Differentiable Programming for Physics-based Hyperspectral Unmixing

by

John Janiczek

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2020 by the
Graduate Supervisory Committee:

Suren Jayasuriya, Co-Chair
Gautam Dasarathy, Co-Chair
Phil Christensen

ARIZONA STATE UNIVERSITY

May 2020

ABSTRACT

Hyperspectral unmixing is an important remote sensing task with applications including material identification and analysis. Characteristic spectral features make many pure materials identifiable from their visible-to-infrared spectra, but quantifying their presence within a mixture is a challenging task due to nonlinearities and factors of variation. In this thesis, physics-based approaches are incorporated into an end-to-end spectral unmixing algorithm via differentiable programming. First, sparse regularization and constraints are implemented by adding differentiable penalty terms to a cost function to avoid unrealistic predictions. Secondly, a physics-based dispersion model is introduced to simulate realistic spectral variation, and an efficient method to fit the parameters is presented. Then, this dispersion model is utilized as a generative model within an analysis-by-synthesis spectral unmixing algorithm. Further, a technique for inverse rendering using a convolutional neural network to predict parameters of the generative model is introduced to enhance performance and speed when training data are available. Results achieve state-of-the-art on both infrared and visible-to-near-infrared (VNIR) datasets as compared to baselines, and show promise for the synergy between physics-based models and deep learning in hyperspectral unmixing in the future.

i

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure   Page

Chapter 1

INTRODUCTION

The journey of light from the surface of a planet, through the atmosphere, and into the detector of a satellite instrument can be described by physics. A conceptual diagram of this journey is shown in Figure 1.1. The central idea of this thesis is to blend these physical models with optimization and machine learning algorithms to extract information from hyperspectral images.

Hyperspectral imaging is a method of imaging where light radiance is densely sampled at multiple wavelengths. Increasing spectral resolution beyond a traditional
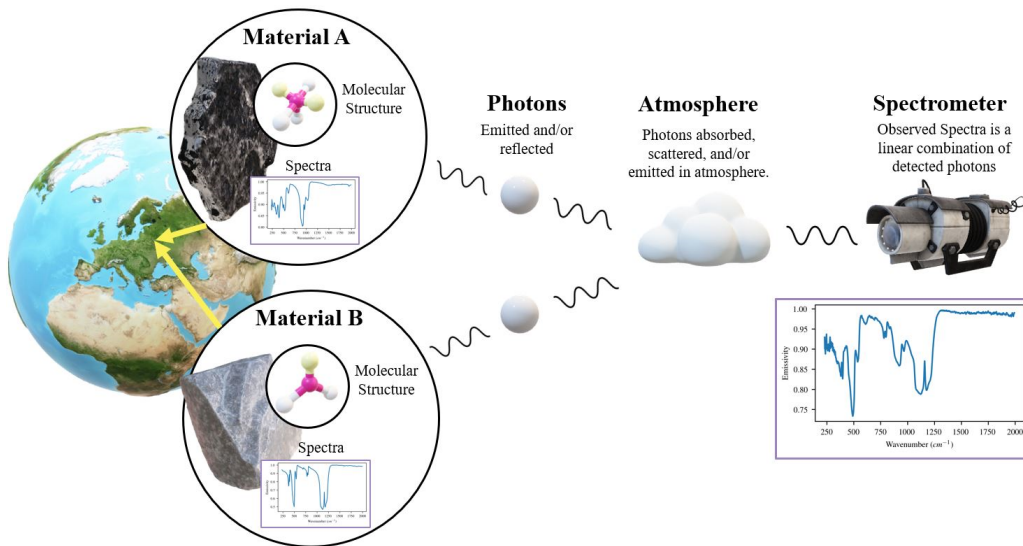
Figure 1.1: Journey of a Photon: A satellite captures a hyperspectral image by detecting photons reflected and emitted from the planetary surface. The molecular structure of each material gives a unique spectral response for these reflections and emissions. Non-linear effects can occur as the photon interacts with the scene. However, the detected photons do combine linearly at the spectrometer.

camera's red, green, and blue spectral bands typically requires more expensive detectors, optics, and/or lowered spatial resolution. However, hyperspectral imaging has demonstrated its utility in computer vision, biomedical imaging, and remote sensing. For example, absorption spectra can show the concentration of oxygen saturation in hemoglobin, which is useful for cancer diagnosis by identifying the related physical processes of angiogenisis and hypermetabolism. Further, using hyperspectral imaging has been shown to improve performance of classical computer vision tasks such as object recognition [54, 13]. In particular, spectral information is critically important for understanding material reflectance and emission properties, important for recognizing materials. Spectral unmixing is a specific task within hyperspectral imaging for identifying material presence in a mixture. This analysis of hyperspectral images, usually from aircraft or satellite based spectrometers, has application to many land classification problems related to ecology, hydrology, and mineralogy [37, 43]. For example, thermal emission spectroscopy was used to map the minerology of Mars including the important discovery of the hematite mineral, which is known to form in liquid water [2]. While pure materials have characteristic spectral features that can be immediately recognized, mixtures require algorithms to identify and quantify material presence. Figure 1.2 shows a conceptual diagram of a spectral unmixing unmixing algorithm.

Many spectral unmixing algorithms are based off of simple physical principles such as the linear mixing model. The model states that an observed spectrum is a linear combination of pure material spectra, scaled by the fractional abundance of each material. Mathematically this is expressed as $\mathbf{b} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}$ where $\mathbf{b}$ is the observed spectra, $\mathbf{A}$ is a matrix whose columns are the pure material spectra, $\boldsymbol{\eta}$ is the measurement noise, and $\mathbf{x}$ is the abundance of each pure material. Nonlinear effects are known to occur when photons interact with multiple materials prior to being detected,

2
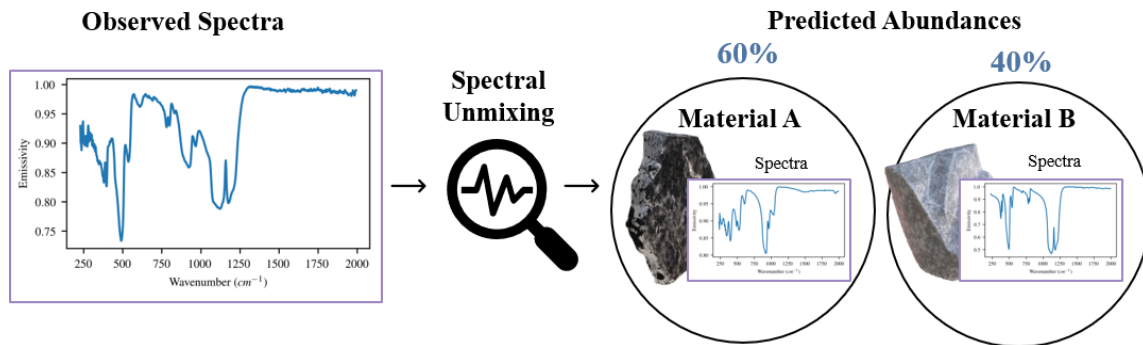
Figure 1.2: Spectral Unmixing: We are given data aquired from a spectrometer. Using spectral unmixing we can predict the materials present in the sample observed by the spectrometer.

which we discuss in more detail in Chapter 2. Many of the techniques for non-linear unmixing encode physics-based radiative transfer models into the algorithms [37].

Recent advances in automatic differentiation [58] have opened the door to include even more sophisticated physical models into predictive algorithms. By making physical models differentiable, the parameters of the models can be predicted with gradient descent optimization. Additionally, differentiable physical models can be blended with machine learning algorithms by backpropagating the error through both the physical and machine learning models, which can improve speed and performance when training data are available. This technique of differentiable programming has recently become a popular research area due to its potential to bridge gaps between physics-based and machine learning-based techniques for computer vision and graphics [74, 32, 1]. Our key insight is to leverage differentiable programming and dispersion theory to address challenges in spectral unmixing.

1.1 Contributions of the Thesis

First, we examine the problem of producing sparse abundance maps using regularization in Chapter 3. A common usage of spectral unmixing is to produce maps

of the fractional abundance of each material on a planetary surface. Typically the spectral unmixing algorithms produce dense maps, meaning that each pixel of the hyperspectral image is predicted to contain many materials. However, it is understood that abundance maps should be sparse, meaning that only a few materials would realistically be present in each region of the scene. Sparse regularization is a method to penalize algorithms from making unrealistically dense predictions. We provide new experiments which show that sparse regularization with the L_p norm has the most performance benefits. We also show that the L_∞ method, which had not been previously tried on spectral unmixing, offers some performance benefits although it is slower than the L_p method.

In Chapter 4, we examine the problem of spectral variability from a physics-based approach. Both linear and non-linear mixing models require the pure material spectra to be specified. However, pure materials have an inherent variability in their spectral signatures, and thus cannot be represented by a single characteristic spectrum. Spectral variability of endmembers is caused by subtle absorption band differences due to factors such as different grain sizes [55, 65, 60, 61] or differing ratios of molecular bonds [12, 70] as shown in Figure 4.1. This variability causes significant errors in unmixing algorithms, and is an active area of research [81, 23, 86].

Our approach is to leverage differentiable programming to incorporate a physics-based dispersion model with parameters that control the strength, shape, and frequency of absorption bands into a spectral unmixing algorithm. Such an approach has the capacity to unmix scenes with a large amount of variability, while constraining the predictions to be physically plausible. These physically plausible variations of endmember spectra also provide additional science data as the variation of absorption bands can reveal properties about the composition and history of the material. To our knowledge, we are the first to use a generative physics model to account for
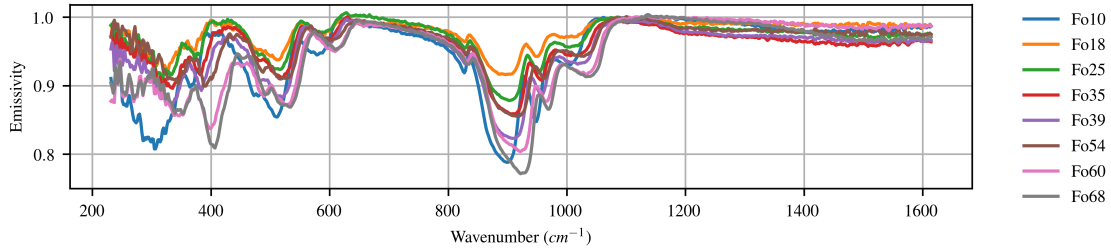
Figure 1.3: Endmember Variation: Several spectra of olivine are plotted to demonstrate it's spectral variability. The olivine mineral is a solid solution with continuous compositional variation of Iron (Fe_2) and Magnesium (Mg_2) bonds. This ratio of bonds (indexed by the Fo number), causes absorption bands to shift in frequency and strength [70].

spectral variability in an unmixing algorithm.

Our specific contributions in this thesis are the following:

- We consider new experiments to compare the sparse regularization techniques and show that L_p is the best performing method. We are also the first to consider the L_∞^{-1} method for spectral unmixing and show it is a viable convex alternative to L_p regularization although it is slower.

- We introduce a physics-based dispersion model (first presented in [67, 45, 76]) to generate and render spectral variation for various pure materials. We provide an efficient optimization method via gradient descent to find dispersion model parameters for this spectral variation.

- We incorporate this dispersion model into an end-to-end spectral unmixing algorithm utilizing differentiable programming to perform analysis-by-synthesis optimization. Analysis-by-synthesis is solved via alternating minimization optimization and requires no training data.

- We further design an inverse rendering algorithm consisting of a convolutional neural network to jointly estimate dispersion model parameters and mineral abundances for spectral unmixing. This method requires training data, but is computationally efficient at test time and outperforms analysis-by-synthesis and other state-of-the-art methods.

We provide extensive analysis of our proposed methods with respect to noise and convergence criteria. To validate our contributions, we test on both synthetic and real datasets using hyperspectral observations in the visible and near infrared (VNIR), and mid to far infrared (IR). The datasets also span three different environments from laboratory, aircraft, and satellite based spectrometers. Our methods achieve state-of-the-art across all datasets, and we compare against several baselines from literature. We hope this work inspires more fusion between physics models and machine learning for hyperspectral imaging and computer vision more generally in the future.

In Chapter 5, we provide a conclusion of our findings and recognize opportunities for future work. Also, we present Appendices for readers interested in specific aspects of this Thesis. In Appendix A, we show a derivation of the dispersion model from first principles. In Appendix B, we provide tables summarizing the dispersion parameters for several important materials. In Appendix C, we provide a discussion on convergence of our alternating minimization approach to our analysis-by-synthesis algorithm. Finally, in Appendix D, we show the mineral maps of Mars produced by our analysis-by-synthesis algorithm.

Chapter 2

RELATED WORK

Linear vs. Non-linear mixing: The models for spectral mixing depend on whether emitted or reflected energy is being measured. Emissions are generally observed in the infrared (IR) wavelengths, around 8-100 μm. Measurements at these wavelengths are dominated by thermal emissions from the top-most surface because most materials are strongly absorbing in the infrared resulting in a low chance of observing reflected photons. This process results in a linear mixing model except when the grain size of the particles approaches optical wavelengths [65, 60, 61].

Reflected sunlight is generally observed in the visual-to-near-infrared (VNIR) wavelengths, 0.4-2.4 μm. In the VNIR, it is common for light to penetrate and interact with a volume of materials prior to reflecting. Volume mixing has been shown to cause non-linearities, particularly if the endmembers are "intimately mixed", meaning that the multiple materials are blended together in the same patch of the surface. This is can lead to non-linear mixing in the VNIR because many materials are relatively transparent at these wavelengths leading to photons penetrating the surface and interacting with a volume. However, in many cases the linear mixing model is applied to VNIR data by assuming the non-linearities are negligible [43, 79]. Other efforts have been made to encode physical knowledge of radiative transfer to account for non-linear mixing by converting the VNIR radiance to single scattering albedo (SSA) [66, 43, 35, 34]. For a review of methods to account for non-linear mixing we refer readers to the review by Heylen et al. [37]. We mainly use thermal emission datasets which are assumed to follow the linear mixing model and VNIR reflected datasets which are assumed to have negligible non-linear effects. Thus, we build our

7

contributions around the linear mixing model, although it would be possible to combine our spectral variability and sparse regularization techniques to non-linear mixing models as well.

Optimization-based Approaches Standard optimization techniques for spectral unmixing include projection, non-negative least squares, weighted least squares, and interior point methods [60, 36, 27, 18, 34]. These family of methods setup an optimization problem based off of physical models describing the observed spectra in terms of the known pure material spectra. For example, in the linear mixing model $\mathbf{b} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}$, where $\mathbf{b}$ is a vector representing the observed spectral observation, $\mathbf{A}$ is a matrix whose columns are the pure material spectra, $\mathbf{x}$ is a vector whose elements are the abundance of the materials corresponding with the columns of $\mathbf{A}$, and $\boldsymbol{\eta}$ is a vector representing noise or other sources of error in the measurement. The matrix-vector multiplication $\mathbf{A}\mathbf{x}$ produces a linear combination of the columns of $\mathbf{A}$ weighted by the elements of $\mathbf{x}$, describing the physics of certain spectral mixing scenarios. Then, this model can be inverted using a least squares approach by finding the abundances that minimize the mean squared error between the modelled and observed spectra, $||\mathbf{b} - \mathbf{A}\mathbf{x}||_2^2$. Note that this notation is simply the squared L_2 norm, where the L_2 norm is defined as $||\mathbf{x}||_2 := (\sum_{i=1}^n |x_i|^2)^{\frac{1}{2}}$. The squared L_2 norm is equivalent to the mean-squared error, and is typically referred to as the least-squares when used in an optimization or curve fitting problem. A closed form solution for this least squares optimization problem exists by multiplying each side of the linear mixing model with the pseudo-inverse of $\mathbf{A}$.

Gradient descent [64] can also be used to solve this optimization problem, and is advantageous in that constraints and penalty terms can be added to improve predictions so long as the penalty terms are differentiable. In short, gradient descent iteratively updates the parameters away from the gradient of the cost function to

converge at a local minimum with respect to the error. When the cost function is a convex function, formally defined as convex if a line connecting any two points on a graph of a function lies above or on the graph (for example, a bowl shaped graph is convex), then if a local minima exists it is also guaranteed to be the global minima. Thus, convex optimization problems can guarantee unique solutions with gradient descent based optimization as they always converge to local minima given sufficient iterations. Non-convex functions, in which there are lines between points on the graph which go underneath the graph, can have many local minima and thus gradient descent may converge to any of these minima and the solutions are not necessarily unique. Note that the basic least squares optimization problem for linear unmixing is a convex problem, and a graph of the cost function indeed looks like a bowl when visualizing three dimensions of the least squares cost function. Of course, there are usually more than three dimensions in a linear unmixing problem, but the cost function is still convex even if it cannot be visualized in higher dimensions. However, as we introduce sparsity promoting penalty terms and dispersion models into the optimization problem, we will deal with non-convex optimization problems.

Note that we also wish to constrain our optimization programs to produce realistic abundance predictions. In particular, the abundances should be non-negative and sum-to-one since they are treated as percentages and it does not make sense to have negative percentages of a material. Constraints are implemented by adding penalty terms to a cost function via the Karush-Kuhn-Tucker approach [9]. Since these additional constraint penalties added to the least squares cost function are also differentiable, the cost function can still be solved with gradient descent based optimization that converges to solutions which meet the prescribed equality (sum is equal to one) and inequality (each element is greater than zero) constraints.

Sparse Spectral Unmixing: It is desired to encode prior knowledge of sparsity

(few materials present) into spectral unmixing algorithms. Adding penalty terms to the cost function of the spectral unmixing optimization problem is one method to produce sparse abundance predictions. These penalty terms are commonly the norm of the abundance vector, where the norm is simply assigning a real number to a vector according to a formula. The general formula for an L_p norm, where p can be replaced with a non-negative number, is defined as $||\mathbf{x}||_p := (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$. Note that when p is between 0 and 1, it breaks the mathematical definition of norm but is still useful to consider for sparse spectral unmixing. Also note that the L_∞ norm is simply the maximum element of the vector, $||\mathbf{x}||_\infty := \max_i |x_i|$.

It is well understood that regularization by adding the L_1 norm to the cost function promotes sparse solutions [72, 39]. This method, referred to as least absolute shrinkage and selection operator (LASSO), has been applied to spectral unmixing by Iordache et al. [40]. However, it is known that L_1 regularization has a conflict with the sum-to-one abundance constraint of the spectral unmixing problem. In order to maintain this constraint, L_p regularization (with $p < 1$) has been proposed for spectral unmixing, although it makes the problem non-convex [14]. Regularizing with the inverse of the L_∞ norm has been proposed as a convex method of promoting sparsity when the L_1 norm is constrained [59] and has not been tried for spectral unmixing. We compare these methods experimentally to find the best method for sparse regularization in spectral unmixing, and find that the L_p method offers the best performance despite the non-convexity, although the L_∞ method also improves performance.

Spectral Variability: Spectral variability has been a topic of recent interest [81, 10]. One approach is to augment the endmember library $\mathbf{A}$ with multiple variations or spectra for each endmember. To do this, multiple endmember spectral mixture analysis (MESMA) [63] and multiple-endmember linear spectral unmixing (MEL-SUM) [22] both require labeled data of the spectral variation for each endmember.

In contrast, unsupervised techniques learn endmember sets from unlabelled hyperspectral images, including semi-automated techniques [6], k-means clustering [6], and the sparsity promoting iterated constrained endmember algorithm (SPICE) [79, 80] which simultaneously finds endmember sets while unmixing for material abundances. These techniques are limited by the amount of sets in the endmember library, and computational complexity increases with more additions. Our method by contrast finds an efficient parameter set to physically model the spectral variation.

Another category of endmember variability techniques models the endmember spectral variation as samples from a multivariate distribution $\mathbf{P}(\mathbf{e}|\theta)$ where $\mathbf{e}$ is the endmember spectra, and θ are the distribution parameters. Common statistical distributions proposed include the normal compositional model [68], Gaussian mixture models [86], and the beta compositional model [23]. These distribution models have large capacity to model spectral variations, however sometimes they can render endmember spectra that are not physically realistic.

Deep Learning for Hyperspectral Classification and Unmixing: Deep learning has recently improved many hyperspectral imaging tasks [83, 49]. Deep learning refers to computational models which connect artificial neurons in a hierarchical pattern and optimizes the weights of the connections between these neurons to produce a desired mapping from input to output (for example, to predict an abundance vector given an input spectrum vector). To produce this mapping, the weights of the neuron connections are optimized given a training dataset of labelled input-output pairs (for example using stochastic gradient descent optimization [11]). While this shallow description only scratches the surface of neural networks and deep learning, a complete explanation requires lengthy descriptions of a long list of concepts combining statistics, optimization, and machine learning for which we refer readers to the text by Lecun et al. [46].

In spectral unmixing, neural networks have been proposed to process hyperspectral pixel vectors using both deep belief networks [52] and convolutional neural networks (CNNs) [38, 16, 77]. For spatial hyperspectral data, CNNs [17], joint spectral-spatial feature extraction [85], and 3D CNNs [51] have been used used. All these methods require large hyperspectral datasets that are annotated correctly, which is currently a bottleneck for spectral unmixing. One of our methods, which we refer to as analysis-by-synthesis, does not require training data and can be directly applied to observed spectra by performing an alternating optimization using physics-based models. This analysis-by-synthesis technique avoids low training data issues, but we also show that the generative physics models can be made complementary to deep learning architectures as we show in our inverse rendering CNN.

Differentiable Programming and Rendering: Differentiable programming refers to writing algorithms which can be fully differentiated end-to-end using automatic differentiation for any parameter [74, 75, 7]. This has been applied for audio [26] and 3D geometry processing [62]. In graphics, differentiable rendering has improved ray tracing [50, 56, 53, 82], solved analysis-by-synthesis problems in volumetric scattering [33, 32], estimated reflectance and lighting [1], and performed 3D reconstruction [73]. We write a forward imaging model utilizing the physics of dispersion in spectral variation to allow our pipeline to be differentiable end-to-end.

Chapter 3

PRODUCING SPARSE ABUNDANCE MAPS

A primary usage of spectral unmixing is to produce maps showing the fractional abundance of each material present on a planetary surface. The spatial distribution of materials is expected to be sparse - only a few materials should be present in each pixel of the scene. However, previous studies using remote sensing data [2, 27, 25] and laboratory data [60, 28, 19, 5, 3] have noted that spectral unmixing algorithms tend to incorrectly predict dense spatial distribution of materials. That is, the abundance maps produced from remote sensing data showed many materials present in every pixel of the scene, contrary to expectation. The laboratory data, where ground truth was available, confirmed that these dense predictions are erroneous. In this chapter, we explore methods to encode prior knowledge of sparsity into spectral unmixing algorithms through regularization. Further, we use synthetic, laboratory, and remote sensing data to validate which regularization terms yield the best results.

3.1   Methods

3.1.1   Regularization Terms

The general form of an L_p norm for any p is simply assigning a real number to a vector according to the formula $||\mathbf{x}||_p := (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$. Note that the L_1 norm is simply the sum of the absolute value of each element of $\mathbf{x}$. In spectral unmixing, the L_1 norm is usually constrained to be equal to one since it does not make sense to have more or less than 100% of total materials.

The most common sparsity promoting technique in signal processing, referred to as the least absolute shrinkage and selection operator (LASSO), penalizes the

cost function of an optimization problem with the L_1 norm of the variable under optimization [72]. However, if L_1 norm is constrained to 1, it is an ineffective penalty term for minimization.

In proxy to penalizing with the L_1 norm, penalizing with the L_p norm where $p < 1$ and with the inverse L_∞ norm are also known to provide sparse solutions and do not conflict with the constraints [59, 14]. Note that the L_p norm with $p < 1$ has previously been tried on spectral unmixing but is non-convex, and thus is prone to converging to local minima. On the other hand, the inverse L_∞ is convex and had not been tried on spectral unmixing before.

Adding a penalty term to the linear unmixing objective function results in the following optimization problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{b} - \mathbf{A}\mathbf{x}||_2^2 + \lambda\mathbf{\Gamma}(\mathbf{x})$$

$$\text{subject to} \quad ||\mathbf{x}||_1 = 1, \quad \mathbf{x} \geq 0.$$

(3.1)

Where $\hat{\mathbf{x}}$ are the predicted abundances that minimize the cost function, $\mathbf{x}$ are the abundances of the linear mixing model, $\mathbf{\Gamma}(\mathbf{x})$ is a penalty that promotes sparse abundance predictions, $\mathbf{b}$ is the observed spectra, and $\mathbf{A}$ is the endmember library. Also note that the term $||\mathbf{b} - \mathbf{A}\mathbf{x}||_2^2$ is simply the squared L_2 norm, equivalent to the mean-squared error, which is commonly referred to as a least-squares optimization problem. For the L_∞^{-1} method, $\mathbf{\Gamma}(\mathbf{x}) = \frac{1}{||\mathbf{x}||_\infty}$. For the L_p method, $\mathbf{\Gamma}(\mathbf{x}) = ||\mathbf{x}||_p$ and $p < 1$. For the LASSO method, $\mathbf{\Gamma}(\mathbf{x}) = ||\mathbf{x}||_1$. Also for LASSO, the sum-to-one constraint is removed during optimization and the outputs are normalized to respect the constraint after optimization. For all methods, λ is the weight of the regularization

The L_∞^{-1} method as expressed in Equation (3.1) is not convex. However, it can be rewritten as n convex problems, where n is the number of endmembers, expressed by the following optimization problem:

14

$$\hat{\mathbf{x}} = \min_{i \in [n]} \{\arg\min_{\mathbf{x}} ||\mathbf{b} - \mathbf{A}\mathbf{x}||_2^2 + \frac{\lambda}{x_i}\}$$

$$\text{subject to:} \quad ||\mathbf{x}||_1 = 1, \quad \mathbf{x} \geq 0.$$

(3.2)

The proof that the L_∞^{-1} method can be globally solved by the convex problems written in equation (3.2) is shown in [59].

Recall from the chapter 2 that the basic least squares optimization problem, without penalty terms, is a convex function so gradient descent will converge to a global minimum. However, since the L_p norm with $p < 1$ is a non-convex function, because it breaks the formal definition that a line between points on it's graph fall below the graph. Therefore, the entire cost function is non-convex, since the sum of convex and non-convex functions are non-convex in general. Thus, the results will be sensitive to the initialization because gradient descent based optimization can converge to a local minima rather than the global minimum, and thus are not unique. In the reported results, $\mathbf{x}$ was initialized such that all elements were set to $\frac{1}{n}$ where n is the number of endmembers.

3.1.2 Implementation Details

The baseline method, referred to as fully constrained least squares (FCLS) [36], solves the least squares optimization problem with the non-negativity and sum-to-one constraints but without any regularization. This baseline method is compared against the LASSO , L_∞^{-1}, and L_p regularization techniques. All of these methods are are implemented using the Sequential Least SQuares Programming (SLSQP) method in the SciPy optimization toolkit [42]. Note that the λ term in Equation (3.1) that gives weight to the penalty term needs to be input to the algorithm to give the appropriate balance between the sparse penalty and the least squares error. This type of parameter is referred to as a hyper-parameter because it cannot be optimized
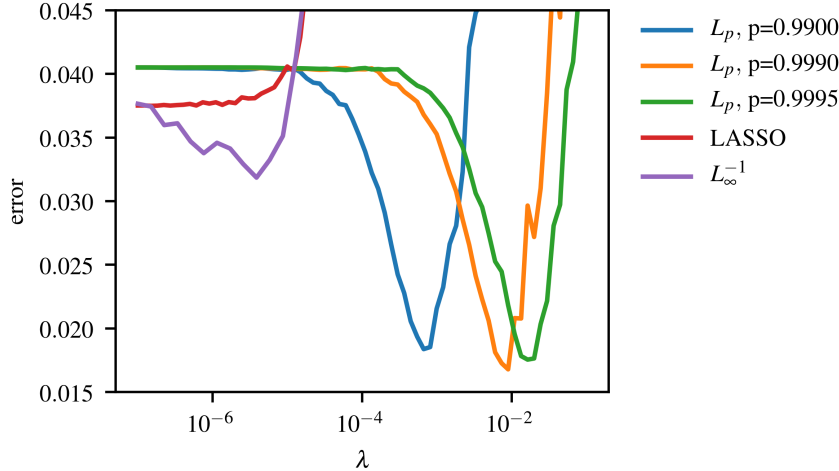
Figure 3.1: Hyperparameter Sweep: The hyperparameters for the regularization weight, λ, is swept through a range of values for each method. The L_p norm has a second hyper-parameter for the value of p which is also swept across several values. The error is calculated by performing the unmixing algorithms on 1000 synthetic mixtures and then is repeated for each trial value of the hyper-parameters. This error vs. parameter value is plotted for the methods under consideration. The optimal hyper-parameters for each method are chosen that minimize the error.

by the algorithm. The hyper-parameters for λ, as well as the p in the L_p norm, are determined by sweeping across a range of values and choosing the parameter that minimizes the average error in an experimental setup using synthetically generated spectral mixtures as shown in Figure 3.1. However, as the amount of noise varies, the optimal parameters do change since there needs to be less weight given to the mean squared error to avoid unwanted fitting to the noisy signal, which is commonly referred to as overfitting.

3.2   Experimental Results

3.2.1   Datasets

We utilize three different datasets to compare the sparsity promoting regularization techniques. The datasets range from synthetic, to laboratory, to remote sensing datasets in the infrared.

Synthetic Mars Dataset: Pure endmembers from the ASU spectral library [19] are combined linearly to form synthetic mixtures that simulate Mars mixtures. The proportions of minerals are meant to simulate three regions of the Martian surface, according to the distrubtion shown in Table 3.1.

The mixtures are perturbed with Gaussian noise to simulate measurement error in spectrometers. The amount of noise is calculated by comparing against calibration results from an exemplary spectrometer, the Osiris-Rex Thermal Emission Spectrometer (OTES) [21], which showed that the variance of the noise due to measurement error was $\sigma < 1e - 8$. In the synthetic data experiments, several tests are repeated where the variance of the noise is increased above this amount.

Feely dataset: We utilize 90 samples from the Feely et al. dataset [28] of thermal emission spectra in the infrared for various minerals measured in the lab. Ground truth was determined via optical petrography [28], and a labeled endmember library is provided. Note that there is approximately 5% error in labelled abundances of this dataset from human error.

Mars TES dataset: The remote sensing data set is provided by the Mars Thermal Emission Spectrometer (TES) measurements [20]. The block diagram representation of raw data processing is shown in Figure 3.2.

For these experiments, the raw TES data is used, and the atmospheric components must be accounted for. To boost efficiency, atmospheric correction and spectral

Table 3.1: Synthetic Data Generation: Synthetic data is generated by averaging endmember spectra together with plausible proportions. Three regions of Mars are simulated, with a probability of sampling each region which roughly approximates the distribution of Mars. In each region there is a different range of proportions for each category of materials. After the proportions are randomly selected from a uniform distribution, they are normalized to follow the sum-to-one constraint.

Region	Feldspar Rich	Hematite Rich	Rare
Probability of region	0.85	0.1	0.05
Hematite	0	10-30%	0
Feldspar	10-40%	0-30%	0-30%
Pyroxene	10-40%	0-30%	0-30%
Olivine	0-20%	0-10%	0-10%
High Si Phases	0-20%	0-10%	0-10%
Carbonate	0-10%	0-10%	0-10%
Sulfate	0-10%	0-10%	0-10%
Other	0%	0-10%	10-40%

unmixing are performed in a single optimization problem. We show that atmospheric correction can be solved simultaneously by the following argument: The observation **b** can be expressed as:

$$\mathbf{b} = \mathbf{A}^*\mathbf{x}^* + \mathbf{A}'\mathbf{x}' + \eta. \tag{3.3}$$

Where the endmember library **A** can be split into two libraries, $\mathbf{A}^*$ and $\mathbf{A}'$ for atmospheric and surface mineral endmembers with corresponding abundances $\mathbf{x}^*$ and $\mathbf{x}'$. This is valid because the atmospheric components have been shown to combine

18
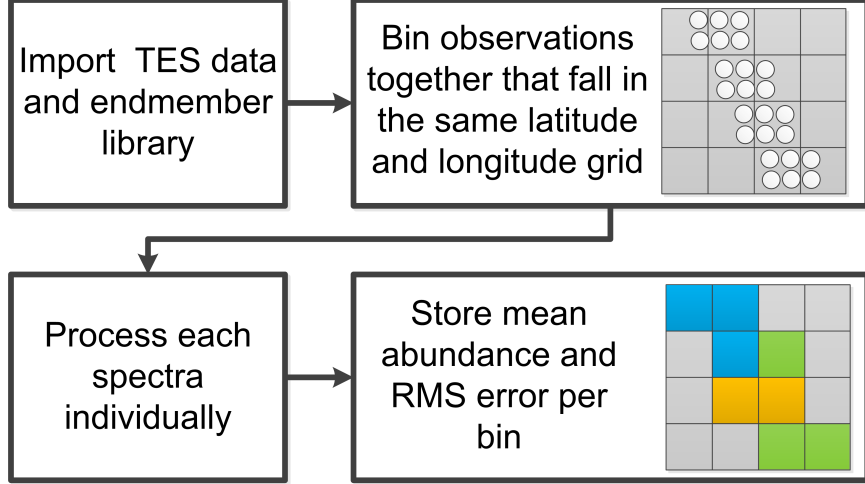
Figure 3.2: TES Data Processing: Block diagram showing the process for raw TES data processing.

extremely linearly [4]. Then, $\mathbf{x}^*$ and $\mathbf{x}'$ can be solved for by the following optimization problem:

$$\underset{\mathbf{x}^*,\mathbf{x}'}{\arg\min} \, ||\mathbf{b} - \mathbf{A}^*\mathbf{x}^* - \mathbf{A}'\mathbf{x}'||_2^2 \tag{3.4}$$
$$\text{subject to}: \quad ||\mathbf{x}'||_1 = 1, \quad \mathbf{x}' \geq 0.$$

Bandfield et al. found the residual signal by subtracting the atmospheric components to determine the surface emissivity, $\mathbf{b} - \mathbf{A}^*\mathbf{x}^* = \mathbf{A}'\mathbf{x}' + \eta$, prior to unmixing a second time to solve for the surface mineral abundances [2, 27]. However, unmixing the residual to solve for $\mathbf{A}'\mathbf{x}'$ results in an equivalent optimization problem to Equation (3.4), so only the first optimization is necessary.

3.2.2 Spectral Unmixing Results

Performance metrics: The performance on the datasets with ground truth is evaluated by the mean squared error (MSE) between the predicted and ground truth abundances. It is also interesting to measure the accuracy, precision, and recall of

19

detecting mineral presence, to show that even small improvements in error can make a big difference in correctly determining which minerals are present. The presence is determined when the abundance is above a threshold, set to 1% in the reported results. The predicted presence is compared against ground truth and uses the standard convention for true positive t_p, true negative t_n, false positive f_p, and false negative f_n to compute accuracy, precision, and recall.

Synthetic Dataset Results: The qualitative behavior of sparse regularization on a single synthetic sample is shown in Figure 3.3. The results summarized in Table 3.2 shows the performance of each regularization taken on 1000 mixtures with noise added where $\sigma = 2.5e - 5$. Table 3.3 shows the performance of 1000 mixtures where $\sigma = 2.5e - 4$. The hyper-parameters of each algorithm were optimized for each of these experiments where $p = 0.999$, $\lambda_{L_p} = 1e-2$, $\lambda_{L_\infty^{-1}} = 1e-5$, and $\lambda_{\text{LASSO}} = 1e-6$. The results show that the L_p method has the lowest error and the L_∞^{-1} method has the second lowest error.

Table 3.2: Synthetic Results with Low Noise: Performance metrics for sparse regularization techniques on the synthetic dataset with low noise.

	FCLS	LASSO	L_∞^{-1}	L_p
MSE	4.1e-04	8.2e-04	2.6e-04	**2.2e-04**
accuracy	0.9973	0.9935	0.9986	**0.9989**
precision	0.9887	0.9719	0.9946	**0.9956**
recall	0.9979	0.9976	**0.9984**	**0.9984**

The results show that the L_p method produces the endmember predictions with the lowest error, and also increases accuracy and precision. FCLS had a higher recall for some results, because it is not penalized for including components, which increases
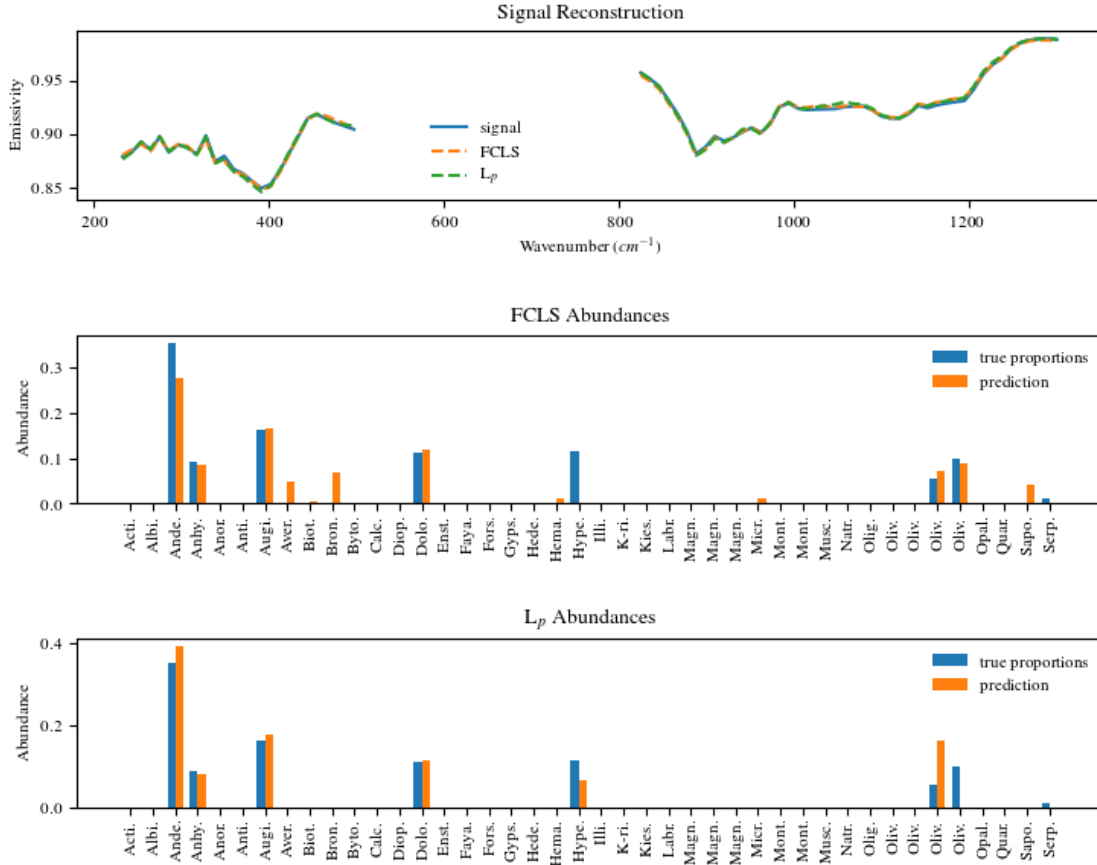
Figure 3.3: Qualitative Analysis of Sparse Regularization: A comparison of FCLS and L_p regularization results for unmixing a single synthetic sample. Note that both signal reconstructions are almost perfect, but the predicted abundances vary. The L_p regularization predicts fewer materials present in the material, driving many of the erroneous predictions to zero. It also correctly finds Hypersthene which FCLS misses. It only predicts one Olivine rather than the combination of two different Olivine, although this negatively impacts the the MSE metric, it can be advantageous to have fewer, more confident, predictions.

chances of recalling the support endmembers.

Feely dataset results: The Feely results are shown in table 3.4. All methods

Table 3.3: Synthetic Results with High Noise: Performance metrics for sparse regularization techniques on the synthetic dataset with higher noise.

	FCLS	LASSO	L_∞^{-1}	L_p
MSE	1.6e-04	5.0e-04	8.5e-05	**5.1e-05**
accuracy	0.9940	0.9908	**0.9963**	0.9959
precision	0.9768	0.9633	0.9873	**0.9884**
recall	**0.9929**	0.9927	0.9928	0.9895

had significantly more error on the laboratory mixture experiment than the synthetic mixture experiment. Still, the L_p method performed better than other regularization methods by lowering the error and increasing the accuracy and precision. FCLS has the highest recall because it is not penalized for including extra components.

Table 3.4: Feely Dataset Results: Performance metrics for sparse unmixing on the Feely laboratory dataset.

	FCLS	LASSO	L_∞^{-1}	L_p
MSE	0.121	0.116	0.110	**0.098**
accuracy	0.7544	0.7421	0.7699	**0.8016**
precision	0.4908	0.4773	0.5171	**0.5529**
recall	**0.8131**	0.8004	0.8015	0.7303

Mars TES dataset results: Because of the large amount of data to be processed, only the L_p method and the FCLS methods were used to generate Mars TES mineral maps. Figure 3.4 shows the amount of materials predicted above 1% present in each pixel of the map (at 1 pixel per degree). As expected, the L_p encourages sparsity

and limits the amount of materials predicted in each scene. The L_p method, with $p = 0.98$ and $\lambda = 0.01$, predicted an average of 11.7 components per pixel and the FCLS method predicted an average of 14.2 components per pixel. This shows that sparse regularization is useful on real data for limiting the number of materials predicted, to better match realistic material distributions. Note that no one knows the exact count of minerals that should be present in each spatial region of Mars, and this is only a qualitative assessment based on the expected distribution of minerals.

Noise sweep results: A noise sweep is run on 100 mixtures, by repeating spectral unmixing algorithms on the same synthetic mixtures and increasing the noise power to observe the effect on error. In the noise sweep, the hyper-parameters are held constant (λ and p optimized for $\sigma^2 = 2.5e\text{-}4$, which is in the middle of the noise sweep). Figure 3.5 shows the error of each method as noise is increased. The results show that the L_p method has the lowest error across all noise levels, followed by the L_∞^{-1} method.

Execution Time: The execution time of the algorithms was measured. The results are summarized in Table 3.5 with varying number of endmembers, n. As expected L_∞^{-1} is the slowest algorithm, since it has to solve n optimization problems, and the other methods are roughly equivalent.

3.3 Discussion

In this chapter we explored sparse regularization techniques for improving spectral unmixing results. We compared methods which had already been proposed for spectral unmixing, FCLS, LASSO and the L_p norm, and also introduced the L_∞ norm which had not been previously applied to spectral unmixing.

We validated these techniques against synthetic, laboratory, and remote sensing datasets in the infrared. In particular, we examined linear unmixing although the
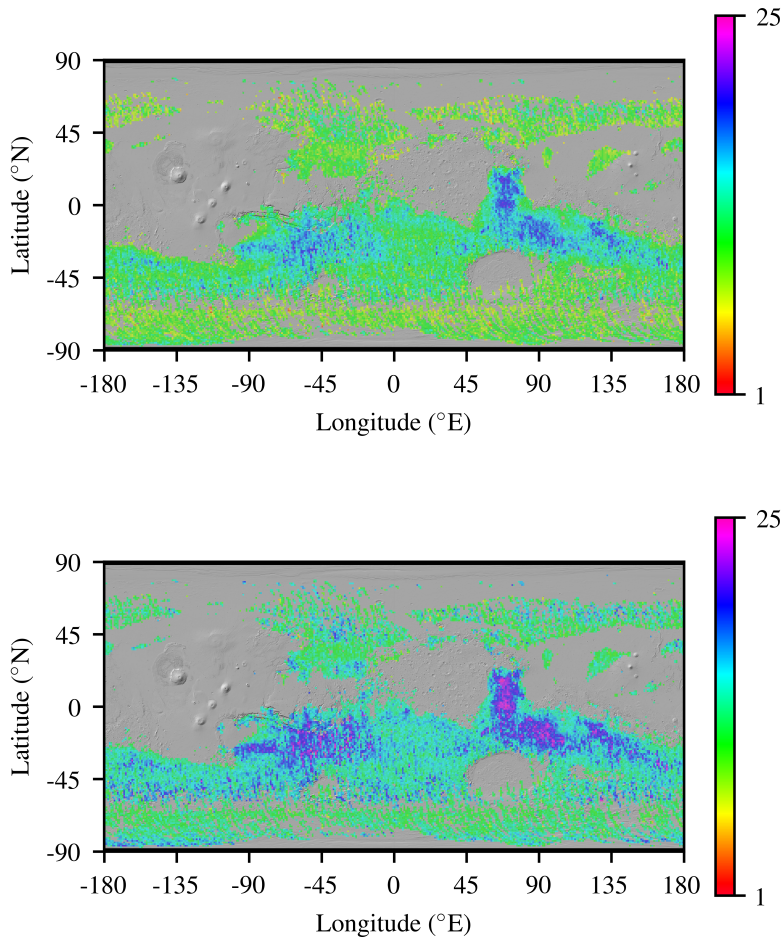
Figure 3.4: Sparsity Maps: The sparsity maps show the count of endmembers predicted above a threshold, set to 1%. The L_p method (top) consistently uses fewer endmembers than FCLS method (bottom), showing that the sparse regularization is working.

techniques could be extended to non-linear unmixing algorithms by adding the proposed penalty terms to cost functions for non-linear unmixing algorithms.

We found that sparse unmixing does indeed improve results, and that L_p regularization seems to be the best method. Regularization with the L_p norm reduces the error, increases the accuracy and precision, and uses fewer endmembers in predic-
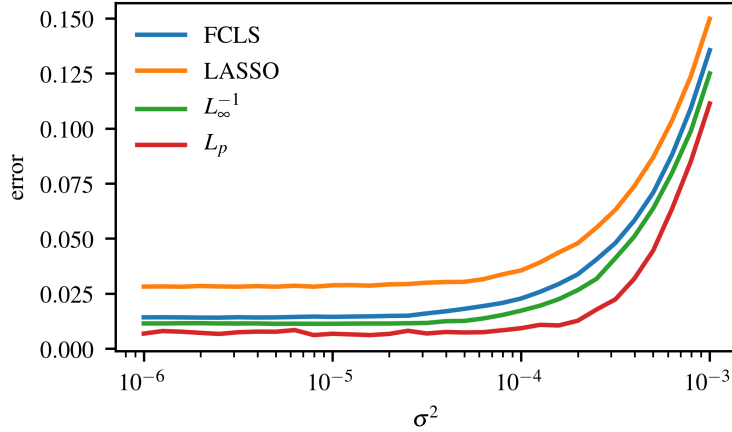
Figure 3.5: Noise Sweep: Sparse regularization performance is evaluated on 100 synthetic mixtures. The noise is varied by several orders of magnitude. The L_p method produces the lowest error predictions across all noise levels.

Table 3.5: Execution Time Comparison: The execution times of sparse regularization methods are compared. The L_∞^{-1} method is much slower than the other methods.

n	FCLS	LASSO	L_∞^{-1}	L_p
5	1 ms	1 ms	5 ms	1 ms
15	5 ms	4 ms	39 ms	4 ms
25	8 ms	4 ms	108 ms	8 ms
50	12 ms	12 ms	607 ms	20 ms

tions. The L_∞^{-1} method also showed some performance benefits, but can be an order of magnitude slower. The L_p method has approximately the same speed as FCLS and LASSO and has the overall best performance.

Further work on solving spectral variation in Chapter 4 incorporates sparse regularization into the end-to-end pipeline. The performance benefits of accounting for both sparsity and spectral variation further improves the results.

25

We hope that this chapter inspires closer looks into remote sensing data by the geoscience community to produce more realistic abundance maps with sparse regularization. Although no one knows the exactly how sparse the abundance predictions should be, providing a "dial" to tune the amount of sparsity can hopefully lead to valuable insights.

Chapter 4

PHYSICS-BASED SPECTRAL VARIABILITY

Pure materials cannot be represented by a single characteristic spectrum, because their absorption bands can change in strength, shape, and frequency due to factors such as different grain sizes [55, 65, 60, 61] or differing ratios of molecular bonds [12, 70] as shown in Figure 4.1. This is problematic to existing spectral unmixing algorithms, because both linear and non-linear mixing models assume that the pure material spectra is known [81, 23, 86]. The objective of this chapter is to find a model that accounts for the absorption band shifts in a physically plausible way. That is, we seek a generative model for spectra with "dials" to tune the frequency, strength, and shape of absorption bands. From the literature on analysis of the formation of spectra from an atomistic perspective [67, 76], we find that the Lorentz-Lorenz dispersion model is the correct approach to take. However, unlike previous works we go further than using the model to derive optical properties of materials, we also incorporate the dispersion model into an end-to-end spectral unmixing pipeline that allows the parameters to be fine-tuned via differentiable programming to account for spectral variability.

We incorporate a physics-based dispersion model with parameters that control the absorption bands into a spectral unmixing algorithm. This allows our algorithm to find the endmember variation which best fits the data, but constrains these predictions to be physically plausible. These variations of endmember spectra also provide additional science data, as the variation of absorption bands can reveal properties about the composition and history of the material. To our knowledge, we are the first to use this approach of optimizing a parametrized physical model to generate spectra
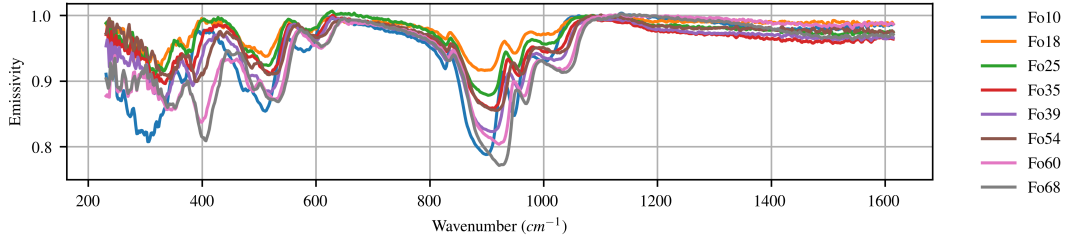
Figure 4.1: Endmember Variation: Several spectra of olivine are plotted to demonstrate it's spectral variability. The olivine mineral is a solid solution with continuous compositional variation of Iron (Fe_2) and Magnesium (Mg_2) bonds This ratio of bonds which is indexed by the Fo number and varies from Fo 100 (pure Forresterite with only Mg bonds) to Fo 0 (pure Fayalite with only Fe bonds), causes absorption bands to shift in frequency and strength.

within a spectral unmixing algorithm, simultaneously solving for the abundances and the variance in absorption bands.

4.1 Method

We account for spectral variability in hyperspectral unmixing with two main components: (1) use of a physically-accurate dispersion model for pure endmember spectra, and (2) a differentiable programming pipeline to perform spectral unmixing. This approach has synergistic benefits of leveraging prior domain knowledge while learning from data. Our first algorithm, which we refer to as analysis-by-synthesis, solves spectral unmixing in a self-supervised fashion by generating endmember spectra with the dispersion model and using optimization to solve for the proportional abundance of each spectra as well as optimizing the parameters of the dispersion model to adjust the absorption bands. Further, we show how to inversely rendering the observed spectra via a convolutional neural network (CNN) to predict both abundances and dispersion parameters, with the dispersion model in the loop to "render" the spectra.

This method improves performance when training data is available and is computationally fast in implementation compared to analysis-by-synthesis.

4.1.1 Dispersion Model

We first describe the dispersion model for generating endmember spectra. Endmember and/or endmember spectra is what we call the spectral curve for emissivity ϵ as a function of wavenumber ω. Each pure material has a characteristic endmember spectrum, although the absorption band can have subtle shifts in frequency and shape, which is the problem we are trying to solve/disambiguate. Let $\varepsilon^{\text{measured}}(\omega)$ be endmember spectra we have measured, typically in a lab or in the field, whose emissivity is sampled at different wavenumbers: $\left[\varepsilon^{\text{measured}}(\omega_1), \cdots \varepsilon^{\text{measured}}(\omega_N)\right]^T$. Our goal is to propose a model $\varepsilon^{\text{model}}(\mathbf{\Lambda}; \omega)$ with parameters $\mathbf{\Lambda}$ such that the following loss is minimized: $L(\mathbf{\Lambda}) = \sum_{i=1}^{N} \left(\varepsilon^{\text{measured}}(\omega_i) - \varepsilon^{\text{model}}(\mathbf{\Lambda}; \omega_i)\right)^2$. That is, we fit the model emissivity of an endmember spectrum to the measured spectrum. In practice, we need to add regularization and constraints to this endmember loss for better fitting that we describe after the derivation of the dispersion model.

Our model of endmember spectra is derived from an atomistic oscillator driven by electromagnetic waves impinging on the molecular structure of the pure material [67, 45]. In Figure 4.2, we show a conceptual diagram of this model, and how it generates emissivity curves as a function of wavelength. For the full derivation of the model from first principles, we refer the reader to Appendix A in the supplementary material. Instead, we outline the model below based on the equations derived from that analysis.

Let $\mathbf{\Lambda} = [\boldsymbol{\rho}, \boldsymbol{\omega_o}, \boldsymbol{\gamma}, \boldsymbol{\epsilon_r}]$ be a matrix of parameters, where $\boldsymbol{\rho}, \boldsymbol{\omega_0}, \boldsymbol{\gamma}, \boldsymbol{\epsilon_r} \in \mathbb{R}^K$ and K is a model hyperparameter corresponding to the number of distinct mass-spring equations used to model the emissivity. $\boldsymbol{\rho}$ is the band strength and as it increases the absorption band becomes deeper. $\boldsymbol{\omega_o}$ is the resonant frequency and as it increases

Molecular Vibration

Atomistic oscillator model

$$qE_0 e^{i\omega t} = m\frac{d^2 x}{dt^2} + R\frac{dx}{dt} + G \cdot x(t)$$

Relation to optical constants

$$n^2 - k^2 = \epsilon_v + \sum_i \frac{4\pi q^2 f_i}{m_i} \frac{\omega_i^2 - \omega^2}{(\omega_i^2 - \omega^2)^2 + \gamma_i^2 \omega^2}$$

$$2nk = \sum_i \frac{4\pi q^2 f_i}{m_i} \frac{\gamma_i \omega}{(\omega_i^2 - \omega^2)^2 + \gamma_i^2 \omega^2}$$

Generation of spectra
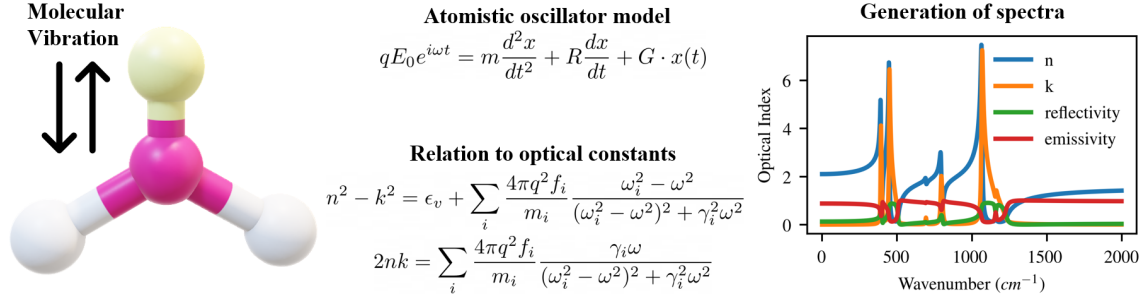
Figure 4.2: Dispersion Model Conceptual Diagram: The insight of the dispersion model is that optical properties can be related to molecular structure through first principles via an atomistic oscillator model. We use this generative model for the formation of spectral variation in our spectral unmixing algorithm.

the absorption band shifts in wavenumber (and also slightly shifts the shape). $\boldsymbol{\gamma}$ is the frictional force (dampening coefficient) and controls the shape/width of the absorption bands. $\boldsymbol{\epsilon_r}$ is relative dielectric permeability and as it increases the entire emissivity curve is shifted downwards. Also note that absorption bands which are close to each other interact in highly non-linear ways. Note: usually $\boldsymbol{\epsilon_r}$ is a constant vector which does not vary with K. In Figure 4.3 we see the effect of perturbing these parameters, and their effect on the strength, shape, and frequency of the absorption bands. This level of control is exactly what we aimed to build into our spectral unmixing algorithm to allow for realistic variation in our endmember library.

In the full derivation of the model shown in Appendix A, optical indices are related to the dispersion parameters from first principles. The refractive index terms n, k are given as follows [67, 45]:

$$n(\boldsymbol{\Lambda};\omega) = \sqrt{\frac{\theta + b}{2}}, \quad k(\boldsymbol{\Lambda};\omega) = \frac{\phi}{n(\boldsymbol{\Lambda};\omega)}, \tag{4.1}$$

where the expressions for θ, b, ϕ are given as follows:

$$\theta = \epsilon_r + \sum_{k=1}^{K} 4\pi \rho_k \omega_{0_k}^2 \frac{(\omega_{0_k}^2 - \omega^2)}{(\omega_{0_k}^2 - \omega^2)^2 + \gamma_k^2 \omega_{0_k}^2 \omega^2}, \tag{4.2}$$

30
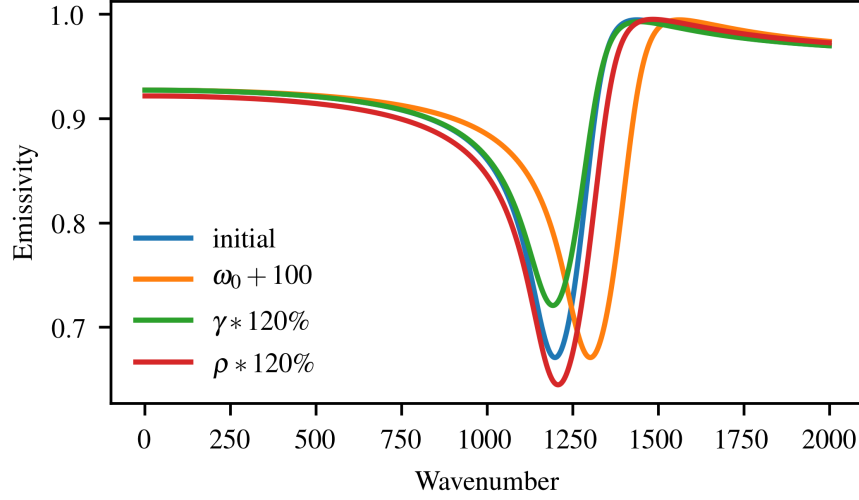
Figure 4.3: Dispersion Parameter Variation: A single absorption band is initialized with $\epsilon_r = 2.356, \omega_0 = 1161, \gamma = 0.1, \rho = 0.67$. Then the parameters are perturbed such that ω_0 is increased by 100, γ is increased to 120%, and ρ is increased to 120%. The plots show the effect of changing each parameter individually to show it's control over the shape and width of the absorption band.

$$b = \sqrt{\theta^2 + 4\phi^2}, \quad \phi = \sum_{k=1}^{K} 2\pi \rho_k \omega_{0_k}^2 \frac{\gamma_k \omega_{0_k} \omega}{(\omega_{0_k}^2 - \omega^2)^2 + \gamma_k^2 \omega_{0_k}^2 \omega^2}. \tag{4.3}$$

We note that subscript k denotes the k-th coordinate of the corresponding vector. Also there is another useful relation (derived in Appendix A) that $n^2 - k^2 = \theta, nk = \phi$. We then define the complex refractive index as $\hat{n}(\boldsymbol{\Lambda}; \omega) = n(\boldsymbol{\Lambda}; \omega) - i \cdot k(\boldsymbol{\Lambda}; \omega)$, where $i = \sqrt{-1}$ is the imaginary number. Hence, we can calculate the emissivity as follows:

$$\epsilon(\boldsymbol{\Lambda}; \omega) = 1 - R(\boldsymbol{\Lambda}; \omega), \qquad \text{where } R(\boldsymbol{\Lambda}; \omega) = \left| \frac{\hat{n}(\boldsymbol{\Lambda}; \omega) - 1}{\hat{n}(\boldsymbol{\Lambda}; \omega) + 1} \right|^2. \tag{4.4}$$

When considering minerals, we introduce $M \in \mathbb{N}$, the number of optical axes of symmetry in crystal structures, (eg. 2 axes of symmetry in quartz [67, 76]), to define the full model:

$$\varepsilon^{\text{model}}(\boldsymbol{\Lambda}; \omega) = \sum_{m=1}^{M} \alpha_m \cdot \epsilon(\boldsymbol{\Lambda_m}; \omega) \quad \text{such that } \sum_{m=1}^{M} \alpha_m = 1, \alpha_m \geq 0, \tag{4.5}$$

where we use a different parameter matrix $\boldsymbol{\Lambda_m}$ and weight α_m for each optical axis.

The dispersion model has been primarily used to analyze optical properties of materials to determine n and k, which then can be subsequently applied to optical models like radiative transfer [67, 76]. After n and k are found, spectra such as reflectance, emissivity, and transmissivity can be generated. In particular, we notice that fine-grained control of the dispersion model parameters can realistically render spectral variation that occurs in hyperspectral data. Our contribution is to leverage these properties in a differentiable programming pipeline for spectral unmixing.

4.1.2   Endmember Fitting

Using the dispersion model presented above, we want to robustly estimate the model parameters to fit the spectra $\varepsilon^{\text{measured}}$ captured in a lab or in the field. To fit the model, we wish to perform gradient descent to efficiently find these parameters. In short, we want to exploit the fact that the negative gradient points in the direction which minimizes the loss function, so we want to iteratively update the parameters away from the gradient to better fit the data and find optimal parameters. Using chain rule on the loss function, we see that $\frac{\partial L}{\partial \boldsymbol{\Lambda}_{ij}} = \frac{\partial L}{\partial \varepsilon^{\text{model}}} \frac{\partial \varepsilon^{\text{model}}}{\partial \boldsymbol{\Lambda}_{ij}}$, where (ij) corresponds to that element of the parameter matrix, and all expressions are scalars once the coordinate is specified. While the partial derivatives can be calculated explicitly via symbolic toolboxes, the resulting expressions are too long to be presented here. For simplicity and ease of use, we use the autograd function [7, 57] in PyTorch [58] to automatically compute derivatives for our model as we are performing backpropagation. Note that this produces the analytic derivative, not an estimated derivative, by pre-programmed derivatives of simple math operations available in the Pytorch library and applying the chain rule to find derivatives of larger more complicated functions. Since the entire dispersion theory can be expressed as a chain, or compu-

tational graph, of simple math operations, it is possible to find the analytic derivative of each parameter with respect to a cost function.

One main challenge in performing endmember fitting is that the dispersion model is not an injective function, meaning that more than one set of parameters $\mathbf{\Lambda}$ can result in the same generated spectrum, and thus the parameters cannot be uniquely identified given a target material spectrum. This can be solved, in-part, through sparse regularization which adds penalty terms to the cost function of the optimization problem to encourage solutions that use fewer parameters, to narrow down the solutions which possibly have the same fit to the target spectra. This is especially applicable since a preference for fewer dispersion parameters has been suggested in literature on finding dispersion parameters which fit a target spectra [67, 76]. In our implementation, we initialize our model with $K = 50$ rows of the parameter matrix. Since the parameter ρ controls the strength of the absorption band, small values of ρ do not contribute much energy to the spectra (unnecessary absorption bands), and can be pruned. After performing sparse regression by penalizing the L_1 norm of ρ, K is typically around 10-15 in our experiments.

Thus, our modified sparse regression problem may be written as

$$\underset{\mathbf{\Lambda}_{\min} \leq \mathbf{\Lambda} \leq \mathbf{\Lambda}_{\max}}{\arg\min} \sum_{i=1}^{N} \left(\epsilon^{real}(\omega_i) - \epsilon^{model}(\mathbf{\Lambda}; \omega_i) \right)^2 + \lambda_\rho ||\rho||_1, \tag{4.6}$$

where $\mathbf{\Lambda}_{\min}$ and $\mathbf{\Lambda}_{\max}$ restrict the variation of the dispersion parameters to a plausible range. In addition, endmembers (particularly minerals) can have multiple optical axes of symmetry described by separate spectra, which has been noted in the literature [67, 76]. Without prior knowledge of the number of axes for every material we encounter, we run this optimization for a single and double axes, and pick the one with the lowest error. See Section 4.3 for results on endmember fitting and Figure 4.7 for examples of modelled vs. measured spectra.

Analysis-by-synthesis

Inputs → Outputs

Inputs: Initial Conditions and Target Spectra

$\Lambda^{(0)}$

Dispersion Model → Generate Endmembers

Linear Unmixing
$\arg\min_{\mathbf{x}} ||\mathbf{b} - \mathbf{A}\mathbf{x}||_2^2$

Parameter Update
$\Lambda_{ij} = \Lambda_{ij} - \lambda \dfrac{\partial L}{\partial \Lambda_{ij}}$

Backpropagation
$\dfrac{\partial L}{\partial \Lambda_{ij}} = \dfrac{\partial L}{\partial \epsilon^{model}} \dfrac{\partial \epsilon^{model}}{\partial \Lambda_{ij}}$

Outputs: Modelled parameters, abundances, and spectra

$\Lambda^{out} \quad \hat{\mathbf{x}}$
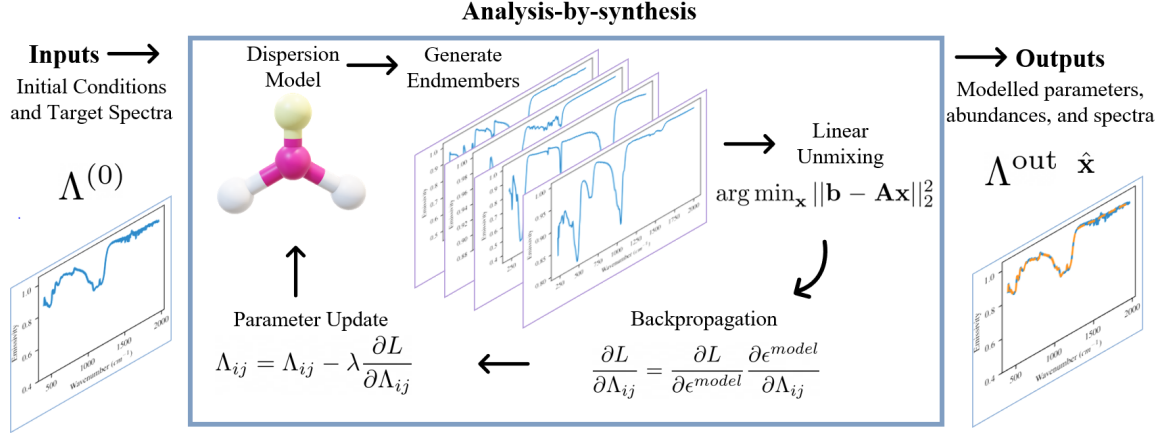
Figure 4.4: Analysis-by-Synthesis: The analysis-by-synthesis algorithm uses differentiable programming to find optimal dispersion parameters and abundances. The initial dispersion parameters and the target spectra are fed as inputs, and the algorithm alternates between optimizing the abundances in the least squares sense and updating the dispersion parameters with respect to the gradient.

This regression problem is non-convex, formally meaning that a line connecting points on the graph of the cost function can fall below the graph, and intuitively meaning that there can be several local minima that gradient descent optimization can converge to resulting in non-unique predictions. However, we still solve the problem using gradient descent with a random initialization; which is known to converge to a local minimum with probability 1 [47]. A global minimum is not necessary at this stage, since we use endmember fitting to provide a good initialization point for the subsequent alternating minimization procedure introduced in the next section 4.1.3.

4.1.3   Differentiable Programming for Spectral Unmixing

4.1.4   Analysis-by-Synthesis Optimization

In Figure 4.4, we show our full end-to-end spectral unmixing pipeline. Here, $\varepsilon^{\text{model}}(\Lambda; \omega)$, which is initially fit to $\varepsilon^{\text{measured}}$, is then aggregated into the columns of $\mathbf{A}$. Then, the observed spectra $\mathbf{b}$ is linearly unmixed by solving a regularized least-squares optimization: $\operatorname{argmin}_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_p$ subject to sum-to-one and non-negativity constraints $\|\mathbf{x}\|_1 = 1, \mathbf{x} \geq 0$. Given these constraints, one cannot impose sparsity with the usual L_1 norm. Instead, we use the L_p norm to induce sparsity for the predicted abundances; this has been proposed before for spectral unmixing [14].

The key to our pipeline is that everything is fully differentiable, and thus we can actually minimize the following equation:

$$\underset{\mathbf{x}, \mathbf{\Lambda} \in [\mathbf{\Lambda}_{min}, \mathbf{\Lambda}_{max}]}{\arg\min} \|\mathbf{b} - \mathbf{A}(\mathbf{\Lambda})\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_p \text{ such that } \|\mathbf{x}\|_1 = 1, \mathbf{x} \geq 0. \qquad (4.7)$$

with respect to both the parameters of the dispersion model $\mathbf{\Lambda}$ and the abundances $\mathbf{x}$. This gives us our recipe for hyperspectral unmixing: first, perform endmember fitting to initialize $\mathbf{A}(\mathbf{\Lambda})$, then, solve Equation 4.7 in an alternating fashion for $\mathbf{x}$ and $\mathbf{\Lambda}$. One could also solve this equation jointly for both unknowns, however, we found that the alternating optimization was faster and converged to better results.

The optimization problem established in equation (4.7) is an alternating minimization problem and is unfortunately not convex [41]. One popular approach to tackle nonconvex problems is to find a good initialization point [24, 8], and then execute a form of gradient descent. Inspired by this, we first initialize $\mathbf{A}(\mathbf{\Lambda})$ by performing endmember fitting using Equation 4.6 as described in Section 4.1.1. Our experiments indicate that this provides a useful initialization for our subsequent step. We then

perform alternating minimization on Equation 4.7 for $\mathbf{x}$ and $\boldsymbol{\Lambda}$. For more details on this, we refer the reader to Appendix C where we discuss on the properties of $\mathbf{A}(\boldsymbol{\Lambda})$ across multiple runs and how they relate to the convergence of the optimization.

In the ideal scenario, this initial matrix $\mathbf{A}(\boldsymbol{\Lambda})$ would consist of the endmember spectra that fully characterizes the mixed spectra $\mathbf{b}$. However, since spectra for similar materials can significantly vary [55, 65, 60, 61, 12, 70] (see Figure 4.1), the initialization can be slightly off and we follow up with (4.7) to obtain a better fit. Note that this optimization problem is solving for the maximum likelihood estimator under a Gaussian noise model. Our optimization technique is performing analysis-by-synthesis, as given a single observation $\mathbf{b}$, the dispersion model synthesizes endmember variation until a good fit is achieved.

4.1.5   Inverse Rendering of Dispersion Model Parameters

The previous analysis-by-synthesis optimization does not require training data in order to perform spectral unmixing. However, there is room for even more improvement by using labeled data to help improve the parameter fitting of the model in the synthesis step. We train a CNN to predict the parameters for a generative model, known as inverse rendering in other domains [78]. In Figure 4.5, we show this inverse rendering conceptually, and how it can be fed into our differentiable programming pipeline for end-to-end spectral unmixing.

Our CNN architecture consists of convolutional layers followed by a series of fully-connected layers. Using a CNN for inverse rendering is significantly faster at test time as compared to the analysis-by-synthesis optimization. However, it does have a drawback of requiring training data which is unavailable for certain tasks/datasets.
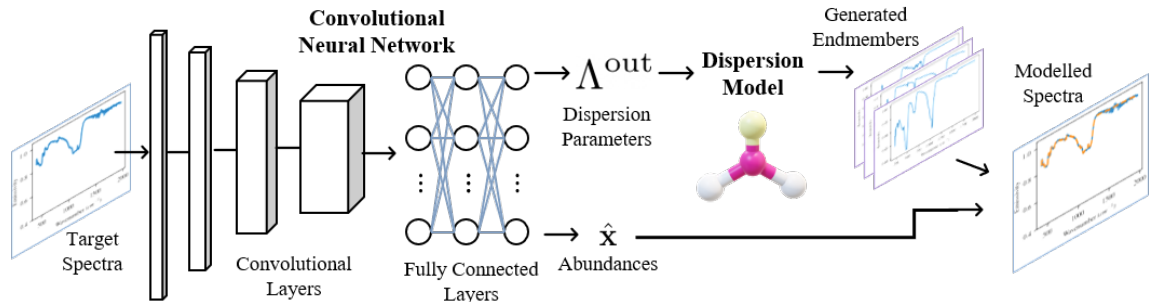
Figure 4.5: Inverse Rendering: A CNN is trained to "inversely render" pixels of the hyperspectral image, by predicting both the dispersion parameters that control the spectral variability, and the abundances that control the mixing model. During training, the reconstruction error is back-propagated through the differentiable dispersion model to boost the performance of the network at making physically realistic predictions.

4.2 Implementation Details

Analysis-by-Synthesis Optimization: For analysis-by-synthesis, the sparse regularization was set with p = 0.95 and λ_p = 0.0001. Dispersion parameters were constrained within a tolerance of their initial conditions with $\rho_{tol} = 0.05$, $\gamma_{tol} = 0.005$, $\epsilon_{tol} = 0.001$, and $\omega_{tol} = 0.0001$. On the Gulfport datasets γ_{tol} and ϵ_{tol} were increased to 0.05 to compensate for increased variation. Analysis-by-Synthesis alternates between finding optimal abundances (solving a regularized least squares problem), and updating the dispersion parameters for 100 iterations using the Adam optimizer with learning rate = 0.01, betas = (0.9, 0.999), and weight decay = 0.

Inverse Rendering CNN Architecture: The input to the inverse rendering CNN is the spectrum (or batch of spectra). There are four convolutional layers with alternating 1x5 and 1x4 kernels and a depth of 3, 6, 12, and 24 kernels in each layer respectively. The convolutional layers also have a rectified linear unit (ReLU) activa-

tions and the last 2 layers have a 1x2 maxpooling layer. The convolutional layers are followed by fully connected layers with 150 hidden units. The final fully connected layer has enough units for the amount of dispersion parameters and abundances depending on the size of the endmember library and number of dispersion parameters per endmember. Then, the dispersion parameters are used to render endmember spectra and the mixture is reconstructed under the linear mixing model with the predicted abundances as inputs. The network only needs the input spectra and the abundances as inputs for training, as the reconstruction error of the spectra is used to back-propagated through the differentiable dispersion model to teach the network to predict good dispersion parameters. Real data (when available) and synthetic data (around 50,000 samples) are used to train the network, which converges after about 100 epochs. An Adam optimizer is used with learning rate set to 1e-3, betas set to (0.9, 0.999), and weight decay set to 0.

4.2.1 Baseline Methods

We compare against several state-of-the-art baselines in the literature. The basic linear unmixing algorithm is Fully Constrained Least Squares (FCLS) [36] which solves least squares with sum-to-one and non-negativity constraints on the abundances. We also implement two state-of-the-art statistical methods for modelling endmember variability as distributions: the Normal Compositional Model (NCM) [68] and the Beta Compositional Model (BCM) [23]. NCM and BCM use a Gaussian and Beta distribution respectively, perform expectation-maximization for unmixing, and require a small amount of training data to determine model parameters.

We also compare against two state-of-the-art deep learning networks by Zhang et al. [84]. The first network utilizes a 1D CNN (CNN-1D) architecture, while the second network utilizes a 3D CNN (but with 1D convolutional kernels) (CNN-3D). CNN-3D is

only applicable to datasets with spatial information, and not testable on the Feely and Gulfport synthetic data. We further created a modified CNN architecture (CNN-1D Modified) to maximize the performance on our datasets by changing the loss function to MSE, removing max-pooling layers, and adding an additional fully connected layer before the output.

Fully Constrained Least Squares: We compare against Fully Constrained Least Squares (FCLS) [36] which is a popular classical unmixing algorithm. FCLS solves for the the (aerial) abundances, $\mathbf{x}$: $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{b} - \mathbf{A}\mathbf{x}||_2^2$ subject to $||\mathbf{x}||_1 = 1, \mathbf{x} \geq 0$. The constraints, referred to as the sum-to-one constraint and the non-negativity constraint, are enforced since abundances are interpreted as percentages.

Normal Compositional Model: The Normal Compositional Model (NCM) [68] is one of the most popular methods for modelling endmember variability via statistical methods. The method requires a small amount of training data (roughly 50 samples per endmember) to learn the mean and variance of reflectivity (or emissivity) of each spectral band, and modelling the variation as a Gaussian distribution. During unmixing, Expectation Maximization is used to simultaneously learn the abundances and the endmember variation, subject to the abundance sum-to-one and non-negativity constraints.

The NCM is run using the Matlab code provided by Du et al. [23]. Training data of about 50 samples of endmember variation were provided to the NCM for each dataset. There are no hyperparameters needed for this method.

Beta Compositional Model: The Beta Compositional Model (BCM) [23] is a more recent method for modelling spectral variability via a statistical method. Similar to the NCM, a small amount of training data are used to learn the beta parameters of each spectral band, and an Expectation Maximization algorithm is used during unmixing. The beta parameters allow each spectral band to be modelled as a more

complex distribution than the NCM and has been shown to increase performance. The BCM is run using the Matlab code provided by Du et al. [23]. Training data of about 50 samples of endmember variation were provided to the BCM for each dataset. For datasets without sufficient endmember samples, we generated synthetic endmember variation with the dispersion model. We search for the optimal hyperparameters through repeated experiments and report the best results. The optimal BCM across all datasets was run with K = 3, $\sigma_V = 100$, and $\sigma_M = 0.001$.

CNN-1D Architecture: A 1x1 hyperspectral pixel is input into the network, four convolutional layers with alternating 1x5 and 1x4 kernels and a depth of 3, 6, 12, and 24 kernels in each layer respectively. All convolutional layers have ReLU activations a 1x2 maxpooling layer. The convolutional layers are followed by a fully connected layer with 150 hidden units, and an output fully connected layer with a size that depends on the number of abundances. Normalization is used to enforce the abundance sum-to-one constraint and the ReLU activation enforces the non-negativity constraint. The CNN is trained to minimize the log loss between the predicted and ground truth abundances. The network converges in about 100 epochs with a learning rate of 1e-3.

CNN-3D Architecture: CNN-3D has an almost identical architecture, although it accepts a 3x3 set of pixels at the input. Although a spatial dimension exists at the input, the convolutions only occur in the spectral dimension. four convolutional layers with alternating 1x5 and 1x4 kernels and a depth of 16, 32, 64, and 128 kernels in each layer respectively. The convolutional layers are followed by a fully connected layer with 150 hidden units, and an output fully connected layer with a size that depends on the number of abundances.

The output abundances are then normalized to enforce the abundance sum-to-one constraint and the ReLU activation of the artificial neurons enforces the non-

negativity constraint. The CNN is trained to minimize the log loss between the predicted and ground truth abundances. The network converges in about 100 epochs with a learning rate of 1e-3.

CNN-1D Modified Architecture: Finally, a modified version of CNN-1D is baselined against to try to find the optimal architecture for performance on our datasets. The first 2 max-pooling layers are removed, an additional hidden fully connected layer with 150 units is added before the output, and a softmax operation is applied to the output to enforce the abundance sum-to-one constraint. The network is trained to minimize the mean squared error between the predicted and ground truth abundances. The network converges in 100 epochs with a learning rate of 1e-3.

4.3 Experimental Results

4.3.1 Datasets

We utilize three separate datasets to validate our spectral unmixing algorithms. In Figure 4.6, we visually represent these datasets and their exemplar data.

Feely Dataset: We utilize 90 samples from the Feely et al. dataset [28] of thermal emission spectra in the infrared for various minerals measured in the lab. Ground truth was determined via optical petrography [28], and a labeled endmember library is provided. The limited amount of data is challenging for machine learning methods, so we utilize the dispersion model to generate 50,000 additional synthetic spectra for dataset augmentation. Note that ground truth for this dataset has about 5% error in each abundance prediction since the optical petrography method is prone to human error.

Gulfport Dataset: The Gulfport dataset from Gader et al. [29] contains hyperspectral aerial images in the VNIR along with ground truth classification labels

segmenting pixels into land types (e.g. grass, road, building). Although the dataset is for spectral classification, it can also be used to benchmark unmixing algorithms by creating synthetic mixtures of pure pixels from the Gulfport dataset with random abundances as done by [23, 86]. We perform both spectral classification (Gulfport) and unmixing (Gulfport synthetic) tasks in our testing. Both datasets are split into a train and test set (although some methods do not require training data), and the training data are augmented with 50,000 synthetically generated mixtures from the dispersion model.

One main difficulty of this dataset is the endmembers identified correspond to coarse materials such as grass and road as opposed to pure materials. Such endmembers can significantly vary across multiple pixels, but this spectral variation is not physically described by the dispersion model. To solve this problem, we utilize K-means clustering to learn examplar endmembers for each category (e.g. grass, road, etc). Then the resulting centroid endmember can be fit to the dispersion model to allow further variation such as absorption band shifts in the spectra. We found that $K = 5$ worked the best for the Gulfport dataset.

TES Martian Dataset: The Thermal Emission Spectrometer (TES) [20] uses Fourier Transform Infrared Spectroscopy to measure the Martian surface. We utilize pre-processing from Bandfield et al. [2], and the endmember library used by Rogers et al. to analyze Mars TES data [27]. There is no ground truth for this dataset, as the true abundance of minerals on the Martian surface is unknown, so other metrics such as reconstruction error of the spectra are considered.

4.3.2   Endmember Fitting Results

To bootstrap both the analysis-by-synthesis and inverse rendering algorithms, good initial conditions for the dispersion parameters need to be input to the model.
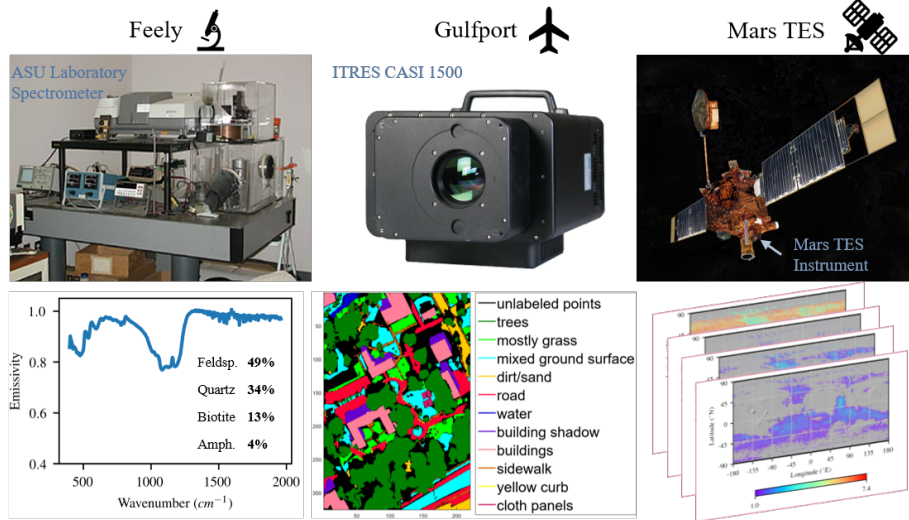
Figure 4.6: Datasets: This figure shows representative data and instrumentation for the three datasets considered in this chapter. Data includes laboratory, aircraft, and satellite measurements, and ground truth ranges from detailed abundance analysis under a microscope (Feely [28]) to pure pixel labels of land type for spectral classification (Gulfport [29]) to no ground truth for the Martian data (TES [20]).

Determining dispersion parameters typically required detailed molecular structure analysis or exhaustive parameter searching methods [67, 76, 48]. One main advantage of our method is that we utilize gradient descent to efficiently find parameter sets for different materials. In Appendix B, we share some of these parameter sets.

In Figure 4.7, we show qualitative results of our endmember fitting by minimizing the loss in Equation 4.6 using gradient descent. The reconstructed spectra achieves a low MSE with the measured spectra with an average MSE of 0.016 for the TES library, 0.0019 for the Feely library, and an MSE of 2.6e-5 on the Gulfport cluster centroids. Note that there is noise in the measurements, and so MSE is not an absolute metric of the fit to the true unknown spectra.
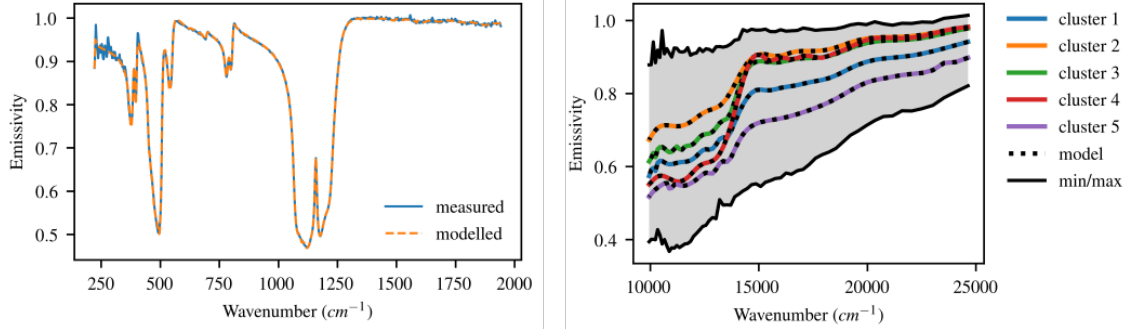
Figure 4.7: Endmember Fitting: (Left) Measured and modelled spectra for a quartz sample in the IR. (Right) Cluster centroids found for pixels labelled as grass in the Gulfport dataset, and the model fit to these centroids. Note the high fidelity of fit via the dispersion model for both these cases.

Table 4.1: Spectral Variation Results: Mean squared error of the abundance predictions vs. ground truth for Feely, Gulfport, and Gulfport synthetic datasets. The bold entries indicate top performance.

Dataset	FCLS [36]	NCM [68]	BCM [23]	CNN-1D [84]	CNN-3D [84]	CNN-1D Modified	Analysis-by-synthesis	Inverse Rendering
Feely [28]	0.121	0.119	0.131	0.469	N/A	0.205	**0.052**	0.188
Gulfport [29]	0.75	0.799	0.800	1.000	0.497	0.297	0.45	**0.272**
Gulfport Synthetic	0.911	0.471	0.136	0.824	N/A	0.148	0.147	**0.059**

4.3.3   Spectral Unmixing Results

In Table 4.1, we show results on the Feely, Gulfport, and the Gulfport synthetic mixture datasets. For Feely, the analysis-by-synthesis method achieved a MSE of 0.052, with the next closest method (NCM) achieving 0.119. Due to the Feely dataset only containing 90 test samples, the machine learning methods were trained on synthetic data which explains their lower performance as data mismatch. Thus, the low error of analysis-by-synthesis shows the utility of the dispersion model for modelling endmember variability, particularly in cases with low training data.

For the Gulfport dataset, the task was to predict the material present since the labeled data are for single coarse materials (e.g. road, grass, etc) at 100% abundance per pixel. Here, the deep learning methods of CNNs and Inverse Rendering have the highest performance. This is expected as there exists a large amount of training data to learn from. Note that Inverse Rendering performs the best at 0.272 MSE, demonstrating that the addition of a generative dispersion model to the output of the CNN improves performance over purely learned approaches. Also note that our analysis-by-synthesis method still has relatively high performance (0.45 MSE) without using any training data at all.

For the Gulfport synthetic mixture dataset, Inverse Rendering achieves the lowest MSE of 0.059, leveraging both physics-based modeling for spectral mixing as well as learns from available training data. The BCM and the analysis-by-synthesis methods both outperform the CNN methods, even though they do not have access to the training data. In fact, BCM even slightly outperforms the analysis-by-synthesis method, which could be because the sources of variation in this data are well-described by statistical distributions.

Speed of Methods: The additional capacity of adding statistical and physical

models usually has a cost of speed in implementation. Averaged over 90 mixtures, the convergence for a single operation was FCLS - 10ms, BCM - 1.23s, NCM - 18ms, CNN - 33ms, Inverse Rendering - 39ms, and analysis-by-synthesis - 10.2s. Future work could potentially increase the speed of analysis-by-synthesis with parallel processing.

4.3.4   Noise Analysis

Prior to spectral unmixing, emissivity is separated from radiance by dividing out the black-body radiation curve at the estimated temperature [60, 21]. In general, a Gaussian noise profile in the radiance space with variance $\sigma^2_{\text{radiance}}$ results in wavenumber dependent noise source in the emissivity space with the profile $\sigma^2(\omega) = \sigma^2_{\text{radiance}} \cdot 1/B(\omega, T)$ where B is the black-body function given by Planck's law. In our noise experiments we use a black-body radiation curve for a 330K target, which is the approximate temperature the Feely dataset samples were held at. In Figure 4.8 left, we see that the emissivity noise is higher where the radiance signal is lower.

We simulated varying the noise power to determine the methods' robustness tested on 30 samples from the Feely dataset. In Figure 4.8, you can see that analysis-by-synthesis still has the best performance in the presence of noise, and is relatively flat as noise increases compared to other methods. We note that statistical methods, while having higher average error, seem to be robust to increased noise as they can handle random perturbations of each spectral band statistically. CNN and Inverse Rendering methods perform the worst for high noise, as these methods were trained on data without noise.

4.3.5   Producing Mars Abundance Maps

The Mars TES data were unmixed using our analysis-by-synthesis method to demonstrate it's utility on tasks where zero training data are available. The method
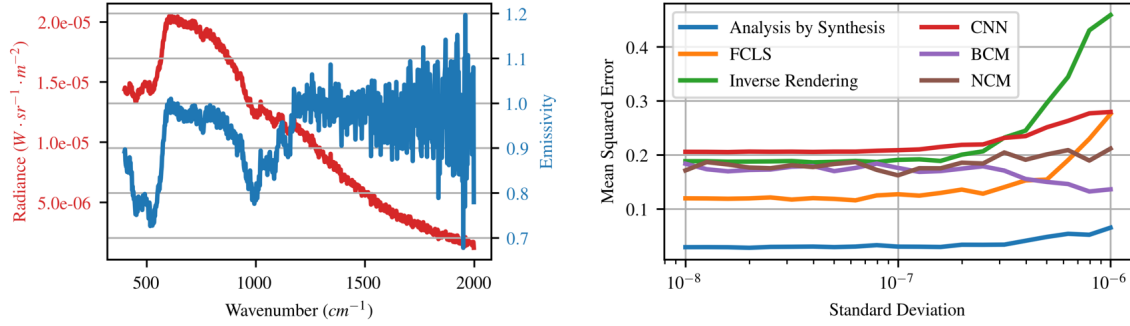
Figure 4.8: Noise Sweep: The left plot shows the radiance profile of a spectra perturbed by Gaussian noise and the resulting emissivity profile after separating out the blackbody radiance. The right figure shows the robustness of the algorithms to increasing amounts of noise.

produces mineral maps which correctly finds abundances of the mineral hematite at Meridiani Planum in Figure 5.1. This is an important Martian mineral which provides evidence for liquid water having existed at some point on Mars, and has been verified by NASA's Opportunity Rover [44]. Note how FCLS predicts many sites for hematite, while our method narrows down potential sites on the Martian surface, which is useful for planetary scientists. By allowing for spectral variation through our physics-based approach, our method has lower RMS reconstruction error than previous analysis of TES data. FCLS, which was previously used on TES because of the zero training-data problem, has an average RMS reconstruction error of 0.0043 while analysis-by-synthesis has an average of 0.0038. This is an exciting result as our methods could provide a new suite of hyperspectral analysis tools for scientists studying the Martian surface.
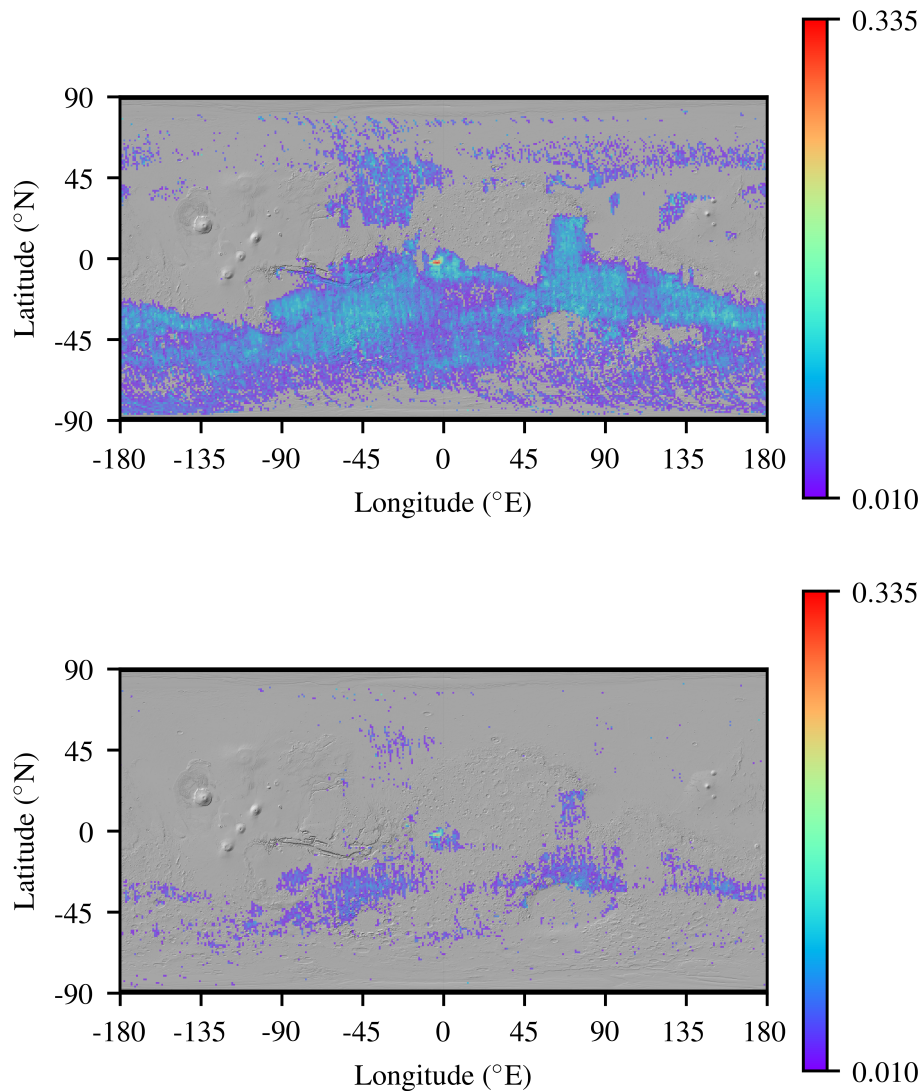
Figure 4.9: Martian Surface Hematite Map: The figure shows a mineral map of hematite on the Martian surface produced by FCLS (top) and analysis-by-synthesis (bottom) using TES data. Both algorithms find the deposit of hematite on Meridiani Planum, as expected, but analysis-by-synthesis predicts a sparser map which matches expectation of mineral distributions. Note that these are preliminary maps and should be inspected by geoscientists who can tune the sparse regularization and dispersion parameter constraints to better match the expected mineral distributions.

4.4 Discussion

This chapter incorporated generative physics models into spectral unmixing algorithms via differentiable programming. We adopt a physics-based dispersion model to simulate spectral variability, and show how this model can realistically fit several real measured spectra via gradient descent. We further show how to jointly optimize for the dispersion parameters and material abundances with an analysis-by-synthesis optimization. A second algorithm is introduced for tasks where additional data is available by training a CNN to "Inversely Render" a hyperspectral image with the differentiable dispersion model in the loop.

We validate these contributions extensively on three datasets ranging from mid to far IR and VNIR, and compared against state-of-the-art optimization, statistical and deep learning benchmarks. From these experiments we observe that analysis-by-synthesis has the best performance when training data is not available, and that Inverse Rendering has the best performance when training data is available. We also see that analysis-by-synthesis is noise resilient, and reconstructs Mars spectra with lower error than previous techniques.

There are still limitations for the methods proposed. First, analysis-by-synthesis has a large computational cost compared to other methods, although this could be mitigated through parallelization of the algorithm. Secondly, the spectral unmixing community is limited by the training data available. This is difficult to overcome, because it is expensive to produce quality datasets. Future work could investigate generating realistic synthetic data suitable for training machine learning based algorithms for better performance. We hope that this work using generative physics based models inspires others to produce work on realistic synthetic data as well as differentiable programming methods which require low training data.

Chapter 5

CONCLUSION

In this thesis we explored methods to encode physical realism into spectral unmixing. Sparse regularization and constraints are used to guide the algorithm to produce believable abundance maps. We compare several existing sparse regularization techniques and show that the L_p norm produces the lowest error. Then, we address the problem of spectral variability from a physics-based approach by allowing absorption bands to shift in strength, shape, and frequency. This is accomplished by using the Lorentz-Lorenz dispersion model to render spectra, with variation in absorption bands controlled by dispersion parameters. The model is fit to several real spectral measurements via an efficient gradient descent optimization. Then, an analysis-by-synthesis optimization approach is used to jointly optimize the dispersion parameters (allowing absorption bands to shift) and the pure material abundances. We introduce a second algorithm that trains a CNN to "inversely render" a hyperspectral image, which improves performance and speed when training data is available. All of the the algorithms, constraints, and regularization terms are implemented by differentiable terms that use the gradient to produce optimal results. This central theme of differentiable programming was successful on incorporating physical knowledge into optimization and machine learning based techniques for spectral unmixing.

We validated our methods on four datasets ranging from synthetic, laboratory, aircraft, and satellite-based spectroscopy. The datasets also covered two important regions of spectroscopy for thermal infrared emissions and visual-to-near-infrared reflected solar spectroscopy. We compared against several optimization, statistical and deep learning benchmarks and achieve state-of-the-art performance by reducing error

and producing sparser abundance maps of Mars with lower reconstruction error. Our experiments were limited to datasets where the linear mixing model was assumed. In the future, work could be done to tie these sparse regularization and spectral variation techniques into non-linear unmixing algorithms.

Differentiable programming is a promising area for many fields, including remote sensing and hyperspectral imaging. Since the physical concepts that model the journey of a photon as it is emitted, reflected, scattered, absorbed, and ultimately observed are well understood, it makes sense to build this knowledge into our algorithms. Still, the amount of information we observe could become dwarfed by the amount of parameters in a full physical simulation, which is a recipe for an under-determined system and non-unique solutions. Of course, if the physical world really does have all of those parameters then the exact interactions of the photons may be too complex to identify uniquely. The options would be to observe more information by increasing spectral resolution/combining observations together or to make prior assumptions to simplify the amount of unknowns. The method of combining several observations could be as simple as including spatial information from adjacent pixels in the analysis, which would be an interesting area for future work. This is a promising avenue to explore as spatial information from adjacent pixels has greatly improved other computer vision tasks such as detecting objects with spatial-convolutional filters.

The method of encoding prior knowledge by making assumptions is tricky, and sometimes we do not realize the assumptions we are making. For example, when we use a simplified physical model, such as linear unmixing (without spectral variability/regularization), it may seem advantageous that the system becomes over-determined and there are unique solutions promised. However, linear unmixing is actually making very strong assumptions that there are no non-linear effects and that the endmember library exactly matches the spectra of the pure materials in the

target sample (no spectral variation). In reality, there is spectral variation, and linear unmixing is just making hard coded assumptions on these parameters which can lead to errors. Another method of encoding prior knowledge is through constraints and regularization. Sparse regularization has made amazing contributions to other fields such as compressed sensing by showing that unique solutions can exist for some underdetermined systems if sparsity is encoded into the algorithm. Although we do not claim that our sparse prior guarantees uniqueness, the experimental evidence showed that regularization and differentiable physical models gave complementary performance benefits.

We hope that this work inspires geoscientists and other remote sensing analysts to use our tools to improve the accuracy of their results. There could even be additional science done by analyzing the physically plausible spectral variation predicted for the pure materials in target mixtures. The spectral variation of the material gives clues about it's exact composition, and the history of the geological processes that formed it. This could be particularly interesting in the study of planets other than our own, such as Mars, where the spectral variation might not match anything we have seen on Earth and could be a telling sign of an important geological process.

We also hope that the central idea of blending physical models with optimization and machine learning algorithms inspires future innovations in hyperspectral unmixing, and computer vision in general. With recent advance in automatic differentiation, it is now convenient to make complex physical models differentiable. There is also an underlying problem in machine learning for remote sensing because of the lack of datasets labelled by experts rather than algorithms. This is hard for the remote sensing community because experts themselves do not know how to label data without further analysis from their algorithms. Unless the algorithm was flawless, training supervised algorithms with this type of data would be unreliable. So, there is clearly
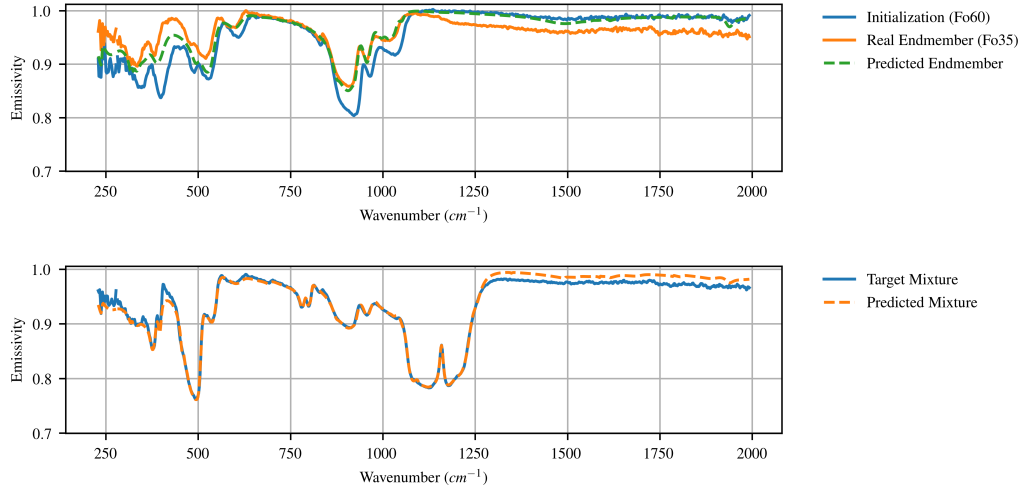
Figure 5.1: Solving for Spectral Variation: The analysis-by-synthesis algorithm is initialized with dispersion parameters that fit Quartz and Olivine Fo60. Then a synthetic target mixture is generated with 60% Olivine Fo35 (which does not match the Fo60 endmember) and 40% Quartz (Which does match the endmember). The analysis-by-synthesis algorithm finds the abundances and updates the dispersion parameters to find the best fit to the target mixture (bottom). Interestingly, after performing the optimization the predicted endmember spectra (top) better fits the real endmember, Olivine Fo35. This is promising that analysis-by-synthesis with the dispersion model could be used to predict the real endmember variation present in an unknown spectra.

a need for physics-based solutions to produce algorithms that provide good results without the need for training data. Also, physics-based simulations could possibly produce synthetic data which is reliable enough to train machine learning algorithms. It is an exciting period of time when physics-driven and data-driven methods are bridging together and we are eager to see where this research area goes.

REFERENCES

[1] Azinovic, D., T.-M. Li, A. Kaplanyan and M. Niessner, "Inverse path tracing for joint material and lighting estimation", in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", pp. 2447–2456 (2019).

[2] Bandfield, J. L., "Global mineral distributions on mars", Journal of Geophysical Research: Planets 107, E6, 9–1 (2002).

[3] Bandfield, J. L., D. Rogers, M. D. Smith and P. R. Christensen, "Atmospheric correction and surface spectral unit mapping using thermal emission imaging system data", Journal of Geophysical Research: Planets 109, E10 (2004).

[4] Bandfield, J. L., D. Rogers, M. D. Smith and P. R. Christensen, "Atmospheric correction and surface spectral unit mapping using thermal emission imaging system data", Journal of Geophysical Research: Planets 109, E10 (2004).

[5] Bandfield, J. L. and M. D. Smith, "Multiple emission angle surface–atmosphere separations of thermal emission spectrometer data", Icarus 161, 1, 47–65 (2003).

[6] Bateson, C. A., G. P. Asner and C. A. Wessman, "Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis", IEEE Transactions on Geoscience and Remote Sensing 38, 2, 1083–1094 (2000).

[7] Baydin, A. G., B. A. Pearlmutter, A. A. Radul and J. M. Siskind, "Automatic differentiation in machine learning: a survey", The Journal of Machine Learning Research 18, 1, 5595–5637 (2017).

[8] Bhojanapalli, S., A. Kyrillidis and S. Sanghavi, "Dropping convexity for faster semi-definite optimization", in "Conference on Learning Theory", pp. 530–582 (2016).

[9] Boltyanski, V., H. Martini and V. Soltan, *Geometric methods and optimization problems*, vol. 4 (Springer Science & Business Media, 2013).

[10] Borsoi, R. A., T. Imbiriba, J. C. M. Bermudez, C. Richard, J. Chanussot, L. Drumetz, J.-Y. Tourneret, A. Zare and C. Jutten, "Spectral variability in hyperspectral data unmixing: A comprehensive review", arXiv preprint arXiv:2001.07307 (2020).

[11] Bottou, L., "Large-scale machine learning with stochastic gradient descent", in "Proceedings of COMPSTAT'2010", pp. 177–186 (Springer, 2010).

[12] Burns, R. G., "Crystal field spectra and evidence of cation ordering in olivine minerals", American Mineralogist: Journal of Earth and Planetary Materials 55, 9-10, 1608–1632 (1970).

[13] Chang, C.-I., *Hyperspectral imaging: techniques for spectral detection and classification*, vol. 1 (Springer Science & Business Media, 2003).

[14] Chen, F. and Y. Zhang, "Sparse hyperspectral unmixing based on constrained lp-l 2 optimization", IEEE Geoscience and Remote Sensing Letters 10, 5, 1142–1146 (2013).

[15] Chen, Y., Y. Chi, J. Fan and C. Ma, "Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval", Mathematical Programming 176, 1-2, 5–37 (2019).

[16] Chen, Y., H. Jiang, C. Li, X. Jia and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks", IEEE Transactions on Geoscience and Remote Sensing 54, 10, 6232–6251 (2016).

[17] Cheng, G., Z. Li, J. Han, X. Yao and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification", IEEE Transactions on Geoscience and Remote Sensing 56, 11, 6712–6722 (2018).

[18] Chouzenoux, E., M. Legendre, S. Moussaoui and J. Idier, "Fast constrained least squares spectral unmixing using primal-dual interior-point optimization", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 1, 59–69 (2014).

[19] Christensen, P. R., J. L. Bandfield, V. E. Hamilton, D. A. Howard, M. D. Lane, J. L. Piatek, S. W. Ruff and W. L. Stefanov, "A thermal emission spectral library of rock-forming minerals", Journal of Geophysical Research: Planets 105, E4, 9735–9739 (2000).

[20] Christensen, P. R., J. L. Bandfield, V. E. Hamilton, S. W. Ruff, H. H. Kieffer, T. N. Titus, M. C. Malin, R. V. Morris, M. D. Lane, R. Clark et al., "Mars global surveyor thermal emission spectrometer experiment: investigation description and surface science results", Journal of Geophysical Research: Planets 106, E10, 23823–23871 (2001).

[21] Christensen, P. R., V. E. Hamilton, G. Mehall, D. Pelham, W. O'Donnell, S. Anwar, H. Bowles, S. Chase, J. Fahlgren, Z. Farkas et al., "The osiris-rex thermal emission spectrometer (otes) instrument", Space Science Reviews 214, 5, 87 (2018).

[22] Combe, J.-P., S. Le Mouélic, C. Sotin, A. Gendrin, J. Mustard, L. Le Deit, P. Launeau, J.-P. Bibring, B. Gondet, Y. Langevin and P. Pinet, "Analysis of omega/mars express data hyperspectral data using a multiple-endmember linear spectral unmixing model (melsum): Methodology and first results", Planetary and Space Science 56, 7, 951–975 (2008).

[23] Du, X., A. Zare, P. Gader and D. Dranishnikov, "Spatial and spectral unmixing using the beta compositional model", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 6, 1994–2003 (2014).

[24] Duchi, J. C. and F. Ruan, "Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval", Information and Inference: A Journal of the IMA 8, 3, 471–529 (2019).

[25] Ehlmann, B. L. and C. S. Edwards, "Mineralogy of the martian surface", Annual Review of Earth and Planetary Sciences 42, 291–315 (2014).

[26] Engel, J., L. H. Hantrakul, C. Gu and A. Roberts, "Ddsp: Differentiable digital signal processing", in "International Conference on Learning Representations", (2020).

[27] et al., R., "Mineralogical composition of sands in meridiani planum determined from mars exploration rover data and comparison to orbital measurements", Journal of Geophysical Research: Planets 113, E6 (2008).

[28] Feely, K. C. and P. R. Christensen, "Quantitative compositional analysis using thermal emission spectroscopy: Application to igneous and metamorphic rocks", Journal of Geophysical Research: Planets 104, E10, 24195–24210 (1999).

[29] Gader, P., A. Zare, R. Close, J. Aitken and G. Tuell, "Muufl gulfport hyperspectral and lidar airborne data set", Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570 (2013).

[30] Garbuny, M., "Optical physics", Optical Physics by Max Garbuny New York, NY: Academic Press, 1965 (1965).

[31] Ge, R., C. Jin and Y. Zheng, "No spurious local minima in nonconvex low rank problems: A unified geometric analysis", in "Proceedings of the 34th International Conference on Machine Learning-Volume 70", pp. 1233–1242 (JMLR. org, 2017).

[32] Gkioulekas, I., A. Levin and T. Zickler, "An evaluation of computational imaging techniques for heterogeneous inverse scattering", in "European Conference on Computer Vision", pp. 685–701 (Springer, 2016).

[33] Gkioulekas, I., S. Zhao, K. Bala, T. Zickler and A. Levin, "Inverse volume rendering with material dictionaries", ACM Transactions on Graphics (TOG) 32, 6, 162 (2013).

[34] Goudge, T. A., J. F. Mustard, J. W. Head, M. R. Salvatore and S. M. Wiseman, "Integrating crism and tes hyperspectral data to characterize a halloysite-bearing deposit in kashira crater, mars", Icarus 250, 165–187 (2015).

[35] Hapke, B., *Theory of reflectance and emittance spectroscopy* (Cambridge University Press, 2012).

[36] Heinz, D. C. *et al.*, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery", IEEE Transactions on Geoscience and Remote Sensing 39, 3, 529–545 (2001).

[37] Heylen, R., M. Parente and P. Gader, "A review of nonlinear hyperspectral unmixing methods", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 6, 1844–1868 (2014).

[38] Hu, W., Y. Huang, L. Wei, F. Zhang and H. Li, "Deep convolutional neural networks for hyperspectral image classification", Journal of Sensors 2015 (2015).

[39] Indyk, P. and M. Ruzic, "Near-optimal sparse recovery in the l1 norm", in "2008 49th Annual IEEE Symposium on Foundations of Computer Science", pp. 199–207 (IEEE, 2008).

[40] Iordache, M.-D., J. M. Bioucas-Dias and A. Plaza, "Hyperspectral unmixing with sparse group lasso", in "Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International", pp. 3586–3589 (IEEE, 2011).

[41] Jain, P., P. Kar *et al.*, "Non-convex optimization for machine learning", Foundations and Trends® in Machine Learning 10, 3-4, 142–336 (2017).

[42] Jones, E., T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python", URL http://www.scipy.org/ (2001–).

[43] Keshava, N. and J. F. Mustard, "Spectral unmixing", IEEE Signal Processing Magazine 19, 1, 44–57 (2002).

[44] Klingelhöfer, G., R. V. Morris, B. Bernhardt, C. Schröder, D. S. Rodionov, P. De Souza, A. Yen, R. Gellert, E. Evlanov, B. Zubkov *et al.*, "Jarosite and hematite at meridiani planum from opportunity's mössbauer spectrometer", Science 306, 5702, 1740–1745 (2004).

[45] Larkin, P., *Infrared and Raman spectroscopy: principles and spectral interpretation* (Elsevier, 2017).

[46] LeCun, Y., Y. Bengio and G. Hinton, "Deep learning", Nature 521, 7553, 436 (2015).

[47] Lee, J. D., M. Simchowitz, M. I. Jordan and B. Recht, "Gradient descent converges to minimizers", arXiv preprint arXiv:1602.04915 (2016).

[48] Lee, S. and C. Tien, "Optical constants of soot in hydrocarbon flames", in "Symposium (international) on combustion", vol. 18, pp. 1159–1166 (Elsevier, 1981).

[49] Li, S., W. Song, L. Fang, Y. Chen, P. Ghamisi and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview", IEEE Transactions on Geoscience and Remote Sensing 57, 9, 6690–6709 (2019).

[50] Li, T.-M., M. Aittala, F. Durand and J. Lehtinen, "Differentiable monte carlo ray tracing through edge sampling", ACM Trans. Graph. (Proc. SIGGRAPH Asia) 37, 6, 222:1–222:11 (2018).

[51] Li, Y., H. Zhang and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network", Remote Sensing 9, 1, 67 (2017).

[52] Liu, P., H. Zhang and K. B. Eom, "Active deep learning for classification of hyperspectral images", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10, 2, 712–724 (2016).

[53] Loubet, G., N. Holzschuch and W. Jakob, "Reparameterizing discontinuous integrands for differentiable rendering", Transactions on Graphics (Proceedings of SIGGRAPH Asia) 38, 6 (2019).

[54] Lu, G. and B. Fei, "Medical hyperspectral imaging: a review", Journal of Biomedical Optics 19, 1, 010901 (2014).

[55] Moersch, J. and P. R. Christensen, "Thermal emission from particulate surfaces: A comparison of scattering models with measured spectra", Journal of Geophysical Research: Planets 100, E4, 7465–7477 (1995).

[56] Nimier-David, M., D. Vicini, T. Zeltner and W. Jakob, "Mitsuba 2: A retargetable forward and inverse renderer", Transactions on Graphics (Proceedings of SIGGRAPH Asia) 38, 6 (2019).

[57] Paszke, A., S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga and A. Lerer, "Automatic differentiation in pytorch", (2017).

[58] Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library", in "Advances in Neural Information Processing Systems", pp. 8024–8035 (2019).

[59] Pilanci, M., L. E. Ghaoui and V. Chandrasekaran, "Recovery of sparse probability measures via convex programming", in "Advances in Neural Information Processing Systems", pp. 2420–2428 (2012).

[60] Ramsey, M. S. and P. R. Christensen, "Mineral abundance determination: Quantitative deconvolution of thermal emission spectra", Journal of Geophysical Research: Solid Earth 103, B1, 577–596 (1998).

[61] Ramsey M., C. P., "Mineral abundance determination: Quantitative deconvolution of thermal emission spectra: Application to analysis of martian atmospheric particulates", Journal of Geophysical Research: Solid Earth (2000).

[62] Ravi, N., J. Reizenstein, D. Novotny, T. Gordon, W.-Y. Lo, J. Johnson and G. Gkioxari, "Pytorch3d", https://github.com/facebookresearch/pytorch3d (2020).

[63] Roberts, D. A., M. Gardner, R. Church, S. Ustin, G. Scheer and R. Green, "Mapping chaparral in the santa monica mountains using multiple endmember spectral mixture models", Remote Sensing of Environment 65, 3, 267–279 (1998).

[64] Ruder, S., "An overview of gradient descent optimization algorithms", arXiv preprint arXiv:1609.04747 (2016).

[65] Salisbury, J. W., D. M. D'Aria and F. F. Sabins Jr, "Thermal infrared remote sensing of crude oil slicks", Remote Sensing of Environment 45, 2, 225–231 (1993).

[66] Shkuratov, Y., L. Starukhina, H. Hoffmann and G. Arnold, "A model of spectral albedo of particulate surfaces: Implications for optical properties of the moon", Icarus 137, 2, 235–246 (1999).

[67] Spitzer, W. and D. Kleinman, "Infrared lattice bands of quartz", Physical Review 121, 5, 1324 (1961).

[68] Stein, D., "Application of the normal compositional model to the analysis of hyperspectral imagery", in "IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003", pp. 44–51 (IEEE, 2003).

[69] Sun, J., Q. Qu and J. Wright, "A geometric analysis of phase retrieval", Foundations of Computational Mathematics 18, 5, 1131–1198, URL http://dx.doi.org/10.1007/s10208-017-9365-9 (2017).

[70] Sunshine, J. M. and C. M. Pieters, "Determining the composition of olivine from reflectance spectroscopy", Journal of Geophysical Research: Planets 103, E6, 13675–13688 (1998).

[71] Thaker, P., G. Dasarathy and A. Nedić, "On the sample complexity and optimization landscape for quadratic feasibility problems", (2020).

[72] Tibshirani, R., "Regression shrinkage and selection via the lasso", Journal of the Royal Statistical Society: Series B (Methodological) 58, 1, 267–288 (1996).

[73] Tsai, C.-Y., A. C. Sankaranarayanan and I. Gkioulekas, "Beyond volumetric albedo — A surface optimization framework for non-line-of-sight imaging", in "IEEE Intl. Conf. Computer Vision and Pattern Recognition (CVPR)", (2019).

[74] Wang, F., J. Decker, X. Wu, G. Essertel and T. Rompf, "Backpropagation with callbacks: Foundations for efficient and expressive differentiable programming", in "Advances in Neural Information Processing Systems", pp. 10180–10191 (2018).

[75] Wang, F., D. Zheng, J. Decker, X. Wu, G. M. Essertel and T. Rompf, "Demystifying differentiable programming: Shift/reset the penultimate backpropagator", Proceedings of the ACM on Programming Languages 3, ICFP, 1–31 (2019).

[76] Wenrich, M. L. and P. R. Christensen, "Optical constants of minerals derived from emission spectroscopy: Application to quartz", Journal of Geophysical Research: Solid Earth 101, B7, 15921–15931 (1996).

[77] Yang, X., Y. Ye, X. Li, R. Y. Lau, X. Zhang and X. Huang, "Hyperspectral image classification with deep learning models", IEEE Transactions on Geoscience and Remote Sensing 56, 9, 5408–5423 (2018).

[78] Yu, Y. and W. A. Smith, "Inverserendernet: Learning single image inverse rendering", in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", pp. 3155–3164 (2019).

[79] Zare, A. and P. Gader, "Sparsity promoting iterated constrained endmember detection in hyperspectral imagery", IEEE Geoscience and Remote Sensing Letters 4, 3, 446–450 (2007).

[80] Zare, A. and P. Gader, "Hyperspectral band selection and endmember detection using sparsity promoting priors", IEEE Geoscience and Remote Sensing Letters 5, 2, 256–260 (2008).

[81] Zare, A. and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing", IEEE Signal Processing Magazine 31, 1, 95–104 (2013).

[82] Zhang, C., L. Wu, C. Zheng, I. Gkioulekas, R. Ramamoorthi and S. Zhao, "A differential theory of radiative transfer", ACM Transactions on Graphics (TOG) 38, 6, 1–16 (2019).

[83] Zhang, L., L. Zhang and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art", IEEE Geoscience and Remote Sensing Magazine 4, 2, 22–40 (2016).

[84] Zhang, S., J. Li, H.-C. Li, C. Deng and A. Plaza, "Spectral–spatial weighted sparse regression for hyperspectral image unmixing", IEEE Transactions on Geoscience and Remote Sensing 56, 6, 3265–3276 (2018).

[85] Zhao, W. and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach", IEEE Transactions on Geoscience and Remote Sensing 54, 8, 4544–4554 (2016).

[86] Zhou, Y., A. Rangarajan and P. D. Gader, "A gaussian mixture model representation of endmember variability in hyperspectral unmixing", IEEE Transactions on Image Processing 27, 5, 2242–2256 (2018).

APPENDIX A

DERIVATION OF DISPERSION MODEL

In this appendix, we derive the dispersion model from first principles, modeling the generation of spectral radiance as a sum of mass-spring oscillations driven by an electromagnetic wave. This induces changes in the index of refraction, which governs the reflectance of the material with respect to light wavelength/frequency. This is based on earlier work by Garbuny and by Spitzer et al. [30, 67].

We first start with the equation for a mass-spring oscillator driven by an external force:

$$F = m\frac{d^2x}{dt^2} + R\frac{dx}{dt} + G \cdot x(t). \tag{A.1}$$

For a charged particle, $F = qE$, where q is charge and $E = E_0 e^{i\omega t}$ for a propagating electromagnetic wave. Thus we can substitute these in to get:

$$F = qE_0 e^{i\omega t} = m\frac{d^2x}{dt^2} + R\frac{dx}{dt} + G \cdot x(t) \tag{A.2}$$

which has the solution:

$$x = \frac{qE_0 e^{i\omega t}}{m}\frac{1}{\frac{G}{m} - \omega^2 + i\frac{R}{m}\omega} = \frac{qE_0 e^{i\omega t}}{m}\frac{1}{\omega_0^2 - \omega^2 + i\gamma\omega}. \tag{A.3}$$

where $\omega_0^2 = G/m$ and $\gamma = R/m$.

At the same time, we can also relate x to E via the band strength:

$$x = \frac{\alpha E}{q} \tag{A.4}$$

where α is the polarizability. Using the identity $\epsilon = 1 + 4\pi\alpha$, we can derive the following band strength equation:

$$x = \frac{(\epsilon - 1)E_0 e^{i\omega t}}{4\pi q}. \tag{A.5}$$

Combining Eq. A.3 and Eq. A.5, we get

$$\frac{(\epsilon - 1)E_0 e^{i\omega t}}{4\pi q} = \frac{qE_0 e^{i\omega t}}{m}\frac{1}{\omega_0^2 - \omega^2 + i\gamma\omega}. \tag{A.6}$$

Solving for ϵ:

$$\epsilon = \frac{4\pi q^2}{m}\frac{1}{\omega_0^2 - \omega^2 + i\gamma\omega} + 1. \tag{A.7}$$

Relating ϵ to the refractive index, $\epsilon\mu = \hat{n}^2$ where $\mu = 1$ and $\hat{n} = n - ik$, we get

$$\epsilon = (n - ik)^2 = n^2 - k^2 - 2nki = \frac{4\pi q^2}{m}\frac{1}{\omega_0^2 - \omega^2 + i\gamma\omega} + 1 \tag{A.8}$$

This yields the refractive index equations:

$$n^2 - k^2 = \frac{4\pi q^2}{m}\frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2} + 1 \tag{A.9}$$

$$2nk = \frac{4\pi q^2}{m} \frac{\gamma\omega}{(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2}. \tag{A.10}$$

Using the Lorentz-Lorenz formula, we can get

$$\hat{n}^2 = 1 + \frac{4\pi q^2}{m} \frac{1}{\omega_1^2 - \omega^2 + i\gamma\omega} \tag{A.11}$$

where $\omega_1^2 = \omega_0^2 - \frac{4\pi q^2}{3m}$ where $\omega_1 < \omega_0$. So plugging in ω_1 for ω_0 yields:

$$n^2 - k^2 = \frac{4\pi q^2}{m} \frac{\omega_1^2 - \omega^2}{(\omega_1^2 - \omega^2)^2 + \gamma^2\omega^2} + 1 \tag{A.12}$$

$$2nk = \frac{4\pi q^2}{m} \frac{\gamma\omega}{(\omega_1^2 - \omega^2)^2 + \gamma^2\omega^2}. \tag{A.13}$$

This entire derivation was for a single oscillator, but in practice, there are multiple oscillators that interact. We write this as a linear superposition given as follows:

$$n^2 - k^2 = \epsilon_r + \sum_i \frac{4\pi q^2 f_i}{m_i} \frac{\omega_i^2 - \omega^2}{(\omega_i^2 - \omega^2)^2 + \gamma_i^2\omega^2} \tag{A.14}$$

$$2nk = \sum_i \frac{4\pi q^2 f_i}{m_i} \frac{\gamma_i\omega}{(\omega_i^2 - \omega^2)^2 + \gamma_i^2\omega^2}. \tag{A.15}$$

where f_i is the strength of each individual oscillator. Using these equations, we have two equations for two unknowns (n and k), which we showed in Section 4.1.1 of Chapter 4 is the basis of calculating reflectance and emission.

APPENDIX B

DISPERSION PARAMETERS

The importance of initializing the alternating optimization with good initial dispersion parameters was emphasized in Chapter 4, as the problem is non-convex and good initialization is essential. It also makes intuitive sense to initialize with parameters fit to an endmember sample to give physical significance to the generated spectra. As shown in the results of Chapter 4, we achieve good fits with low MSE on endmember libraries used to analyze the Martian surface as well as a semi-urban university scene. The endmember libraries used to fit the minerals to analyze the Mars TES data are of high quality from the Arizona State University Thermal Emission Spectral Library [19]. The resulting parameters from a few of the important materials from this endmember library are provided in the following tables. The remaining dispersion parameters are available in the project repository folder named dispersionParameters.

Table B.1: Olivine Parameters: Dispersion parameters found for Olivine Fo10

Axis	Index	ω_0	γ	ρ	ϵ_r
0	0	258.45	0.018	0.022	1.07
0	1	272.71	0.038	0.070	1.07
0	2	285.33	0.027	0.035	1.07
0	3	340.81	0.021	0.015	1.07
0	4	361.06	0.067	0.187	1.07
0	5	467.03	0.060	0.091	1.07
0	6	589.36	0.032	0.043	1.07
0	7	826.60	0.011	0.015	1.07
0	8	863.05	0.030	0.083	1.07
0	9	934.94	0.018	0.038	1.07
0	10	1068.56	0.009	0.001	1.07
0	11	1349.50	0.043	0.009	1.07
0	12	1400.46	0.057	0.026	1.07
0	13	1452.82	0.064	0.020	1.07
0	14	1518.96	0.079	0.025	1.07
0	15	1597.62	0.018	0.001	1.07
0	16	1694.56	0.043	0.007	1.07
0	17	1794.69	0.032	0.002	1.07
0	18	1837.96	0.009	0.001	1.07
0	19	1934.50	0.056	0.020	1.07
1	0	293.77	0.042	0.240	1.99
1	1	303.28	0.058	0.263	1.99
1	2	317.16	0.137	0.356	1.99
1	3	473.47	0.006	0.002	1.99
1	4	496.39	0.029	0.030	1.99
1	5	504.45	0.062	0.302	1.99
1	6	562.92	0.055	0.057	1.99
1	7	577.32	0.027	0.008	1.99
1	8	891.85	0.023	0.189	1.99
1	9	990.28	0.047	0.086	1.99
1	10	1108.25	0.023	0.006	1.99

Table B.2: Biotite Parameters: Dispersion parameters found for Biotite

Axis	Index	ω_0	γ	ρ	ϵ_r
0	0	235.91	0.066	0.2343	1.31
0	1	432.39	0.056	0.4040	1.31
0	2	439.80	0.039	0.4131	1.31
0	3	446.34	0.014	0.0385	1.31
0	4	451.92	0.042	0.4797	1.31
0	5	594.57	0.073	0.0147	1.31
0	6	954.50	0.036	0.2510	1.31
0	7	1008.94	0.014	0.0578	1.31
0	8	1013.39	0.017	0.0184	1.31
0	9	1041.20	0.048	0.0178	1.31
0	10	1075.68	0.025	0.0198	1.31
0	11	1116.66	0.007	0.0003	1.31
0	12	1152.61	0.019	0.0012	1.31
0	13	1390.98	0.044	0.0177	1.31
0	14	1460.91	0.061	0.0280	1.31
0	15	1524.44	0.065	0.0676	1.31
0	16	1629.72	0.025	0.0271	1.31
0	17	1661.44	0.007	0.0034	1.31
0	18	1687.84	0.068	0.0723	1.31
0	19	1772.30	0.074	0.0877	1.31
0	20	1813.27	0.006	0.0009	1.31
0	21	1865.48	0.064	0.0731	1.31
0	22	1964.44	0.055	0.0131	1.31
1	0	268.77	0.073	0.4634	2.61
1	1	294.51	0.045	0.1965	2.61
1	2	313.92	0.064	0.3242	2.61
1	3	337.12	0.093	0.4930	2.61
1	4	362.24	0.062	0.1954	2.61
1	5	400.00	0.209	0.5174	2.61
1	6	462.66	0.065	0.4399	2.61
1	7	492.95	0.080	0.3498	2.61
1	8	510.47	0.061	0.0664	2.61
1	9	653.21	0.078	0.0611	2.61
1	10	718.49	0.040	0.0331	2.61
1	11	873.68	0.115	0.3343	2.61
1	12	928.32	0.048	0.0488	2.61
1	13	991.97	0.015	0.3550	2.61
1	14	1588.86	0.040	0.0607	2.61
1	15	1963.15	0.004	0.0023	2.61
1	16	1989.53	0.001	0.0002	2.61

Table B.3: Hematite Parameters: Dispersion parameters found for Hematite

Axis	Index	ω_0	γ	ρ	ϵ_r
0	0	258.29	0.11	0.110	1.27
0	1	279.35	0.13	0.141	1.27
0	2	294.73	0.11	0.149	1.27
0	3	335.86	0.08	0.130	1.27
0	4	471.32	0.07	0.098	1.27
0	5	526.58	0.05	0.029	1.27
0	6	543.94	0.07	0.062	1.27
0	7	563.14	0.08	0.067	1.27
0	8	609.37	0.04	0.041	1.27
0	9	619.61	0.04	0.041	1.27
0	10	632.43	0.07	0.067	1.27
0	11	654.46	0.09	0.054	1.27
0	12	686.74	0.12	0.038	1.27
0	13	798.98	0.04	0.011	1.27
0	14	890.21	0.03	0.009	1.27
0	15	916.82	0.02	0.005	1.27
0	16	958.26	0.04	0.014	1.27
0	17	1002.55	0.04	0.010	1.27
0	18	1100.72	0.03	0.022	1.27
0	19	1167.07	0.02	0.010	1.27
0	20	1238.37	0.01	0.005	1.27
0	21	1282.36	0.03	0.019	1.27
1	0	234.31	0.02	0.007	1.25
1	1	238.56	0.06	0.031	1.25
1	2	312.13	0.09	0.255	1.25
1	3	356.47	0.04	0.032	1.25
1	4	430.53	0.09	0.085	1.25
1	5	444.75	0.06	0.032	1.25
1	6	457.95	0.04	0.011	1.25
1	7	486.07	0.03	0.019	1.25
1	8	577.56	0.08	0.160	1.25
1	9	727.69	0.06	0.049	1.25
1	10	748.13	0.07	0.040	1.25
1	11	773.90	0.06	0.013	1.25
1	12	1049.92	0.10	0.058	1.25
1	13	1069.60	0.01	0.003	1.25
1	14	1140.36	0.02	0.012	1.25
1	15	1197.28	0.04	0.022	1.25
1	16	1256.54	0.02	0.010	1.25

APPENDIX C

ALTERNATING OPTIMIZATION CONVERGENCE

Our ultimate goal is to solve the spectral unmixing problem which can be formulated as $\min_{\mathbf{A},\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2$, where the minimization occurs over both the matrix $\mathbf{A}$ and the unmixing vector $\mathbf{x}$. This is a standard case of alternate minimization which is known to be nonconvex [41]. In practice, alternating minimzation are particularly hard to tackle due to the presence of suboptimal local minimas. Recent progress on tackling nonconvex problems involves either characterizing the optimization landscape [31, 71, 69] or providing initialization to descent algorithms [8, 15] to assure convergence to the global optimum. It is known that gradient descent applied to alternate minimization problem faces the issue of getting stuck at local minimas [41] and hence initialization plays an important role in solving Equation 7 in Chapter 4. With that in mind, we provide a mechanism to provide good initialization to gradient descent algorithm with the hope of tackling the alternating minimization problem effectively.

Initialization using dictionary $A(\epsilon^{\text{model}})$: We investigate the properties of matrix A as relates to the convergence of the alternating optimization. We denote the measured the emissivity spectrum of various materials in the lab as $\epsilon^{\text{measured}}$, and the physics-based dispersion model as ϵ^{model}. We then use these emissivity spectrum to construct a dictionary $A(\epsilon)$ which servers as the initialization for A in the alternating minimization approach in Equation 7 of Chapter 4. The intuition behind this revolves around the ability of matrix A as a dictionary of known emissivity spectra and we expect that the unknown spectra $\epsilon^{\text{unknown}}$ would be described as a linear combination of columns from matrix A.

Consider the following subproblem of the alternating minimization:

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}(\Lambda)\mathbf{x}\|_2^2$$

without the regularization terms. In order to ensure the uniqueness of the solution $\mathbf{x}^*$, we need to ensure that the matrix A is full rank. The rate of convergence for the above minimization is inversely dependent on the condition number of the matrix $\mathbf{A}(\Lambda)$. While it is difficult to analyze this matrix analytically, we perform an experimental characterization of the rank, condition number, and eigenvalues of the matrix for several different runs of the optimization algorithm with random initializations.

From the plots, we can note that the matrix $\mathbf{A}(\Lambda)$ has full rank with condition number of around 6000. The minimum and maximum eigenvalues are not showing drastic difference which goes well with our motive to incorporate small changes using alternate minimization to fit the spectral differences due to geographic differences. The high condition number is the reason for the relatively slow performance for running the alternating minimization framework, with our method taking tens of seconds to converge.
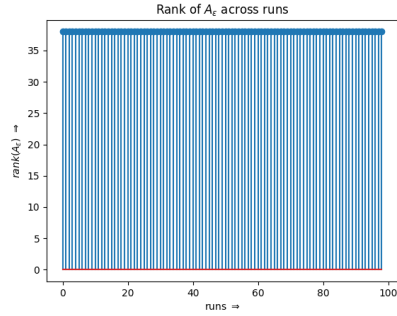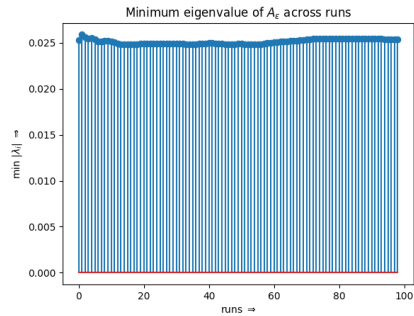
(a) Rank of **A** across runs
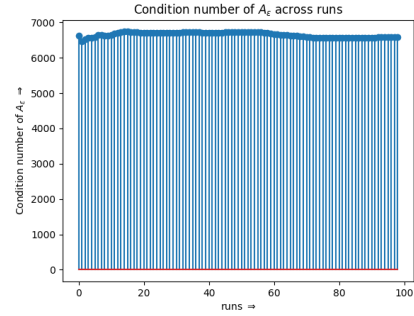
(b) Condition number of **A** across runs
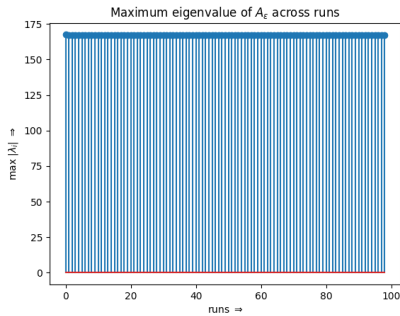
(c) Max eigenvalue of **A** across runs

(d) Min eigenvalue of **A** across runs

Figure C.1: Condition Number Behavior: The endmember library **A** behavior is plotted while the dispersion parameters are optimized during end-to-end spectral unmixing of a random observation.

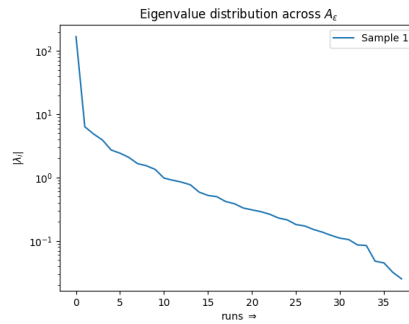Figure C.2: Eigenvalue Distribution: The endmember library **A** eigenvalue distribution is plotted while the dispersion parameters are optimized during end-to-end spectral unmixing of a random observation.

APPENDIX D

ADDITIONAL RESULTS

Some mineral maps produced using analysis-by-synthesis on the Mars TES data are shown in Figures D.1. - D.6 Additional mineral maps and numerical abundance files are located in the project repository. Code to reproduce all experiments is also uploaded to the project repository with instructions on how to use the code and apply the algorithm to new datasets. Additional results such as saved experiment performance metrics and dispersion parameters are also located in the project repository.
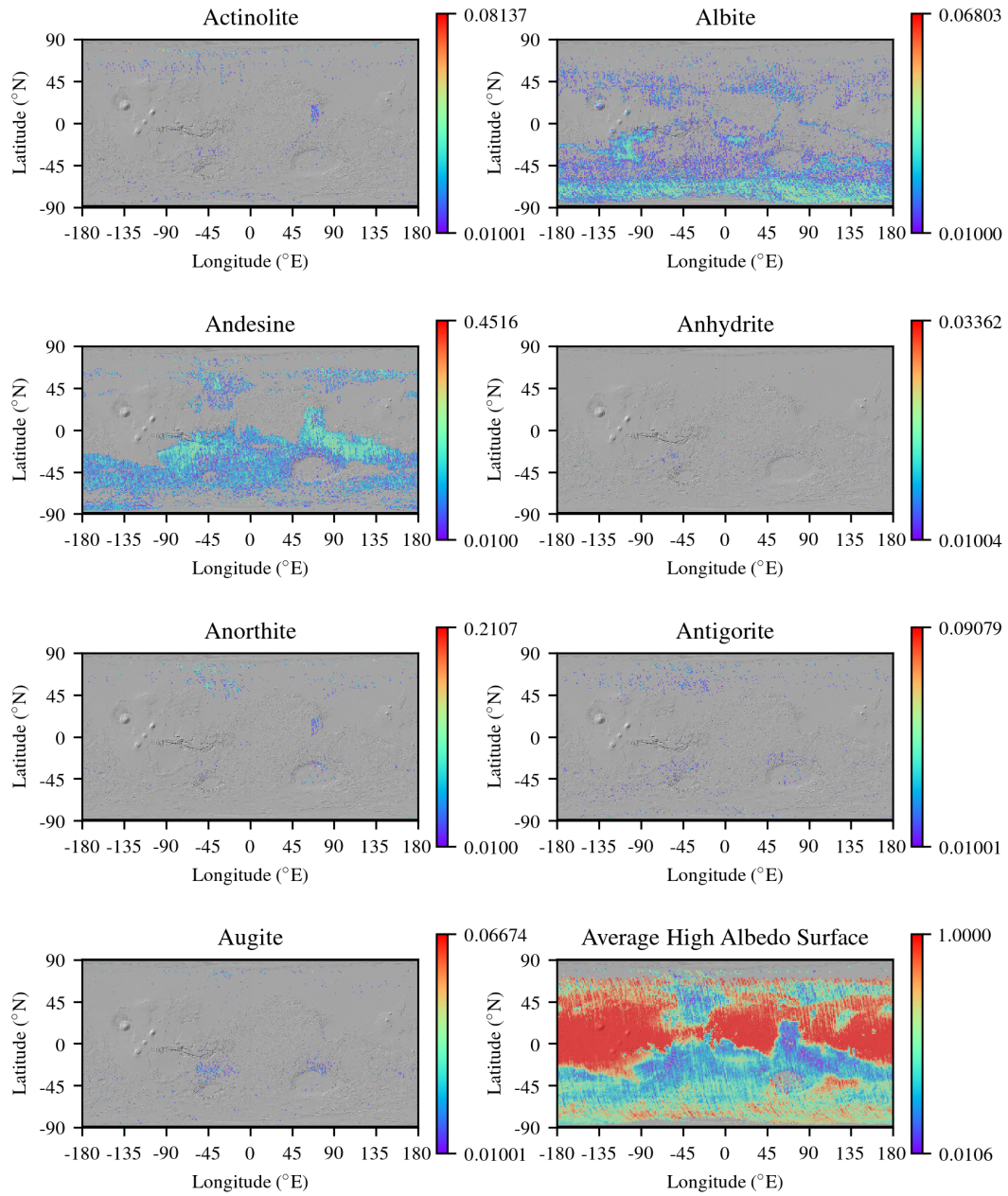
Figure D.1: Mineral Maps 1
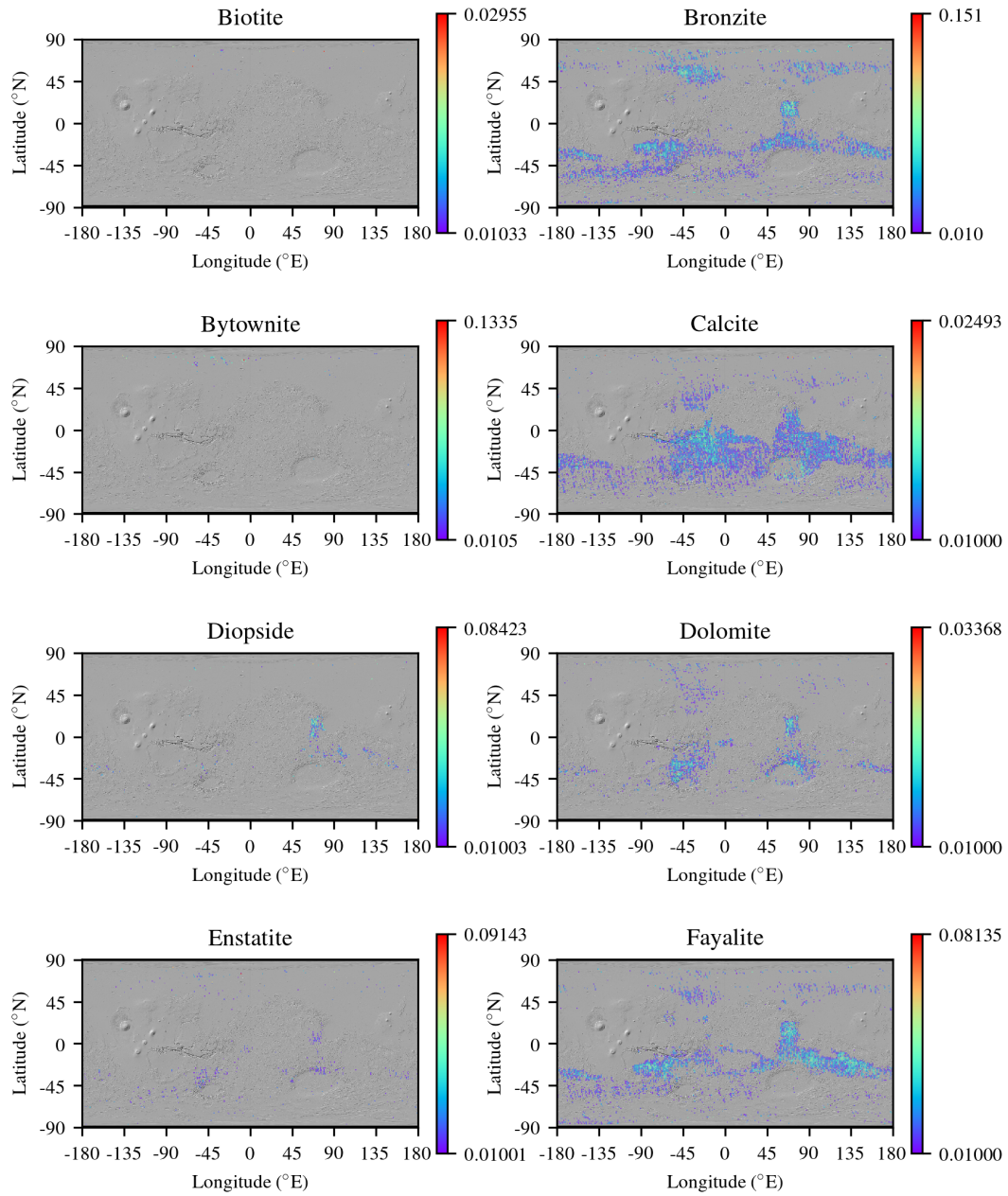
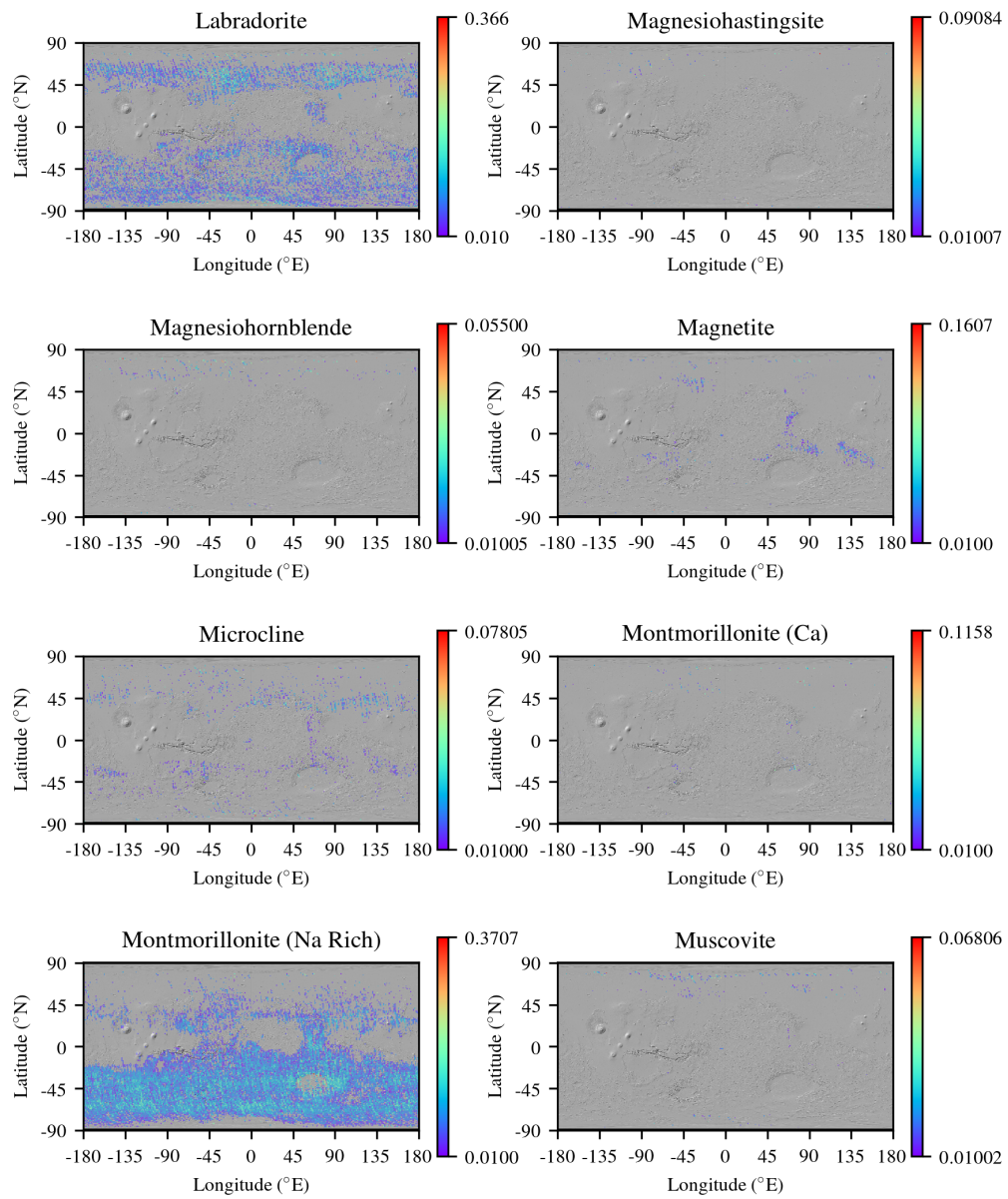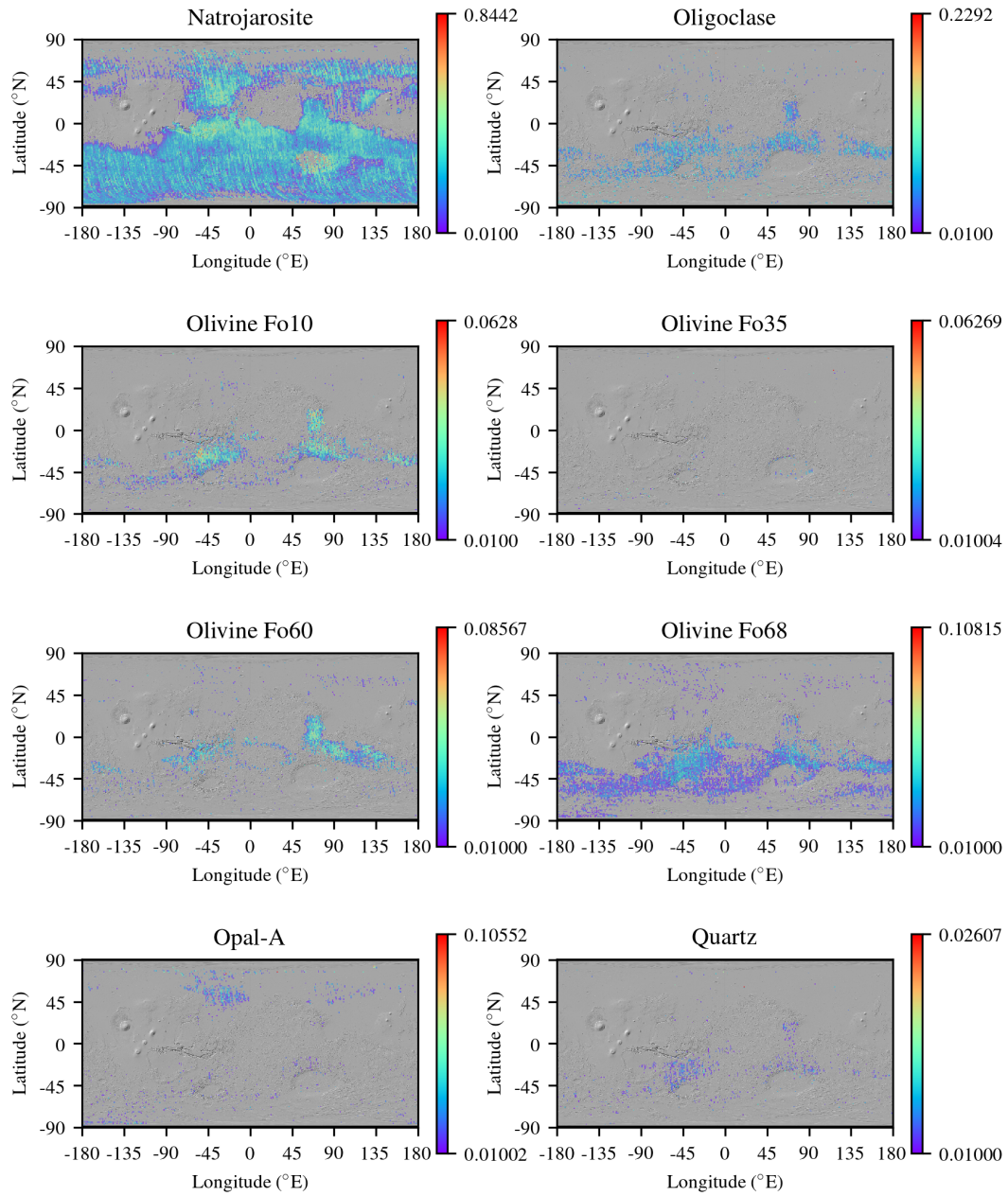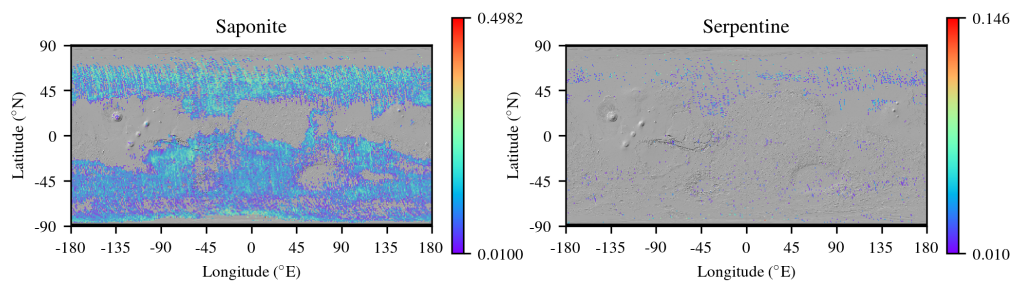Figure D.2: Mineral Maps 2
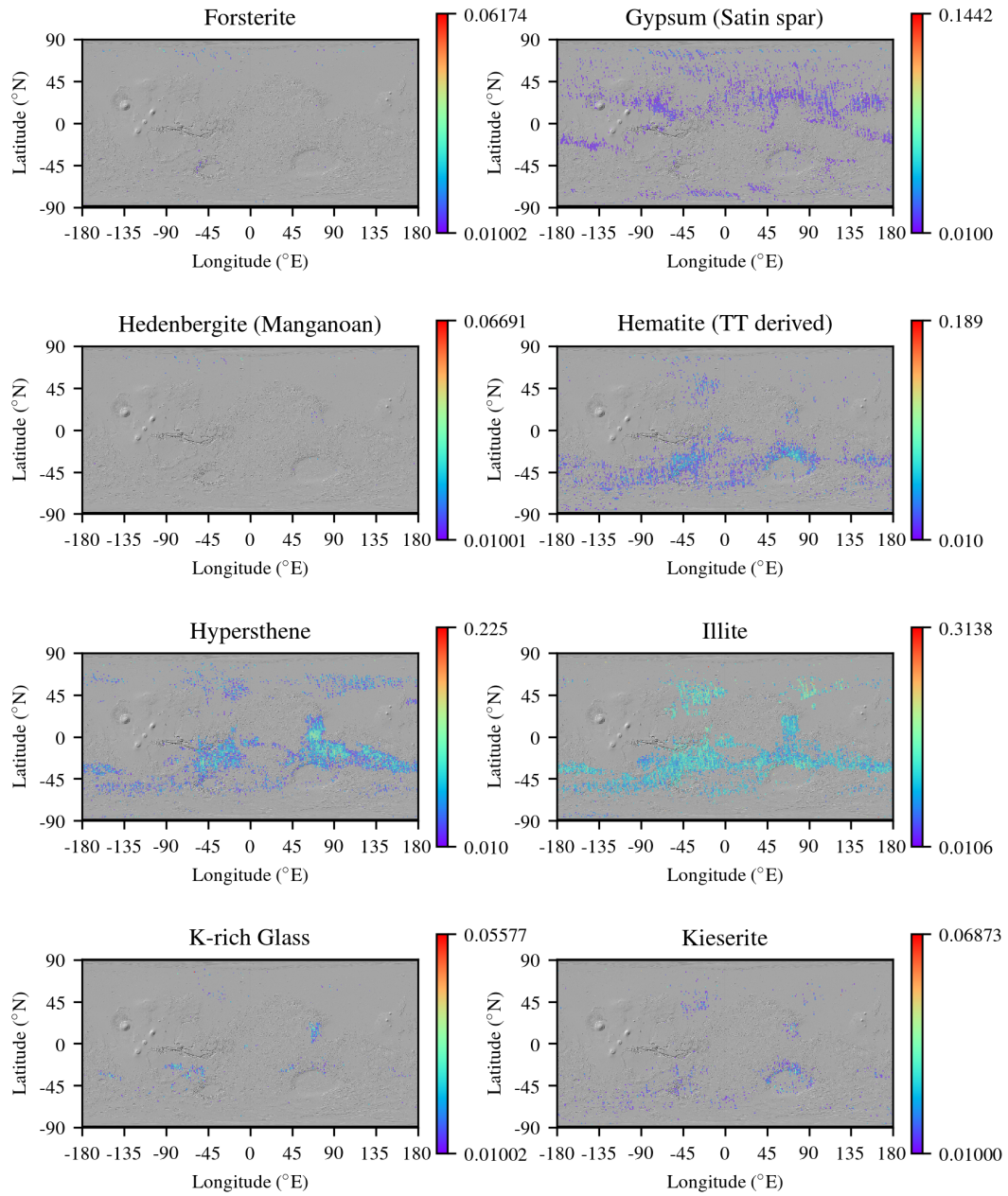
Figure D.3: Mineral Maps 3

Figure D.4: Mineral Maps 4

Figure D.5: Mineral Maps 5

Figure D.6: Mineral Maps 6