the setting where pairwise distances might be strongly distorted. For this reason, let us look more closely at the Grammian matrix produced by concatenation, let $\bar{\mathbf{V}} = [\mathbf{V}_1\, \mathbf{V}_2\, \ldots\, \mathbf{V}_m]^T$ and define the Grammian to be $\boldsymbol{G} = \bar{\mathbf{V}}\bar{\mathbf{V}}^T$. Again, $\boldsymbol{G}$ is a block matrix

where $\boldsymbol{G}_{ij} = \mathbf{V}_i^T \mathbf{V}_j$. The dimensions of $\boldsymbol{G}$ are $mN \times mN$ where you can think of it as an $m \times m$ block matrix with blocks of size $N \times N$. Each of the blocks has rank K. In [8] they recall from linear algebra that a positive semidefinite block matrix, with identity matrices as the diagonal blocks can be factored into a particularly useful form via eigenvalue decomposition. By construction the matrix $\boldsymbol{G}$ is positive semidefinite and can be factored as $\boldsymbol{X}^T \boldsymbol{X}$ where $\boldsymbol{X}$ is a $K \times mN$ configuration matrix. That is, the columns of $\boldsymbol{X}$ form orthogonal bases for m different K dimensional subspaces of $\mathbb{R}^N$. Given that the SVD of the off diagonal blocks of $\boldsymbol{G}$ produce the principal angles, we can also generate the chordal distances between points from these off diagonal blocks.

Returning to the packing problem, instead of trying to solve the max-min problem directly, consider $\rho > 0$ and repose the problem as looking for a configuration of subspaces, $\bar{\mathbf{V}}$, on $Gr(N, K)$ satisfying that $\text{pack}_{\text{chord}}(\bar{\mathbf{V}}) \geq \rho$. A configuration satisfying this is said to be feasible. The value of ρ is incremented until a feasible configuration cannot be found. The paper [8] presents an algorithm for finding a feasible configuration for a given ρ and we refer the interested reader to the paper for exact details. We now begin a discussion about ways to modify this problem formulation to help move in the direction of proving the existence of a nearly isometric map between Grassmannians. For the remainder of our discussion we will denote the packing distance of a configuration of points $\bar{\mathbf{V}}$ on $Gr(N, K)$ by $\text{pack}_{N,K}(\bar{\mathbf{V}})$ (or $\delta_{N,K}$ when convenience dictates) and assume that the packing distance is computed with respect to the chordal distance.

For the first case let us consider the case where we fix N and reduce the dimension of the subspace, K.

Theorem 5.4.3 (Nearly Distance Preserving and Packing Distance)**.** *Let $\bar{\mathbf{V}}$ be a configuration of points on $Gr(N, K)$. Consider a map $f : Gr(N, K) \to Gr(n, k)$ satisfying*

1. $|\textit{pack}_{N,K}(\bar{\mathbf{V}}) - \textit{pack}_{n,k}(f(\bar{\mathbf{V}}))| \leq \epsilon_1/3$,
2. $\max_{i \neq j}\{|\textit{pack}_{N,K}(\bar{\mathbf{V}}) - d_c(\mathbf{V}_i, \mathbf{V}_j)|\} \leq \epsilon_2/3$, *and*
3. $\max_{i \neq j}\{|\textit{pack}_{n,k}(f(\bar{\mathbf{V}})) - d_c(f(\mathbf{V}_i), f(\mathbf{V}_j))|\} \leq \epsilon_3/3$.

If $\epsilon = \max\{\epsilon_1, \epsilon_2, \epsilon_3\}$*,* $pack_{N,K}(\bar{\mathbf{V}})) \geq 1$*, and* $(\epsilon)(pack_{N,K}(\bar{\mathbf{V}})) < 1$*, then this map is nearly distance preserving for* ϵ*.*

Proof. Let $\bar{\mathbf{V}}$ be a configuration of points on $Gr(N, K)$. Consider a map $f : Gr(N, K) \to Gr(n, k)$ satisfying

$$3|\text{pack}_{N,K}(\bar{\mathbf{V}}) - \text{pack}_{n,k}(f(\bar{\mathbf{V}}))| \leq \epsilon_1.$$

Additionally, let

$$3(\max_{i \neq j}\{|\text{pack}_{N,K}(\bar{\mathbf{V}}) - d_c(\mathbf{V}_i, \mathbf{V}_j)|\}) \leq \epsilon_2$$

and

$$3(\max_{i \neq j}\{|\text{pack}_{n,k}(f(\bar{\mathbf{V}})) - d_c(f(\mathbf{V}_i), f(\mathbf{V}_j))|\}) \leq \epsilon_3.$$

We further let $\epsilon = \max\{\epsilon_1, \epsilon_2, \epsilon_3\}$. Lastly, let $\delta_{N,K} = \text{pack}_{N,K}(\bar{\mathbf{V}})$ and $\delta_{n,k} = \text{pack}_{n,k}(f(\bar{\mathbf{V}}))$. Assume that $\delta_{N,K} \geq 1$ and that $\epsilon\delta_{N,K} < 1$. We then have

$$\begin{aligned}
|d_c(f(\mathbf{V}_i), f(\mathbf{V}_j)) - d_c(\mathbf{V}_i, \mathbf{V}_j)| &= |d_c(f(\mathbf{V}_i), f(\mathbf{V}_j)) - \delta_{n,k} + \delta_{n,k} - \delta_{N,K} + \delta_{N,K} - d_c(\mathbf{V}_i, \mathbf{V}_j)| \\
&\leq |d_c(f(\mathbf{V}_i), f(\mathbf{V}_j)) - \delta_{n,k}| + |\delta_{n,k} - \delta_{N,K}| + |\delta_{N,K} - d_c(\mathbf{V}_i, \mathbf{V}_j)| \\
&\leq \epsilon_2/3 + \epsilon_1/3 + \epsilon_3/3 \leq \epsilon
\end{aligned}$$

for all $i \neq j$. Since $\delta_{N,K} \geq 1$ we have

$$|d_c(f(\mathbf{V}_i), f(\mathbf{V}_j)) - d_c(\mathbf{V}_i, \mathbf{V}_j)| \leq \epsilon\delta_{N,K}.$$

From the fact that $\delta_{N,K} \leq d_c(\mathbf{V}_i, \mathbf{V}_j)$for all $i \neq j$ we have $\epsilon\delta_{N,K} \leq \epsilon d_C(\mathbf{V}_i, \mathbf{V}_j)$ for all $i \neq j$. Consequently,

$$\begin{aligned}
|d_c(f(\mathbf{V}_i), f(\mathbf{V}_j)) - d_c(\mathbf{V}_i, \mathbf{V}_j)| &\leq \epsilon d_c(\mathbf{V}_i, \mathbf{V}_j) \\
-\epsilon d_c(\mathbf{V}_i, \mathbf{V}_j) \leq d_c(f(\mathbf{V}_i), f(\mathbf{V}_j)) - d_c(\mathbf{V}_i, \mathbf{V}_j) &\leq \epsilon d_c(\mathbf{V}_i, \mathbf{V}_j) \\
(1 - \epsilon)d_c(\mathbf{V}_i, \mathbf{V}_j) \leq d_c(f(\mathbf{V}_i), f(\mathbf{V}_j)) &\leq d_c(\mathbf{V}_i, \mathbf{V}_j)(1 + \epsilon).
\end{aligned}$$

Thus, the map is nearly distance preserving for ϵ $\square$

This proof goes through whether $n < N$ and K is fixed, $k < K$ and N is fixed, or if $n < N$ and $k < K$. Undoubtably the ϵ's in the proofs will depend on these dimensions. We will use $\delta_{N,K}$ to denote the packing distance of the original configuration $\bar{\mathbf{V}}$ and $\delta_{n,k}$ to be the packing distance of the configuration under the map f. How is this connected to the packing problem? For a particular configuration on $Gr(N, K)$ we can readily compute the packing distance. Thus for a collection of points we can determine whether or not the second property is satisfied. We are then looking for a configuration of points in the image space whose packing distance is bounded based on property one. Finally, the third property may be able to be interpreted as a spectral condition (a function of the singular values). Through formalization of these ideas as added constraints, we may be able to use the algorithm in [8] to search for a feasible configuration satisfying these conditions (as well as a couple others).

If we can find a feasible configuration then we, by construction, have proven the existence of a nearly distance preserving map. We need to consider the answers to a few different questions. First, what do the assumptions mean and can they be simplified? Second, what are the conditions that result from the assumptions? Third, what are the conditions arising from the image space of the theorized mapping? Once these three questions are answered, the formulation of a feasible configuration based optimization problem can be completed.

Notably, there are several non trivial assumptions being made to make the proof go through. Is there a way to simply the assumptions used? Starting from perhaps the most unlikely of the assumptions to be satisfied: $\epsilon\delta_{N,K} < 1$. Equivalently, $\epsilon \leq 1/\delta_{N,K}$. Since we take ϵ to be equal to the maximum of $\{\epsilon_1, \epsilon_2, \epsilon_3\}$ the following three properties must be satisfied

1. $|\delta_{n,k} - \delta_{N,K}| = |\delta_{N,K} - \delta_{n,k}| \leq \epsilon_1/3 \leq \epsilon/3 \leq 1/3\delta_{N,K}$,
2. $\max\limits_{i \neq j}\{|d_c(\mathbf{V}_i, \mathbf{V}_j) - \delta_{N,K}|\} = \max\limits_{i \neq j}\{|\delta_{N,K} - d_c(\mathbf{V}_i, \mathbf{V}_j)|\} \leq \epsilon_2/3 \leq \epsilon/3 \leq 1/3\delta_{N,K}$, and
3. $\max\limits_{i \neq j}\{|\delta_{n,k} - d_c(f(\mathbf{V}_i), f(\mathbf{V}_j))|\} \leq \epsilon_3/3 \leq \epsilon/3 \leq 1/3\delta_{N,K}$.

In some sense, the first property might be the most intuitively connected to being distance preserving. You are considering the smallest pairwise distance in the original space and comparing it to the smallest pairwise distance in the image space. However, what is different about it is that in distance preserving maps we are comparing the distances between two points in the domain to the distance between the images of the same points. Without adding additional requirements to the function f there is no reason to assume that the packing diameter in the image space is attained by the images of the points achieving the packing diameter in the original space.

Unlike the first property, the second and third properties pertain to distances in one space (domain or image) only. In fact, both properties can be thought of as a deviation or variation of pairwise distances relative to the minimim pairwise distance in that space. This seems to suggest a constraint involving the variance of the distance matrix for the configuration we seek in the image space. The deviation constraint will either be satisfied or it won't, for a given epsilon, for the starting configuration. If it isn't satisfied then there will be no optimization to be performed using the combination of our proposed approach and that of Strohmer.

Closely tied to the first assumption is the assumption that $\delta_{N,K} \geq 1$. What does this mean? The minimum chordal distance between any two points in the initial configuration must be greater than or equal to one. This immediately yields that, for our construction, $\epsilon_1, \epsilon_2,$ and ϵ_3 are bounded above by 1/3. This simplification makes the movement towards formalization of an optimization problem much simpler. The computational results shown in the next section give some sense of how realistic this is for random samplings of Grassmannians and using canonical projections as our function mapping between the Grassmannians.

If using some combination of approaches a suitable configuration of points can be found in the image space, it is shows the existence of a map f with the property we seek. There some properties that f might have that would be ideal. Particularly, if f were a canonical projection or could be a random projection. The following section will present the results of several numerical experiments exploring how distances change under canonical projections. These experiments were performed to see if we can reveal characteristics that might push us further towards the existence.

5.4 Computational Results

In [7] they generate a random point on the unit sphere and then use canonical projections to go through their proof. As we were initially inspired by their approach, we conducted a similar experiment. We began by generating a uniform random (with respect to Haar measure) set of points on $Gr(N, K)$. In our experiments we choose $N \times K \times 5$ points on $Gr(N, K)$. These points are generated as described in the background section using the procedure described in [65]. That is to say that we generate a random matrix who's entries are drawn from the standard normal distribution and then perform a QR decomposition and use the orthogonal basis produced to represent a single point on the Grassmannian.

Once the set of points has been generated we consider three scenarios:

1. fix the ambient dimension (N) and reduce the subspace dimension (K),
2. fix the subspace dimension and reduce the ambient dimension, or
3. reduce both the subspace dimension and the ambient dimension.

Dimension reduction is achieved by projecting onto the first subset of the columns to reduce the subspace dimenion or the first subset of rows to reduce the ambient dimension. As the dimension (either N or K) is reduced, we generate an orthonormal basis for the point in the new space. On each new Grassmannian we compute several values. For all the experiments performed distances are measured using the chordal distance. Under each projection we compute several values. Included in this set of values are the largest and smallest principal angles between each pair of points, the packing distance of the new configuration, the maximum deviation from the packing distance for each configuration relative to its packing distance, and also the difference between the current configuration's packing distance and the packing distance of the original configuration.

Figures 5.4.4, 5.4.5, 5.4.6, 5.4.8, 5.4.9, 5.4.7, 5.4.10, 5.4.11, and 5.4.6 contain plots for an initial configuration consisting of 1000 uniform random points on $Gr(20, 10)$. In Figures 5.4.4, 5.4.5, 5.4.6, 5.4.8, 5.4.9, and 5.4.7 only one dimension is changed at a time. For these figures the top subfigure shows the effect of varying the subspace dimension (K) while the bottom subfigures

show the impact of varying the ambient dimension (N). The surfaces shown in Figures 5.4.10, 5.4.11, and 5.4.6 demonstrate the effect of reducing both dimensions.

The distribution of the largest principal angle between pairs of points as the dimensions are reduced are shown in Figures 5.4.8 while the distribution of the smallest principal angle is shown in Figure 5.4.9. For each pair of points (i.e. pair of subspaces) the largest or smallest principal angle is recorded. The set of angles is then histogrammed and normalized based on the number of unique pairs of points. Principal angles take on values in $[0, \pi/2)$ and this interval was broken into 10 evenly spaced bins and the recorded angles were histogrammed accordingly. Along the horizontal axis the angle size increases from left to right. The vertical axis indicates the dimension being changed and it decreases from $N-1$ (or $K-1$) to $K+1$ (or 1) from bottom to top. The value of the pixel in the image indicates the percentage of the principal angles of interest that fell in the range of values indicated along the $x-$axis.

The distributions of the largest principal angle between subspaces also appears to present a trend. The peak of the distribution shifts left (towards a smaller angle), the peak takes on a smaller value, and the width of the distribution increases as either the subspace or the ambient dimension is reduced. As the dimensions are reduced the distribution of the largest angle appears to tend towards uniformity. Alternatively, the distributions of the smallest principal angle change very differently from the those of the largest and the trends are different as the subspace or ambient dimension are changed. Whether the subspace or ambient dimension is reduced, the width of the distribution appears to be constant as does the value of the peak. Unlike the largest principal angle, the peak of the distribution of smallest angles moves to the left (smaller angle) as the ambient dimension is reduced but to the right (a larger angle) as the subspace dimension is reduced.

The distribution of the largest principal angle has been studied in [71] but to our knowledge there has been no formal exploration as to how these distributions change under dimension reduction. All of these results reflect what someone familiar with, or with a developed intuition, would expect as the dimension is changed. As the subspace dimension is reduced via canonical projection, columns are being removed and the points which we at first a uniform sampling of $Gr(N, K)$

are now a uniform sampling of $Gr(N,k)$. The expectation is that random subspaces will be orthogonal in at least one dimension assuming that the ambient dimension is sufficiently large. When the ambient dimension is reduced there is less space for the subspaces to be distributed over and as the subspaces get closer together one expects overlap and consequently a smaller smallest angle between the spaces.

In Figure 5.4.7 we see the percentage of the pairwise distances that are preserved for isometry values ranging from 0 to 1. Along the horizontal axis is the isometry value (ϵ in the JLL). The vertical axis indicates the dimension being changed and it decreases from $N-1$ (or $K-1$) to $K+1$ (or 1) from bottom to top. The value of the pixel in the image indicates the percentage of pairwise distances that were preserved up to the isometry constant relative to the distances of the original configuration. A bright yellow value indicates that 100% of the pairwise distances were preserved. Notice that as one dimension is changed, the majority of times nearly all the distances are preserved.

A horizontal red line was added to Figures 5.4.5 and 5.4.6 to indicate the largest upper bound for the corresponding criterion in Theorem 5.4.3. In both cases this value is 1/3. Figure 5.4.5 shows how $\max_{i\neq j}\{|d_c(f(\mathbf{V}_i),f(\mathbf{V}_j))-\delta_{n,k}$ changes as either the subspace or ambient dimension is reduced. The difference between the packing distance of the configuration in the reduced dimension space and the original packing distance, $|\delta_{n,k}-\delta_{N,K}|$, is shown in Figure 5.4.6. Actual packing distances are shown in Figure 5.4.4.

We see that the packing distance changes in a nonincreasing manner as either the subspace or ambient dimension are reduced. Notice that they also appear to decrease at a similar rate. The behavior seen in Figure 5.4.4 is seen again in Figure 5.4.6. Also of note in Figure 5.4.6 is that there is a range of dimensions for which canonical projections satisfy the first condition of Theorem 5.4.3. However, as we see in Figure 5.4.5, the third condition of Theorem 5.4.3 is never met via our canonical projections. Furthermore, the maximum deviation away from the packing distance is roughly constant as the ambient dimenion decreases but is almost nonincreasing as the subspace

dimension is reduced. This suggests that it may be easier to focus, at least in the immediate future, on maps which only reduce K.

Trends seen in the results produced when one dimension is fixed and the other is reduced also occur when both dimensions are reduced. This can be seen in surfaces in Figures 5.4.10, 5.4.11, and 5.4.12. The red line indicating Theorem 5.4.3 criteria being met now manifest as the red planes in these figures. The surface in Figure 5.4.12 reveals that there is a set of reduced dimension configurations that satisfy the first condition of the theorem. Again, we see that there are no reduced dimension configurations produced via canonical projections that satify the third condition of the theorem. From Figure 5.4.12 one might hypothesize that there is a relationship between n and k which result in the determining whether the first condition is met. As a point of interest, one might also be curious about whether or not the number of subspaces also contributes to this bound.

To explore these ideas we look to the same surface plots generated for the following experiments: 200 uniform random points on $Gr(20, 10)$, 500 points on $Gr(20, 5)$, and 200 points on $Gr(20, 5)$. The pair of surface plots of interest for each of these experiments are shown in Figures 5.4.13 and 5.4.14, Figures 5.4.15 and 5.4.16, and Figures 5.4.17 and 5.4.18, respectively. Across these additional experiments we see a few trends appearing. First, for none of the executed combinations is the deviation constraint, in the current space, met. However, the constraint is closer to being met when there are fewer subspaces involved in the configuration. Second, the region of ambient and subspace dimensions which satisfy the second constraint are larger for smaller numbers of points in the configuration. Computational results suggest that with further investigation a function of the dimension of the ambient space, the dimension of the subspace, and the number of points is likely to exist that would determine whether or not the packing distance based criterion can be met and hence if we can construct a nearly distance preserving map.

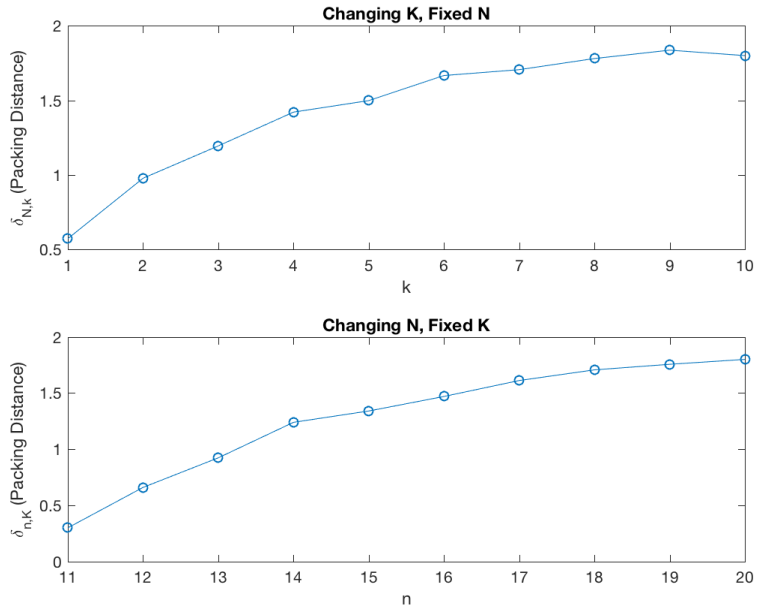


Figure 5.4.4: Packing distance of the configurations as the subspace or ambient dimension are changed with the initial configuration generated as a uniform random sample of 1000 points on $Gr(20, 10)$.

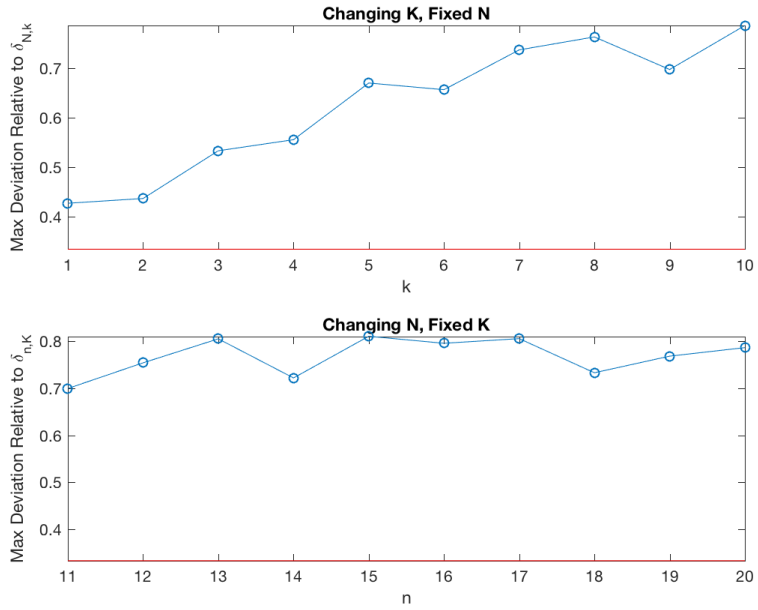


Figure 5.4.5: Maximum deviation away from the packing distance of the configuration, in the current space, as the subspace or ambient dimension are changed. The initial configuration is a uniform random sample of 1000 points on $Gr(20, 10)$. The red horizontal line marks the largest upperbound that would satisfy the theorem's criteria.

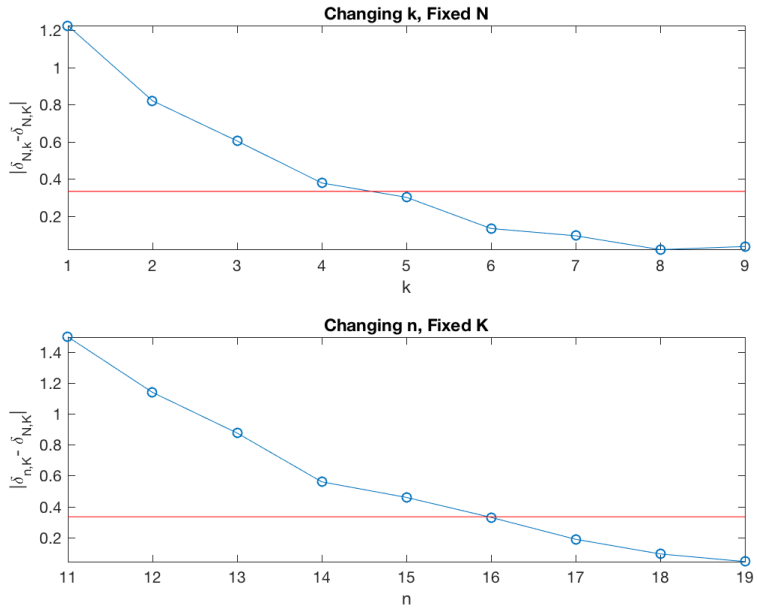


Figure 5.4.6: Change in the packing distance of the configuration in its current space relative to the packing distance of the original configuration as the subspace or ambient dimension are changed. The initial configuration is a uniform random sample of 1000 points on $Gr(20, 10)$. The red horizontal line marks the largest upperbound that would satisfy the theorem's criteria.

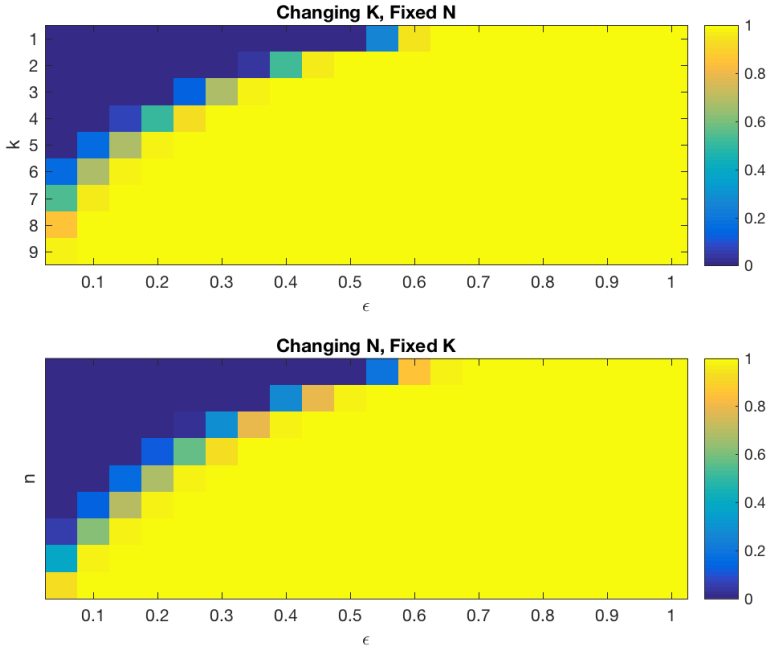


Figure 5.4.7: Percentage of pairwise distances that are preserved for different isometry values (ϵ) as the subspace or ambient dimension are changed. The initial configuration is a uniform random sample of 1000 points on $Gr(20, 10)$.

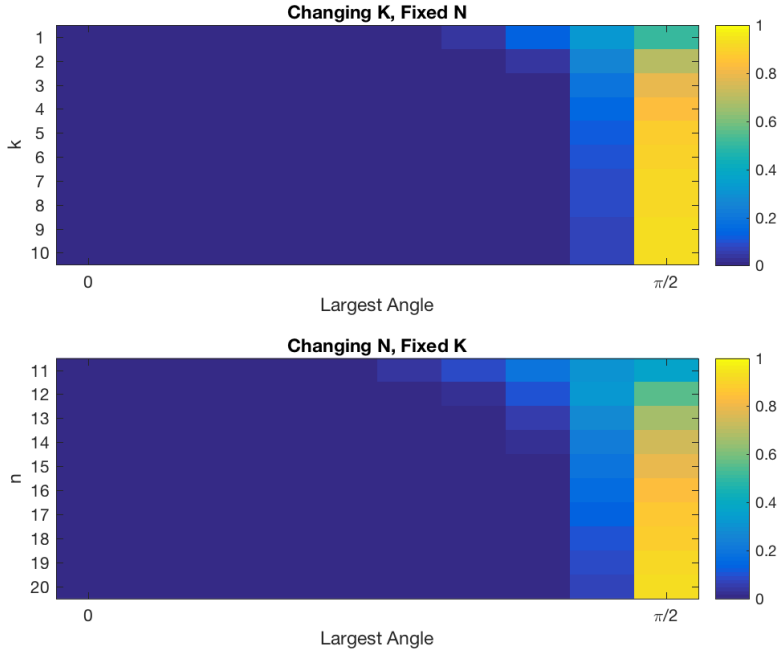


Figure 5.4.8: Distribution of the largest principal angle between points in the configuration as the subspace or ambient dimension are changed. The initial configuration is a uniform random sample of 1000 points on $Gr(20, 10)$.

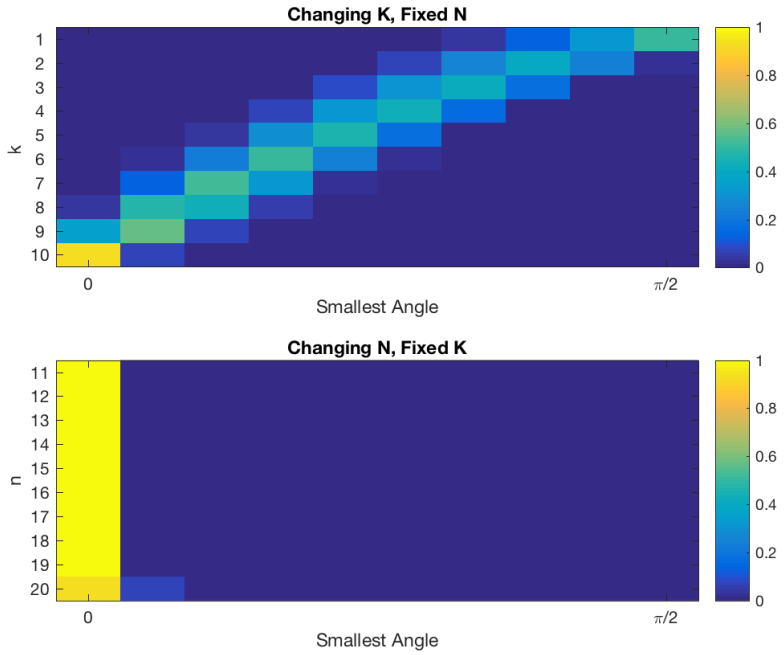


Figure 5.4.9: Distribution of the smallest principal angle between points in the configuration as the subspace or ambient dimension are changed. The initial configuration is a uniform random sample of 1000 points on $Gr(20, 10)$.

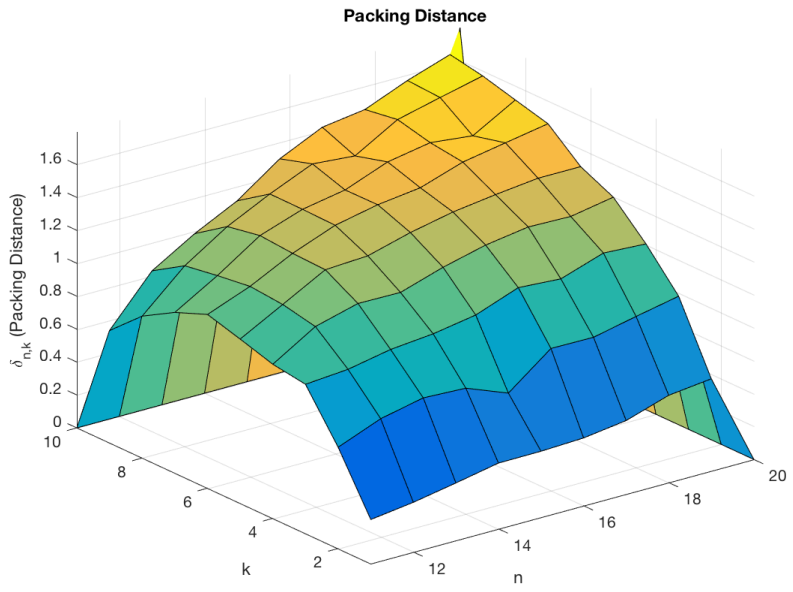


Figure 5.4.10: Packing distance of the configurations as both the subspace and ambient dimension are changed. The initial configuration is a uniform random sample of 1000 points on $Gr(20, 10)$.

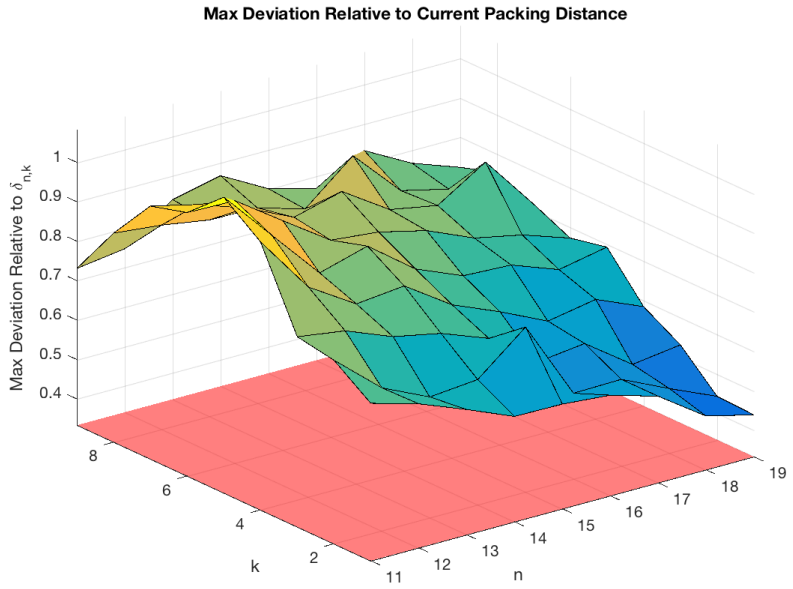


Figure 5.4.11: Maximum deviation away from the packing distance of the configuration, in the current space, as both the subspace and ambient dimension are changed. The initial configuration is a uniform random sample of 1000 points on $Gr(20, 10)$. The red horizontal plane marks the largest upperbound that would satisfy the theorem's criteria.

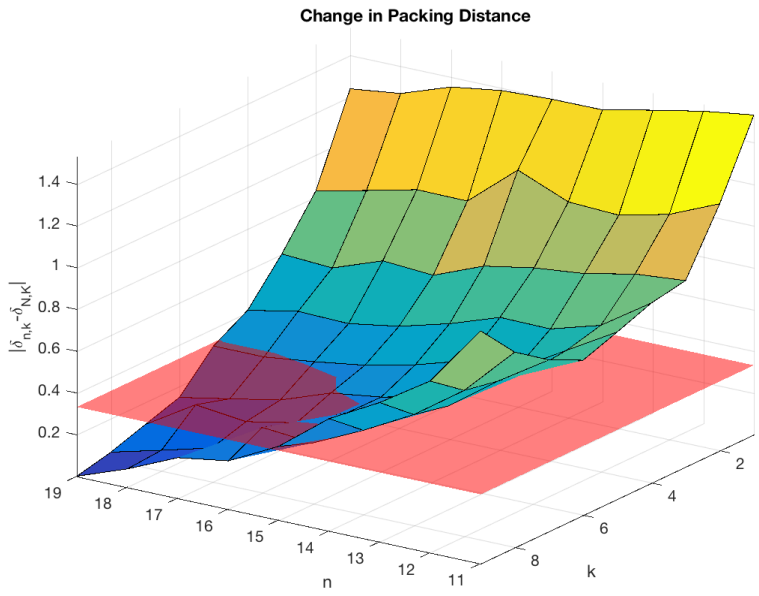


Figure 5.4.12: Change in the packing distance of the configuration in its current space relative to the packing distance of the original configuration as both the subspace and ambient dimension are changed. The initial configuration is a uniform random sample of 1000 points on $Gr(20, 10)$. The red horizontal plane marks the largest upperbound that would satisfy the theorem's criteria.

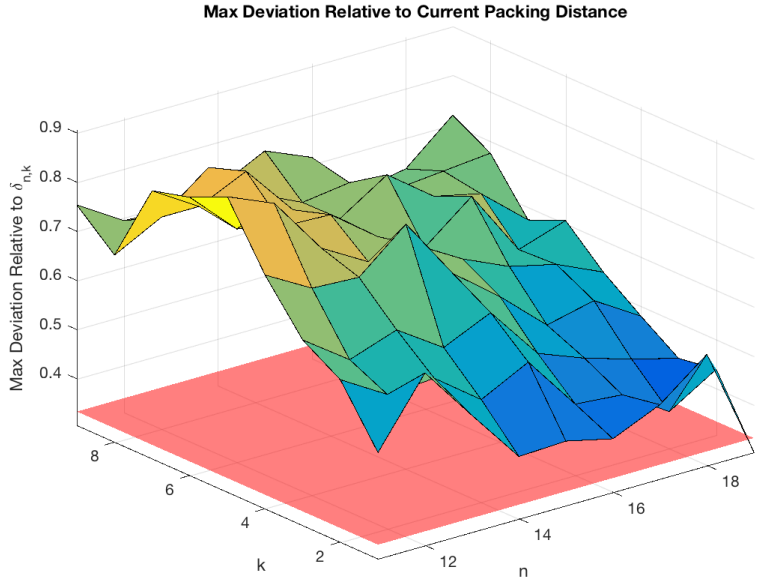


Figure 5.4.13: Maximum deviation away from the packing distance of the configuration, in the current space, as both the subspace and ambient dimension are changed. The initial configuration is a uniform random sample of 200 points on $Gr(20, 10)$. The red horizontal plane marks the largest upperbound that would satisfy the theorem's criteria.

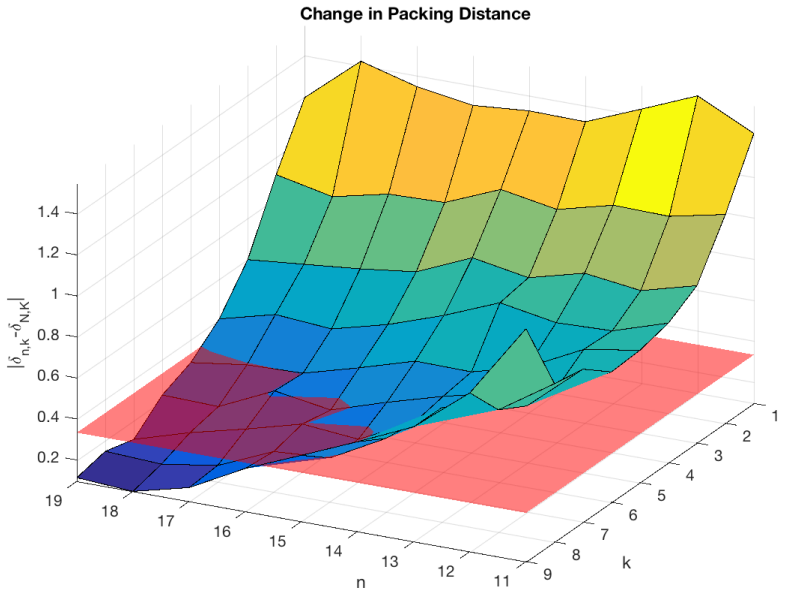


Figure 5.4.14: Change in the packing distance of the configuration in its current space relative to the packing distance of the original configuration as both the subspace and ambient dimension are changed. The initial configuration is a uniform random sample of 200 points on $Gr(20, 10)$. The red horizontal plane marks the largest upperbound that would satisfy the theorem's criteria.

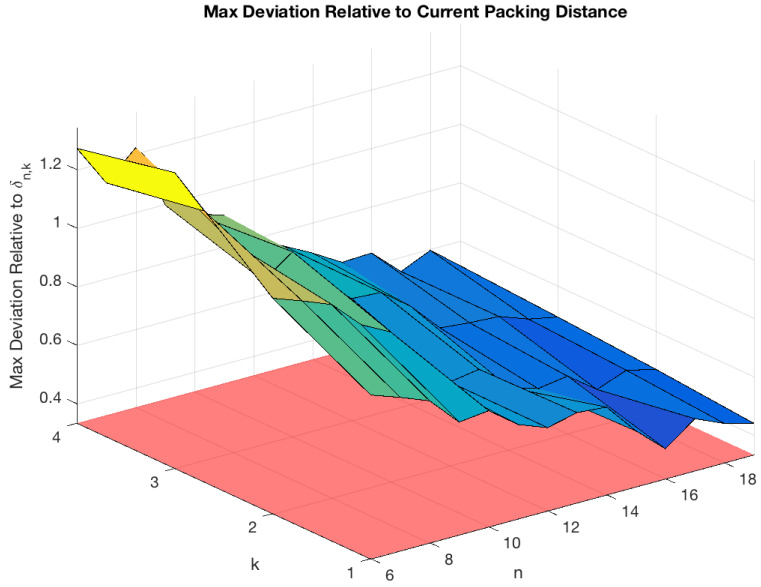


Figure 5.4.15: Maximum deviation away from the packing distance of the configuration, in the current space, as both the subspace and ambient dimension are changed. The initial configuration is a uniform random sample of 500 points on $Gr(20, 5)$. The red horizontal plane marks the largest upperbound that would satisfy the theorem's criteria.

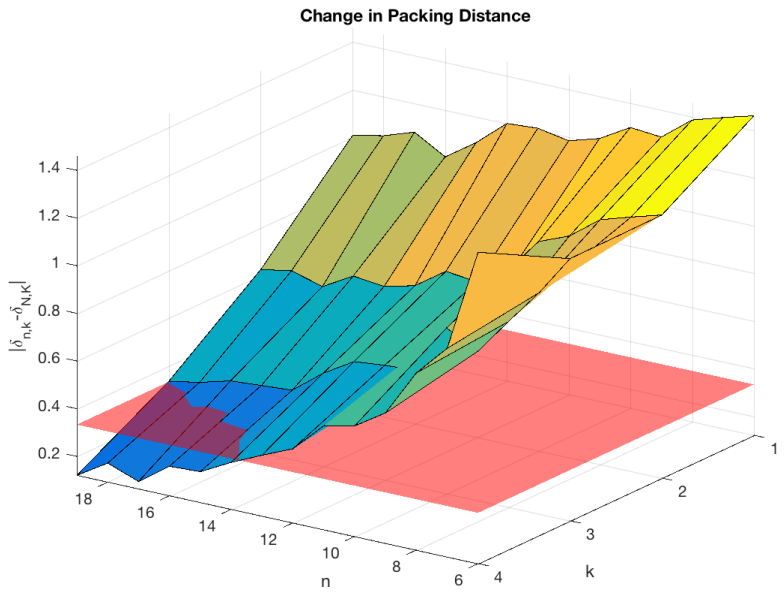


Figure 5.4.16: Change in the packing distance of the configuration in its current space relative to the packing distance of the original configuration as both the subspace and ambient dimension are changed. The initial configuration is a uniform random sample of 500 points on $Gr(20, 5)$. The red horizontal plane marks the largest upperbound that would satisfy the theorem's criteria.

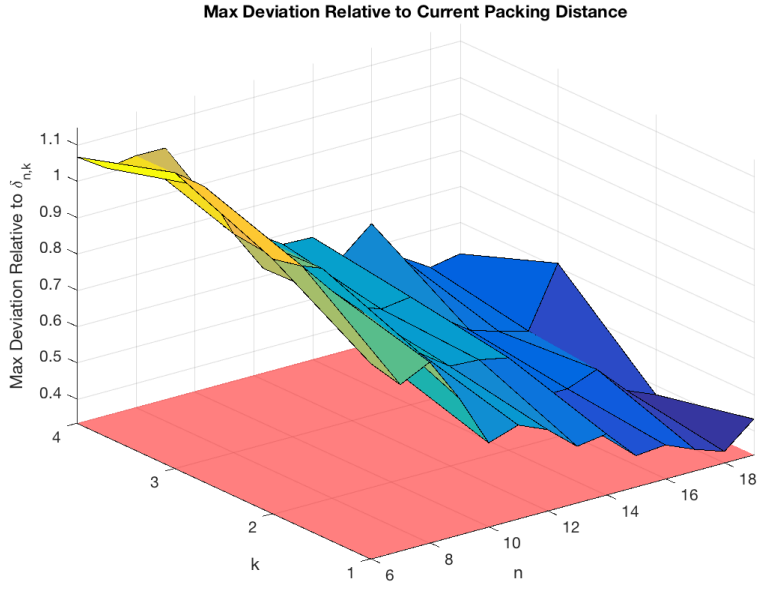


Figure 5.4.17: Maximum deviation away from the packing distance of the configuration, in the current space, as both the subspace and ambient dimension are changed. The initial configuration is a uniform random sample of 200 points on $Gr(20, 5)$. The red horizontal plane marks the largest upperbound that would satisfy the theorem's criteria.

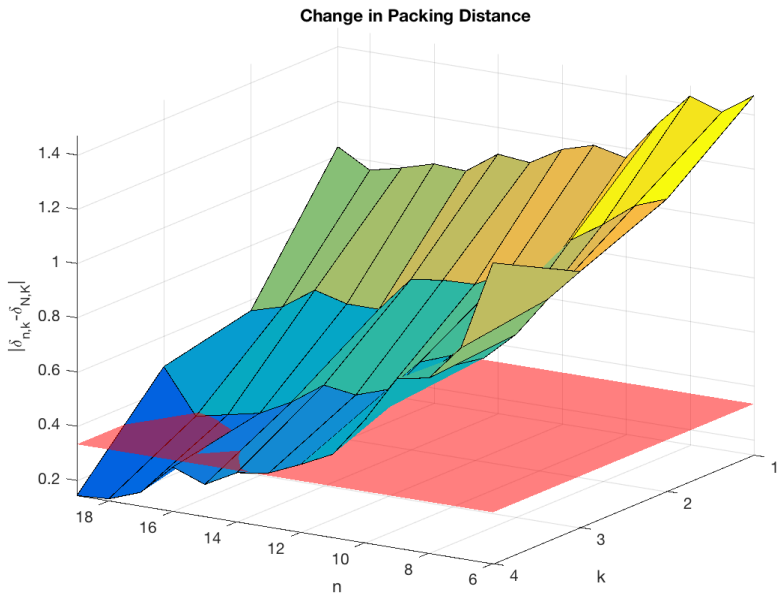


Figure 5.4.18: Change in the packing distance of the configuration in its current space relative to the packing distance of the original configuration as both the subspace and ambient dimension are changed. The initial configuration is a uniform random sample of 200 points on $Gr(20, 5)$. The red horizontal plane marks the largest upperbound that would satisfy the theorem's criteria.

5.5 Discussion

Although we have not concretely proved the existence of a nearly isometric mapping between Grassmannians characterized by different dimensions, we have taken the first steps (down multiple paths) that may lead to such a proof. Of particular promise is the approach tying together the formulation of a Grassmannian packing problem to nearly distance preserving maps. While the approach is promising, the formalizing of an optimization problem that would find a suitable configuration is not trivial. Nor is it obvious how one might implement some of the neccessary constraints. In the paper [8] there are six properties that a matrix needs to satisfy to be able to be factored into a feasible configuration on $Gr(N, K)$. If $\boldsymbol{G}$ is that matrix whose factorization realizes a feasible configuration of m points on $Gr(N, K)$ whose packing distance is greater than ρ, then $\boldsymbol{G}$ must satisfy

1. $\boldsymbol{G}$ is Hermitian,

2. each diagonal block of $\boldsymbol{G}$ must be an identity matrix,

3. each off diagonal block $\boldsymbol{G}_{ij}$ must satisfy $||\boldsymbol{G}_{ij}||_F \leq \sqrt{K-\rho}$,

4. $\boldsymbol{G}$ must be positive semidefinite,

5. $rank(\boldsymbol{G}) \leq m$, and

6. $tr(\boldsymbol{G}) = NK$.

For the problem we consider the matrix $\boldsymbol{G}$ as corresponding to a configuration on $Gr(n,k)$. The first, second, and fourth constraints would be unchanged. The third property would be modified to account for the criteria needed to satisfy $|\text{pack}_{N,K}(\bar{\mathbf{V}}) - \text{pack}_{n,k}(f(\bar{\mathbf{V}}))| \leq \epsilon_1/3$. The rank constraint would become an equality constraint and the trace constraint would be modified to $tr(\boldsymbol{G}) = nk$. Finally, an additional constraint (likely spectral) would need to be added to ensure

$$\max_{i \neq j}\{|\text{pack}_{n,k}(f(\bar{\mathbf{V}})) - d_c(f(\mathbf{V}_i), f(\mathbf{V}_j))|\} \leq 1/3.$$

Unlike the packing problem, we only need to find a feasible configuration once because our packing distance is strictly bounded.

Our computational results hint strongly at trends between the change of values of interest and a reduction in either, or both, subspace or ambient dimension. The results suggest that there is a region of combinations of reduced ambient and subspace dimensions that can satisfy one of the needed criteria for our Theorem 5.4.3 under canonical projections. However, the results highlight the difficulty in finding configurations (at least via canonical projections) that satisfy the deviation about the packing distance constraints. These computational experiments will be extended to a larger set of combinations of ambient and subspace dimensions in ongoing and future work. Exploration of alternative mappings between Grassmannians are still needed. Additionally, exploration of whether or not Theorem 5.4.3's criteria are more easily satisfied by computing distances using an alternative distance measure on the Grassmannian will be computed. Alternative distance

measures were also considered in [8] and produced different formulations of the optimization problem we propose to modify for our purposes.

Chapter 6
Conclusion

A desire to perform dimensionality reduction can be prompted by many things. Some common motivations for dimensionality reduction come in the form of feature selection, data visualization, building interpretable models, computational time, and data storage. In medical and biological sensing there are often a very large number of variables or measurements available and dimensionality reduction becomes particularly important due to the human factor. Humans can only interpret models that consider a relatively small number of variables at a time (feature selection) and are unlikely to trust their health to uninterpretable models. Moreover, in a world filled with "math-phobia" an ability to visualize data and models also contributes to earning a human user's trust.

While there are many reasons one may want to perform dimensionality reduction, it should not be done without careful consideration. There are several methods for dimensionality reduction that involve ad hocly chosen thresholds, parameters, or ranks and can be implemented as a black box. Such approaches may be useful in specific applications but it can be difficult to determine when they will be beneficial. Each of the methods developed in this dissertation have a lesser dependence on ad hoc choices and more readily lend themselves to thorough mathematical analysis. The methods are developed with the intention of preserving mathematical relationships between: numerical relationships (like pairwise distance) or class/task based relationships (being able to distinguish or identify group membership).

Using the mathematical properties of the Fourier transform we were able to develop a rotationally invariant, vector representation of rare circulating cells of interest. By building multiple models, each of which used a $1-$norm regularization to promote feature selection, and looking at features that were consistently given large weights across the models we were able to identify a small set of features that could be inverted to aid in visualization of differentiating structure between cell types. The ability to visualize the information on which automated classification was being performed allowed for fruitful interdisciplinary dialogue. As a result of this dialogue we

were able to interpret the model we built and validate that the model is consistent with classification criteria used by an expert pathologist. Unlike a human expert, our purely mathematical classification does not suffer from fatigue and is consequently more trust worthy in some regard.

Across an application in ground cover classification and an application in generalized modal analysis, we saw improved results achieved using reduced dimension estimators. Reduced dimension estimators were produced by controlling the bias and variance of the estimators when written as a function of the rank of particular matrices. A rich theoretical foundation gave way to rigid but interpretable order determination rules for identifying the optimal rank as well as which dimensions should be retained to produce the optimal rank. Determining optimal (by some measure) reduced rank matrices had previously been considered. However, to our knowledge these frameworks focus on how to produce the optimal reduced rank matrix for a specific rank. They do not explicitly formulate rules that simultaneously identify what reduced rank is optimal and how to produce that optimal rank version of the matrix.

Frequency Agile Lidar is a multispectral laser radar system that was designed to aid in the tracking and detection of aerosolized bio-agents. Algorithms were implemented in 2012 that analyze the data acquired using this system. These algorithms begin with a complex, ad hoc parameter dependent, multi stage preprocessing. Cleaned data is then analyzed using a non-negative matrix factorization problem that can be solved using the popular convex optimization technique known as Split Bregman. There are two glaring problems that prevent this system from being usable on-line. One of the hangups is that the preprocessing requires knowledge of data acquired in the absence of aerosol. The more significant challenge that prevents the on-line use of the system is that the algorithm assumes that the internal factoring dimension in the non-negative matrix factorization problem is known. In the context of this problem the internal factoring dimension is interpreted as the number of aerosols present. Although we have not completely solved this problem, we have proposed a solution that could (in theory) determine the internal factoring dimension and produce a superior factorization. In addition to addressing one of the on-line challenges through the introduction of a novel $1-$norm regularized term in the objective function we found that removal of one

of the previously existing $1-$norm regularization terms from the objective function improved the results of factorization. These improvements included decreased within class variance of feature vectors used for classification as well as increased serparation between classes of feature vectors.

The culminating work of this dissertation considered nearly distance preserving maps between Grassmannian manifolds. Of particular interest are maps that both nearly preserve pairwise distances as well as reduce the dimension of the Grassmannian. A detailed reproduction of a statistical proof of the well known Johnson-Lindenstrauss Lemma is presented. The Johnson-Lindenstrauss Lemma states the existence of nearly distance preserving maps between Euclidean spaces when the dimension of the image space is bounded as a function of the number of points and a distortion factor. It was this lemma that provided the inspiration for our exploration. We first considered an analagous statistical approach to the proof in Euclidean space. Next, and more significantly, we established a connection between packing distances and nearly distance preserving maps. Moreover, we proposed modifications to an exisiting algorithm that addresses the Grassmannian packing problem that if solvable would prove, by construction, the existence of a nearly distance preserving, dimension reducing map. Computational experiments presented provide insight into the feasibility of different assumptions of our theorem being met. Although no concrete application was considered, potential applications were proposed and will be considered in future work.

All of the projects comprising this dissertation are connected through the mathematical concept of dimension though it manifests differently in each project. Challenges related to estimating dimension and reducing dimension will only become more relevant as we continue moving forward in the age of big data. These challenges are interesting from an application based standpoint and the math theoretic perspective. This dissertation is bookended by projects that evolved from each of these perspectives. In the first project we were motivated by a specific task in automated classification and then used existing techniques in dimensionality reduction and properties of our novel representation to visualize and interpret a successful mathematical model. The final project was inspired purely by mathematical theory but we believe it will be widely applicable to tasks rang-

ing from mental task identification to identification of important spatial or temporal dimensions in video data.

Dimensionality reduction has shown itself to be a versitile and useful tool across many fields and a large array of applications. Some methods for dimensionality reduction are used like a black box without consideration as to whether or not the mathematical/geometric assumptions are being met. Other methods rely heavily on ad hocly chosen parameter values. Further development of new algorithms that aim to preserve mathematical relationships in data with minimal dependence on ad hoc parameters, like those in this dissertation, is needed. As more types of mathematical relationships are considered the breadth of data sets which can benefit from dimensionality reduction will increase. These benefits include improved classification accuracies, increased detection rates, decreased computational time, and reduction in storage space needed.

What is data? Previously we defined data as a collection of numerical representations which can be analyzed using mathematical techniques to reveal facts and patterns on which reasonable, and useful, models can be built. We leave the reader now with a less esoteric and more ethereal definition. Pythagoras is credited with the statement "All is number." From this perspective a data set is merely a subset of everything. Mathematics is the language of everything and contains the tools needed to translate these subsets into information which can be understood.

Bibliography

[1] T Emerson, M Kirby, K Bethel, A Kolatkar, M Luttgen, S O'Hara, P Newton, and P Kuhn. Fourier-ring descriptor to characterize rare circulating cells from images generated using immunofluorescence microscopy. *Computerized Medical Imaging and Graphics*, 40:70–87, 2015.

[2] T Emerson. Automated detection of circulating cells using low level features. Master's thesis, Colorado State University, 2013.

[3] R Warren, S Osher, and R Vanderbeek. Multiple aerosol unmixing by the split bregman algorithm. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(9):3271–3279, 2012.

[4] T Emerson, M Kirby, L Scharf, and C Peterson. Reduced dimension estimators in matched subspace detection. In *8th Workshop on Hyperspectral Image and Signal Processing: Evolutions in Remote Sensing*, August 2016.

[5] T Emerson, M Kirby, L Scharf, and C Peterson. Reduced dimension estimators for controlling bias and variance in oblique pseudo inverses and oblique projections. In *IEEE Transactions on Signal Processing*, Submitted 2016.

[6] W Johnson and J Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[7] S Dasgupta and A Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures and Algorithms*, 22(1):60–65, 2003.

[8] I Dhillon, R Heath, T Strohmer, and J Tropp. Constructing packings in grassmannian manifolds via alternating projection. *Experimental Mathematics*, 17(1):9–35, 2008.

[9] M Giuliano, A Giordano, S Jackson, K Hess, U De Giorgi, M Mego, B Handy, N Ueno, R Alvarez, M De Laurentiis, et al. Circulating tumor cells as prognostic and predictive

markers in metastatic breast cancer patients receiving first-line systemic treatment. *Breast Cancer Res*, 13(3):R67, 2011.

[10] S Cohen, C Punt, N Iannotti, B Saidman, K Sabbath, N Gabrail, J Picus, M Morse, E Mitchell, M Miller, et al. Prognostic significance of circulating tumor cells in patients with metastatic colorectal cancer. *Annals of Oncology*, 20(7):1223–1229, 2009.

[11] J Nieva, M Wendel, M Luttgen, D Marrinucci, L Bazhenova, A Kolatkar, R Santala, B Whittenberger, J Burke, M Torrey, et al. High-definition imaging of circulating tumor cells and associated cellular events in non-small cell lung cancer patients: A longitudinal analysis. *Physical biology*, 9(1):016004, 2012.

[12] D Danila, G Heller, G Gignac, R Gonzalez-Espinoza, A Anand, E Tanaka, H Lilja, L Schwartz, S Larson, M Fleisher, et al. Circulating tumor cell number and prognosis in progressive castration-resistant prostate cancer. *Clinical Cancer Research*, 13(23):7053–7058, 2007.

[13] K Pachmann, O Camara, A Kavallaris, S Krauspe, N Malarski, M Gajda, T Kroll, C Jörke, U Hammer, A Altendorf-Hofmann, et al. Monitoring the response of circulating epithelial tumor cells to adjuvant chemotherapy in breast cancer allows detection of patients at risk of early relapse. *Journal of Clinical Oncology*, 26(8):1208–1215, 2008.

[14] S Riethdorf, H Fritsche, V Müller, T Rau, C Schindlbeck, B Rack, W Janni, C Coith, F Beck, Kand Jänicke, et al. Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: A validation study of the CellSearch system. *Clinical Cancer Research*, 13(3):920–928, 2007.

[15] D Marrinucci, K Bethel, A Kolatkar, M Luttgen, M Malchiodi, F Baehring, K Voigt, D Lazar, J Nieva, L Bazhenova, et al. Fluid biopsy in patients with metastatic prostate, pancreatic and breast cancers. *Physical Biology*, 9(1):016003, 2012.

[16] S Chen, M Zhao, G Wu, C Yao, and J Zhang. Recent advances in morphological cell image analysis. *Computational and Mathematical Methods in Medicine*, 2012, 2012.

[17] D Zhang and G Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.

[18] P Elbischger, S Geerts, K Sander, G Ziervogel-Lukas, and P Sinah. Algorithmic framework for hep-2 fluorescence pattern classification to aid auto-immune diseases diagnosis. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pages 562–565. IEEE, 2009.

[19] P Perner, H Perner, and B Müller. Mining knowledge for hep-2 cell image classification. *Artificial Intelligence in Medicine*, 26(1):161–173, 2002.

[20] A Wiliem, Y Wong, C Sanderson, P Hobson, S Chen, and B Lovell. Classification of human epithelial type 2 cell indirect immunofluoresence images via codebook based descriptors. *arXiv preprint arXiv:1304.1262*, 2013.

[21] D Comaniciu, P Meer, and D Foran. Image-guided decision support system for pathology. *Machine Vision and Applications*, 11(4):213–224, 1999.

[22] Y-L Huang, Y-L Jao, T-Y Hsieh, and C-W Chung. Adaptive automatic segmentation of hep-2 cells in indirect immunofluorescence images. In *Sensor Networks, Ubiquitous and Trustworthy Computing, 2008. SUTC'08. IEEE International Conference on*, pages 418–422. IEEE, 2008.

[23] C Wählby, J Lindblad, M Vondrus, E Bengtsson, and L Björkesten. Algorithms for cytoplasm segmentation of fluorescence labelled cells. *Analytical Cellular Pathology*, 24(2):101–111, 2002.

[24] C Wählby, I-M Sintorn, F Erlandsson, G Borgefors, and E Bengtsson. Combining intensity, edge and shape information for 2d and 3d segmentation of cell nuclei in tissue sections. *Journal of Microscopy*, 215(1):67–76, 2004.

[25] R Hiemann, T Büttner, T Krieger, D Roggenbuck, U Sack, and K Conrad. Challenges of automated screening and differentiation of non-organ specific autoantibodies on hep-2 cells. *Autoimmunity Reviews*, 9(1):17–22, 2009.

[26] P Agrawal, M Vatsa, and R Singh. Hep-2 cell image classification: A comparative analysis. In *Machine Learning in Medical Imaging*, pages 195–202. Springer, 2013.

[27] M Abramoff, P Magalhaes, and S Ram. Image processing with imagej. *Biophotonics International*, 11(7):36–42, 2004.

[28] D-M Tsai and Y-H Tsai. Rotation-invariant pattern matching with color ring-projection. *Pattern Recognition*, 35(1):131 – 141, 2002.

[29] D-M Tsai and C-H Chiang. Rotation-invariant pattern matching using wavelet decomposition. *Pattern Recognition Letters*, 23(1â3):191 – 201, 2002.

[30] J Hipp, J Cheng, J Hanson, W Yan, P Taylor, N Hu, J Rodriguez-Canales, M Tangrea, M Emmert-Buck, U Balis, et al. Sivq-aided laser capture microdissection: A tool for high-throughput expression profiling. *Journal of Pathology Informatics*, 2(1):19, 2011.

[31] J Hipp, J Cheng, M Toner, R Tompkins, and U Balis. Spatially Invariant Vector Quantization: A Pattern Matching Algorithm for Multiple Classes of Image Subject Matter Including Pathology. *Journal of Pathology Informatics*, 2(1):13, 2011.

[32] R-E Fan, K-W Chang, C-J Hsieh, X-R Wang, and C-J Lin. Liblinear: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.

[33] M Kirby. *Geometrical Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns*. John Wiley and Sons, Inc., 2001.

[34] R Warren, S Osher, and R Vanderbeek. Multiple aerosol unmixing by the split bregman algorithm. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(9):3271–3279, 2012.

[35] R Warren, R Vanderbeek, A Ben-David, and J Ahl. Simultaneous estimation of aerosol cloud concentration and spectral backscatter from multiple-wavelength lidar data. *Applied Optics*, 47(24):4309–4320, 2008.

[36] T Goldstein and S Osher. The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[37] H Zou and T Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

[38] R Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[39] P Hoyer. Non-negative matrix factorization with sparseness constraints. *The Journal of Machine Learning Research*, 5:1457–1469, 2004.

[40] R Warren, R Vanderbeek, and J Ahl. Estimation and discrimination of aerosols using multiple wavelength lwir lidar. In *SPIE Defense, Security, and Sensing*, pages 766504–766504. International Society for Optics and Photonics, 2010.

[41] D Manolakis, C Siracusa, and G Shaw. Hyperspectral subpixel target detection using the linear mixing model. *Geoscience and Remote Sensing, IEEE Transactions on*, 39(7):1392–1409, 2001.

[42] B Holben and Y Shimabukuro. Linear mixing model applied to coarse spatial resolution data from multispectral satellite sensors. *Remote Sensing*, 14(11):2231–2240, 1993.

[43] J Settle and N Drake. Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing*, 14(6):1159–1177, 1993.

[44] P Puyou-Lascassies, A Podaire, and M Gay. Extracting crop radiometric responses from simulated low and high spatial resolution satellite data using a linear mixing model. *International Journal of Remote Sensing*, 15(18):3767–3784, 1994.

[45] L Scharf and B Friedlander. Matched subspace detectors. *Signal Processing, IEEE Transactions on*, 42(8):2146–2157, 1994.

[46] S Kraut, L Scharf, and T McWhorter. Adaptive subspace detectors. *Signal Processing, IEEE Transactions on*, 49(1):1–16, 2001.

[47] S Kraut, L Scharf, and R Butler. The adaptive coherence estimator: A uniformly most-powerful-invariant adaptive detection statistic. *Signal Processing, IEEE Transactions on*, 53(2):427–438, 2005.

[48] R Behrens and L Scharf. Signal processing applications of oblique projection operators. *Signal Processing, IEEE Transactions on*, 42(6):1413–1424, 1994.

[49] G Seber and A Lee. *Linear Regression Analysis*, volume 936. John Wiley & Sons, 2012.

[50] S Wicker and V Bhargava. *Reed-Solomon Codes and their Applications*. John Wiley & Sons, 1999.

[51] G Forney. On decoding bch codes. *IEEE Transactions on Information Theory*, 11(4):549–557, 1965.

[52] B Gao, G Morison, and P Kundur. Voltage stability evaluation using modal analysis. *Power Systems, IEEE Transactions on*, 7(4):1529–1542, 1992.

[53] B Peeters and G De Roeck. Reference-based stochastic subspace identification for output-only modal analysis. *Mechanical Systems and Signal Processing*, 13(6):855–878, 1999.

[54] Z Wu and N Huang. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(01):1–41, 2009.

[55] L Scharf. *Statistical Signal Processing*, volume 98. Addison-Wesley Reading, MA, 1991.

[56] Y Hua, M Nikpour, and P Stoica. Optimal reduced-rank estimation and filtering. *IEEE Transactions on Signal Processing*, 49(3):457–469, 2001.

[57] L Scharf and T McWhorter. Adaptive matched subspace detectors and adaptive coherence estimators. In *Signals, Systems and Computers, 1996. Conference Record of the Thirtieth Asilomar Conference on*, pages 1114–1117. IEEE, 1996.

[58] M. Woodbury. Inverting modified matrices. *Memorandum Report Statistical Research Group*, 42, 1950.

[59] D Manolakis and G Shaw. Detection algorithms for hyperspectral imaging applications. *Signal Processing Magazine, IEEE*, 19(1):29–43, 2002.

[60] M Baumgardner, L Biehl, and D Landgrebe. 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3, Sep 2015. https://purr.purdue.edu/publications/1947/1.

[61] 224 band arvis hyperspectral image data set: Salinas, valley. http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.

[62] G Golub and C Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 2715 North Charles Street, Baltimore, Maryland, 21218, fourth edition, 2013.

[63] S Krantz. *Essentials of Topology with Applications*. Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, Florida, 33487, 2010.

[64] D Dummit and R. Foote. *Abstract Algebra*. John Wiley and Sons, 111 River Street, Hoboken, New Jersey, 07030, third edition, 2004.

[65] Y Chickuse. *Statistics on Special Manifolds*. Springer-Verlag New York, 175 Fifth Avenue, New York, New York, 10010, 2003.

[66] T Sarlos. Improved approximation algorithms for large matrices via random projections. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 143–152. IEEE, 2006.

[67] P Pakrooh. Chapter 1: Doctoral candidate preliminary exam. Preliminary Exam.

[68] J Conway, R Hardin, and N Sloane. Packing lines, planes, etc.: Packings in grassmannian spaces. *Experimental Mathematics*, 5(2):139–159, 1996.

[69] K Mardia, J Kent, and J Bibby. *Multivariate Analysis*. Academic Press, Inc., San Diego, California, 92101, 1979.

[70] S Chepushtanova and M Kirby. Classification of hyperspectral imagery on embedded grassmannians. *arXiv preprint arXiv:1502.00946*, 2015.

[71] P Absil, A Edelman, and P Koev. On the largest principal angle between random subspaces. *Linear Algebra and its Applications*, 414(1):288–294, 2006.