A BENCHMARK LINEAR UNMIXING DATASET WITH SPECTRAL VARIABILITY AND GROUND TRUTH

Xander Haijen, Bikram Koirala, Xuanwen Tao, and Paul Scheunders

Imec-Visionlab, University of Antwerp (CDE), Universiteitsplein 1, B-2610 Antwerp, Belgium

ABSTRACT

This paper introduces the Checkerboard dataset, a new benchmark for linear unmixing of hyperspectral images that incorporates spectral variability and provides reliable ground truth. This dataset addresses the critical need for realistic evaluation data, as existing real hyperspectral images often lack ground truth and synthetic data generation faces challenges in terms of realism. The Checkerboard dataset contains linear mixtures of four endmembers and includes spectral variability. We validate several aspects of the unmixing process on this dataset, specifically endmember extraction and abundance estimation. This work aims to provide a valuable resource for researchers to thoroughly evaluate and compare newly proposed unmixing methods, particularly those designed to handle spectral variability. The dataset is publicly available via Zenodo: `https://doi.org/10.5281/zenodo.15632214`.

***Index Terms*—** Hyperspectral unmixing, Spectral variability, Ground truth, Endmember extraction

1. INTRODUCTION

Since the hyperspectral signatures of pure materials present in hyperspectral images, known as endmembers (EMs), can vary significantly, researchers have developed a wide range of methods to incorporate this spectral variability into the EM extraction and abundance estimation process, which are the two most important steps when performing spectral unmixing. A popular method for modeling this variability involves scaling the EMs, as scaling factors effectively capture variability caused by factors such as changes in illumination and incident light angle, which may be induced by, for example, topographic variations [1].

Validating new unmixing methods is difficult due to the scarcity of realistic hyperspectral images with reliable ground truth. While synthetic data can be generated with controllable ground truth, achieving realism remains a challenge. Conversely, real hyperspectral images lack associated ground truth abundances. This limitation hinders researchers from thoroughly evaluating and comparing newly proposed unmixing methods.

In this paper, we present the **Checkerboard dataset**, which is a dataset containing several linear mixtures of four EMs. The dataset has a reliable ground truth and contains spectral variability, mainly scaling variability. In this way, we meet the need for a realistic linear unmixing dataset with a reliable ground truth. We validate the tasks of EM extraction and abundance estimation on this dataset.

2. BACKGROUND

2.1. Endmember extraction

EM extraction algorithms (EEAs) select or extract EMs from a hyperspectral image. These methods often assume that the number of EMs is known a priori. A first group of EEAs are pure-pixel-based algorithms. These rely on the *pure pixel assumption*, i.e., there exists at least one pure pixel for each EM in the scene [2]. A popular pure-pixel-based method is Vertex Component Analysis (VCA) [3]. VCA assumes the presence of pure pixels and iteratively identifies EMs by projecting the data onto a direction orthogonal to the subspace spanned by previously found EMs, selecting the extreme of each projection until the desired number of EMs is obtained.

When the pure pixel assumption is violated, pure-pixel-based methods become inaccurate. This has led to the development of volume-based methods. Because of the physical constraints of fractional abundances, namely, non-negativity and the sum-to-one constraints (see Eq. (2)), linearly mixed pixels are confined to a simplex, with EMs corresponding to its vertices (extreme points). Most volume-based methods look for the minimal volume simplex that contains all data, even when there are no pixels located at the vertices. For a guaranteed EM recovery, the abundances have to be *sufficiently scattered*: enough pixels should be located on the faces of the simplex to allow for simplex identification [2]. The methods differ mainly in how exactly they calculate the volume of the simplex. The most common approach is using a determinant formulation (see, e.g., Simplex Identification via Split Augmented Lagrangian (Sisal) [4] and Robust Minimum Volume Estimation (RMVE) [5]) or quadratic surrogates (e.g. those described in [6], including Iterated Con-

The research presented in this paper is funded by the Research Foundation Flanders (project G031921N). Bikram Koirala is a postdoctoral fellow of the Research Foundation Flanders, Belgium (FWO: 1250824N-7028).

strained EMs (ICE) [7]).

In the case of scaling variability, the data no longer lies in a simplex, but instead occupies a polyhedral cone [2, 8]. This can lead to inaccuracies in volume-based EEAs, as the origin, which is not an EM, becomes an extreme point of the cone. To mitigate this, we use a perspective projection on the image data. For an image pixel $\mathbf{x} \in \mathbb{R}^P$, where P denotes the number of spectral bands, the perspective projection $\mathbf{proj}(\mathbf{x})$ is given by

$$\mathbf{proj} : \mathbb{R}^P \to \mathbb{R}^P : \mathbf{x} \mapsto \mathbf{proj}(\mathbf{x}) = \frac{\mathbf{x}}{\mathbf{x}^\top \mathbf{v}} \tag{1}$$

for some fixed $\mathbf{v} \in \mathbb{R}^P$ such that $\mathbf{x}^\top \mathbf{v} \neq 0$ for all $\mathbf{x}$. It can be shown that this projects any polyhedral cone onto a simplex, whose extreme points are some scaled version of the EMs [8]. We will show in the experiments that applying this perspective projection prior to using a volume-based EEA results in accurate EM estimations. A drawback of the perspective projection is that pixels with low magnitudes can result in significant errors, because of possible near-zero values in the denominator $\mathbf{x}^\top \mathbf{v}$.

2.2. Linear unmixing under variability

Over the past years, several approaches have been proposed to mitigate the effects of hyperspectral variability [1]. They are mostly based on the Linear Mixing Model (LMM), which assumes that each measured spectrum $\mathbf{x}_n, n = 1, \ldots, N$ is a linear combination of the EMs $\mathbf{E} \in \mathbb{R}^{P \times K}$ and that the weights of this combination are the abundances $\mathbf{a}_n \in \mathbb{R}^K$, where N denotes the number of pixels and K denotes the number of EMs. The fractional abundances satisfy both the abundance non-negativity constraint (ANC) and the abundance sum-to-one constraint (ASC). The estimation of these abundances from linearly mixed data is accomplished by solving the Fully Constrained Least Squares Unmixing (FCLSU) problem:

$$\begin{aligned} &\min_{a_n} \frac{1}{2} \|\widehat{\mathbf{x}}_n - \mathbf{E}\mathbf{a}_n\|_2^2 \\ &\text{s.t. } \mathbf{a}_n \in \mathbb{R}_+^K \text{ (ANC)}, \quad \mathbf{1}_K^\top \mathbf{a}_n = 1 \text{ (ASC)} \end{aligned} \tag{2}$$

where $\widehat{\mathbf{x}}_n$ denotes a measured spectrum, $\mathbb{R}_+^K$ denotes the K-dimensional nonnegative reals and $\mathbf{1}_K$ is a K-dimensional vector of all ones.

Eq. (2) assumes that the same set of EMs is valid across the entire image, which is not the case when variability is present. To account for spectral variability, the term $\mathbf{E}$ must be replaced by a pixel-dependent term $\mathbf{E}_n = f(\mathbf{E}_0, \boldsymbol{\theta}_n)$, where $\mathbf{E}_0$ are reference EMs and $\boldsymbol{\theta}_n$ are variability parameters. The perturbations in EMs can be either additive (for modeling intrinsic variability) or multiplicative (for scaling variability). The scaled LMM (SLMM) incorporates a pixel-wise scaling factor, and assumes that all EMs are scaled similarly in a given pixel: $\mathbf{E}_n = \mathbf{E}_0 s_n$. This works very well to model uniform changes in illumination conditions. The extended LMM (ELMM) extends this to allow for EM-specific scaling effects and intrinsic variability, which is useful in the presence of topographic variation and in urban environments [9]: $\mathbf{E}_n = (\mathbf{E}_0 + \delta\mathbf{E}_n)\mathrm{diag}(\mathbf{s}_n)$, where diag is the diagonalization operator. In some cases, such as illumination with diffuse light, only certain wavelength regions are scaled. To model this, the generalized LMM (GLMM) incorporates band-wise scaling effects [10]: $\mathbf{E}_n = (\mathbf{E}_0 + \delta\mathbf{E}_n) \odot \mathbf{S}_n$, where the "$\odot$" operator denotes element-wise product. We recently proposed the two-step LMM (2LMM), considering one image-wide scaling step for the EMs to correct for scaling discrepancies, followed by a pixel-wise scaling term [11]: $\mathbf{E}_n = \mathbf{E}_0 \mathrm{diag}(\mathbf{s}_\mathbf{E}) s_n$. The model has a complexity between SLMM and ELMM and models realistic scenarios in which the reference EMs originate from a spectral library, or when they are obtained by an EEA, without having control over their initial scaling.

A second approach is using deep learning (DL) strategies. Several deep learning methods are designed to handle spectral variability, by incorporating physical models (such as the ELMM or GLMM) into their network architecture, or by including dedicated variability-learning components. Most available techniques are based on autoencoders (AEs), and recurrent or convolutional neural networks [12].

2.3. Hyperspectral data

The validation of unmixing methods is often done on two types of datasets. The first is real data obtained by airborne sensors or satellites, which does not have a reliable ground truth. The second is synthetically generated data with controllable ground truth. This type of data has the advantage of having known ground truth, but generating realistic data is a challenge. For the generation of abundances, common choices are Dirichlet distributions, Gaussian Random Fields (GRFs), or the abundance maps resulting from performing unmixing on a real image [1]. For generating spectral variability, a physical model has to be chosen, even though there is no agreement in the community on which physical model is best for modeling variability. The simplest way is to choose a model from Sec. 2.2 and generate the variability parameters from some statistical distribution such as the uniform or normal distribution, or from GRFs.

The absence of real data with ground truth limits the researchers in the validation of their method. This is exacerbated by the fact that a good reconstruction of the image (which is the only metric that can be measured for real data) does not imply a good abundance estimation (which is the main quantity of interest). Furthermore, good performance in synthetic data does not necessarily translate into good performance on real data. To the authors' knowledge, no dataset containing real data and abundance ground truth has been developed yet. The authors in [13] made a dataset in this spirit,

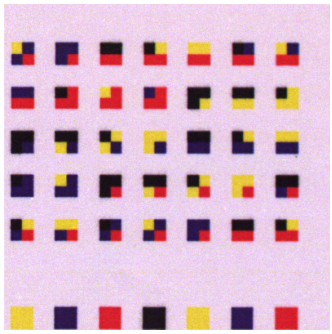

Fig. 1: Printed sample (RGB composite, angle 2)

but it lacks pixel-by-pixel ground truth. The authors in [14] also created a linear unmixing dataset with ground truth using a checkerboard, but unlike our dataset, it does not contain spectral variability, and the data (without EMs) is not sufficiently scattered, so it cannot be used for validating volume-based EEAs.

3. DATA DESIGN AND METHODOLOGY

For the dataset, we generated a sample containing 42 colored squares, each 2 × 2 mm² in size and spaced to avoid overlap. Among these, seven squares consisted of a single solid color, while the remaining 35 squares were subdivided into four smaller grids of 1 × 1 mm², each filled with one of four colors: red, blue, yellow, or black. We included all possible mixtures of the four materials with abundances 0, $1/4$, $1/2$ and $3/4$, where the order inside the square does not matter (31 combinations in total; four were repeated to fill the grid). The sample was printed on white paper, as shown in Fig. 1.

The sample was scanned using the *Imec Snapscan VNIR* camera[1] with two halogen lamps acting as the light source, placed on either side of the sample. This camera operates from 475 to 845 nm, capturing 150 spectral bands with a spectral resolution of approximately 3 nm. Visual inspection of the recorded spectra revealed that the final few bands were noisy and unreliable, so the last 20 bands are discarded in our experiments (see Fig. 2). The camera was positioned 30.5 cm above the sample and the lights were initially placed approximately 10 cm above the sample. To introduce variability, the lamps were raised 2 cm before each new acquisition, until a total of seven acquisitions was obtained. In each scenario, both the white calibration panel and the sample were scanned. We emphasize that the positions of both the camera and the sample remained fixed throughout the experiment. These scenarios were labeled from *angle 1* (lowest position) to *angle 7* (highest position). This variation alters both the illumination intensity and the angle of incidence; two factors that can be modeled well by scaling variability. This is illustrated by Fig. 2, which shows the seven acquired instances of the red EM. To convert the measured radiance to reflectance, the radiance of the white calibration panel from the second acquisition set-up was used. Each hyperspectral image contains 400 × 400 pixels, corresponding to approximately 28 × 28 pixels for each 2 × 2 mm² square. The experimental setup for the seventh acquisition scene is shown in Fig. 3.

[1] https://www.imechyperspectral.com/en/cameras/snapscan-vnir

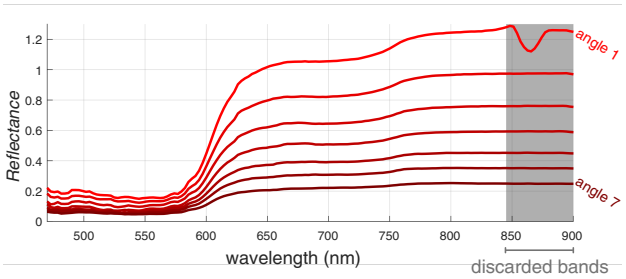

Fig. 2: The red EM for the seven acquisition scenarios

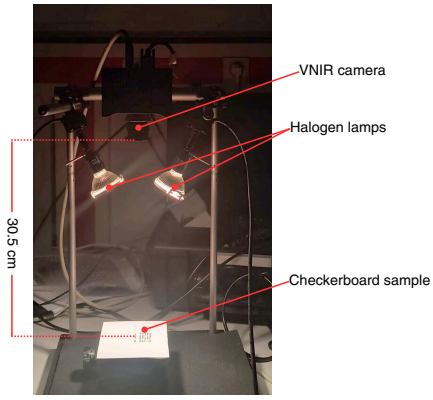

Fig. 3: Experimental setup (angle 7)

Finally, to generate the mixed and pure spectra, we manually selected 20 × 20 regions in each square, with the center of the square coinciding with the point where the sub-squares meet. This way, we ensure that the abundance ground truth is accurate for the selected region. We then took the mean of these 400 pixels to construct the spectrum. This resulted in a dataset containing 294 spectra (7 times 42), of which 245 are mixtures and 49 contain only one material.

4. VALIDATION

We evaluate two unmixing tasks using this dataset: (1) EM extraction and (2) spectral unmixing. Since the extremely low reflectance of black across all bands led to near-zero denominators and numerical issues in all methods, we exclude all spectra containing black. This resulted in a dataset that uniformly covers a 3-dimensional simplex, consisting of 140 spectra: 42 pure spectra and 98 mixed spectra.

EEA	case 1	case 2	case 3
VCA	**0.43**	0.94	5.78
Sisal	52.3	1.25	**2.48**
RMVE	65.7	**0.72**	2.56
ICE	36.7	0.97	2.68

Table 1: SAD (in degrees) between manually selected EMs and estimated EMs for a dataset with pure pixels (case 1), after using a perspective projection (case 2), and without pure pixels (case 3). The best results are highlighted in bold.

4.1. Endmember extraction

In this section, we will demonstrate the following three claims: (1) accurately extracting EMs requires addressing variability, (2) a perspective projection increases EM accuracy for volume-based methods, and (3) volume-based methods remain accurate in the absence of pure pixels. To show this, we run one pure-pixel based EEA (VCA) and three volume-based EEAs (Sisal, RMVE & ICE) on three variations of the Checkerboard dataset, without black spectra; *case 1*: unprocessed spectra, including pure spectra; *case 2*: data from case 1, pre-processed using a perspective projection; and *case 3*: data from case 2, without pure spectra. To evaluate the results, we utilize the mean spectral angle distance (SAD) between ground-truth EMs $\mathbf{E}$ (manually selected from the image) and extracted ones $\widehat{\mathbf{E}}$:

$$\mathrm{SAD}(\mathbf{E}, \widehat{\mathbf{E}}) = \frac{1}{K} \sum_{k=1}^{K} \arccos\left(\frac{\mathbf{e}_k \cdot \widehat{\mathbf{e}}_k}{\|\mathbf{e}_k\|_2 \|\widehat{\mathbf{e}}_k\|_2} \right) \times \frac{180^\circ}{\pi}. \quad (3)$$

Table 1 presents the mean SAD values achieved by the different EEAs on this dataset. VCA is the best performing algorithm in case 1, since VCA implicitly includes a perspective projection step and thus accounts for variability, which supports the first claim. Second, the decrease in SAD from the first case to the last two cases in all volume-based methods supports the second claim. Fig. 4 further supports this claim, showing estimated versus ground truth EMs for cases 1 & 2. Third, there is a noticeable performance drop for VCA when pure pixels are removed from the dataset (case 3), while volume based-methods still perform well. This supports the third claim.

4.2. Spectral unmixing

We conducted two types of experiments to evaluate the ability of state-of-the-art methods to accurately estimate fractional abundances: (a) using manually selected pure pixels and (b) using Sisal for EM extraction. We utilized the mean root mean squared error (RMSE) between ground-truth abundances $\mathbf{A}$ and estimated ones $\widehat{\mathbf{A}}$:

$$\mathrm{RMSE}(\mathbf{A}, \widehat{\mathbf{A}}) = \frac{1}{N} \sum_{n=1}^{N} \sqrt{\frac{1}{K} \|\mathbf{a}_n - \widehat{\mathbf{a}}_n\|_2^2} \quad (4)$$

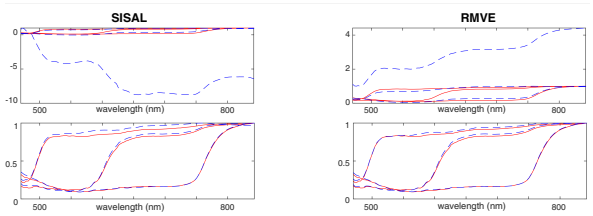

Fig. 4: Top row: real (solid red) and estimated (dashed blue) EMs without perspective projection for Sisal & RMVE. Bottom row: real and estimated EMs with perspective projection.

model	$\mathrm{RMSE}(\mathbf{A}, \widehat{\mathbf{A}})$	$\mathrm{SAD}(\mathbf{X}, \widehat{\mathbf{X}})$	Δt (s)
LMM	0.3243	7.1537	0.24
SLMM	0.2298	1.1530	0.001
ELMM	0.2365	**0.0028**	14.5
2LMM-IP	**0.0668**	1.1496	0.33
2LMM-LBFGS	0.1120	1.1522	0.16
GLMM	0.2799	0.0885	0.63
ReDSUNN	0.1867	9.2261	2.95

Table 2: Unmixing results with fixed EMs. The EMs were manually selected from the first, third and seventh acquisition scenes. The best results are highlighted in bold.

to evaluate the performance of the following seven methods: LMM, SLMM, ELMM, 2LMM-IP, 2LMM-LBFGS, GLMM and ReDSUNN. The method 2LMM-IP is the method proposed in [11], the method 2LMM-LBFGS is a method that uses a slightly adapted version of the preconditioned limited-memory BFGS (LBFGS) algorithm for alternating least-squares problems proposed in [15], and ReDSUNN [12] is a DL method with a recurrent neural network at its core.

For the first experiment, we manually select the EMs from the first, third, and seventh acquisition scene in order to introduce significant scale discrepancies. This situation directly mirrors the challenges posed by using EMs from spectral libraries. The 2LMM is specifically designed for such scenarios, and as expected, it outperforms all competing algorithms. Table 2 shows the abundance RMSE, reconstruction SAD, and computational time (Δt) and highlights the superior performance of 2LMM. While some methods, like ELMM, achieved good reconstruction errors, this did not translate to accurate abundance estimations. The poor result for SLMM can be linked to its underlying assumption of similarly scaled EMs within a pixel. The ELMM and GLMM rely heavily on the SLMM for a good initialization, so performed poorly as well. ReDSUNN also struggled due to the limited dataset (98 pixels), highlighting the need of deep neural networks for large amounts training data. Other deep learning methods similarly failed, yielding abundance RMSEs exceeding 40%.

In the second experiment, where EMs were estimated using Sisal with perspective projection, the 2LMM demonstrated its suitability for blind unmixing under variability, despite suffering from the inaccurate EM estimates provided

model	$\mathrm{RMSE}(\mathbf{A}, \widehat{\mathbf{A}})$	$\mathrm{SAD}(\mathbf{X}, \widehat{\mathbf{X}})$	Δt (s)
LMM	0.5457	18.583	0.21
SLMM	0.1289	1.0381	0.001
ELMM	0.3518	**0.5271**	3.18
2LMM-IP	0.1289	1.0387	0.28
2LMM-LBFGS	**0.1006**	1.0381	0.04
GLMM	0.4441	8.4976	13.5
ReDSUNN	0.2543	11.623	3.44

Table 3: Results for blind unmixing in the absence of pure pixels. The best results are highlighted in bold.

by Sisal due to the presence of intrinsic variability. Table 3 shows that, while the difference is not as large as in the first experiment, the 2LMM still outperformed the other methods, most of which exhibited very high error rates.

5. CONCLUSION

In this paper, we introduced the Checkerboard dataset, a novel hyperspectral unmixing benchmark that incorporates both spectral variability and reliable ground truth. To the best of our knowledge, it is the first dataset of its kind. We have demonstrated that volume-based methods, when combined with perspective projection as a preprocessing step, can accurately extract EMs from the dataset. Furthermore, we have shown the effectiveness of 2LMM in addressing spectral variability. We hope this dataset becomes a valuable resource for researchers seeking to validate newly proposed unmixing methods, particularly those aimed at handling spectral variability.

6. REFERENCES

[1] Ricardo Augusto Borsoi et al., "Spectral Variability in Hyperspectral Data Unmixing: A comprehensive review," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 223–270, Dec. 2021.

[2] Xiao Fu et al., "Nonnegative Matrix Factorization for Signal and Data Analytics: Identifiability, Algorithms, and Applications," *IEEE Signal Process. Mag.*, vol. 36, no. 2, pp. 59–80, Mar. 2019.

[3] José M.P. Nascimento and José M. Bioucas-Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[4] José M. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. 2009 Workshop Hypersp. Image and Signal Process.: Evol. Remote Sens.*, Aug. 2009, pp. 1–4.

[5] Xiao Fu et al., "Robust Volume Minimization-Based Matrix Factorization for Remote Sensing and Document Clustering," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6254–6268, Dec. 2016.

[6] Lina Zhuang et al., "Regularization Parameter Selection in Minimum Volume Hyperspectral Unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9858–9877, Dec. 2019.

[7] Mark Berman et al., "ICE: an automated statistical approach to identifying endmembers in hyperspectral images," in *Proc. 2003 IEEE Int. Geosci. and Remote Sens. Symp. (IGARSS)*, Toulouse, France, July 2003, vol. 1, pp. 279–283.

[8] Lucas Drumetz et al., "Spectral Variability Aware Blind Hyperspectral Image Unmixing Based on Convex Geometry," *IEEE Trans. Image Process.*, vol. 29, pp. 4568–4582, 2020.

[9] Lucas Drumetz et al., "Blind Hyperspectral Unmixing Using an Extended Linear Mixing Model to Address Spectral Variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, Aug. 2016.

[10] Tales Imbiriba, Ricardo Augusto Borsoi, and José Carlos Moreira Bermudez, "Generalized Linear Mixing Model Accounting for Endmember Variability," in *Proc. 2018 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1862–1866.

[11] Xander Haijen, Bikram Koirala, Xuanwen Tao, and Paul Scheunders, "A Two-step Linear Mixing Model for Unmixing under Hyperspectral Variability," Feb. 2025, arXiv preprint, arXiv:2502.17212.

[12] Ricardo A. Borsoi, Tales Imbiriba, and Pau Closas, "Dynamical Hyperspectral Unmixing With Variational Recurrent Neural Networks," *IEEE Trans. Image Process.*, vol. 32, pp. 2279–2294, 2023.

[13] Daniele Cerra et al., "DLR HySU—A Benchmark Dataset for Spectral Unmixing," *Remote Sens.*, vol. 13, no. 13, pp. 2559, Jan. 2021.

[14] Min Zhao, Jie Chen, and Zhe He, "A Laboratory-Created Dataset With Ground Truth for Hyperspectral Unmixing Evaluation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2170–2183, July 2019.

[15] Hans De Sterck and Alexander J. M. Howse, "Nonlinearly preconditioned L-BFGS as an acceleration mechanism for alternating least squares with application to tensor decomposition," *Numer. Linear Algebra Appl.*, vol. 25, no. 6 SI, 2018.