Probabilistic Mixture Modeling For End-Member Extraction in Hyperspectral Data

Oliver Hoidn
SLAC National Accelerator Laboratory
Menlo Park, CA-94025

Aashwin Ananda Mishra
SLAC National Accelerator Laboratory
Menlo Park, CA-94025

Apurva Mehta
SLAC National Accelerator Laboratory
Menlo Park, CA-94025

Abstract

Imaging spectrometers produce data with both spatial and spectroscopic resolution, a technique known as hyperspectral imaging (HSI). In a typical setting, the purpose of HSI is to disentangle a microscopic mixture of several material components in which each contributes a characteristic spectrum–often confounded by self-absorption effects, observation noise and other distortions. We outline a Bayesian mixture model enabling probabilistic inference of end member fractions while explicitly modeling observation noise and resulting inference uncertainties. We generate synthetic datasets and use Hamiltonian Monte Carlo to produce posterior samples that yield, for each set of observed spectra, an approximate distribution over end member coordinates. We find the model robust to the absence of pure (i.e. unmixed) observations as well as to the presence of non-isotropic Gaussian noise, both of which cause biases in the reconstructions produced by N-FINDER and other widespread end-member extraction algorithms.

1 Introduction

A wide range of scientific and applied fields–such as astronomy, geophysics, materials science, and agriculture–rely on the collection and interpretation of spatially-resolved spectral images as a workhorse. While the modality of the spectroscopic measurement varies widely across these fields, in all cases the observation consists of a mixture of a sparse set of "end-members". The scientific challenge is inverting the mixtures and identify the end-members. If the observations in the mixture set are homogeneously mixed then the task of inversion is impossible. On the other hand, if the end-members are immiscible, then the task of inversion is trivial. Here we address the intermediate case where observations are heterogeneously mixed across the observation set.

A simple way of representing a heterogeneously mixed observation is by a set of multi-dimensional arrays of weights, and the end-members identifiable by a set of distinct spectra. An example of a heterogeneously mixed observation set is a color photograph represented by 3 distinct colors: red, blue and green, and the monochrome images (arrays of weights) associated with the three colors. (Note that a sepia-toned monochrome photograph does not have three end-members, but just one, sepia.)

In many scientific applications, the end-member spectra are not as clearly distinct, and the array of weights can be more than 2D and contain non-spatial components, including time, scattering vectors, chemical compositions, red-shifts, etc. Thus, while the case of HSI is illustrative, we emphasize a more encompassing perspective where the observations need not be a spatial map, and can just

Machine Learning and the Physical Sciences workshop, NeurIPS 2022.

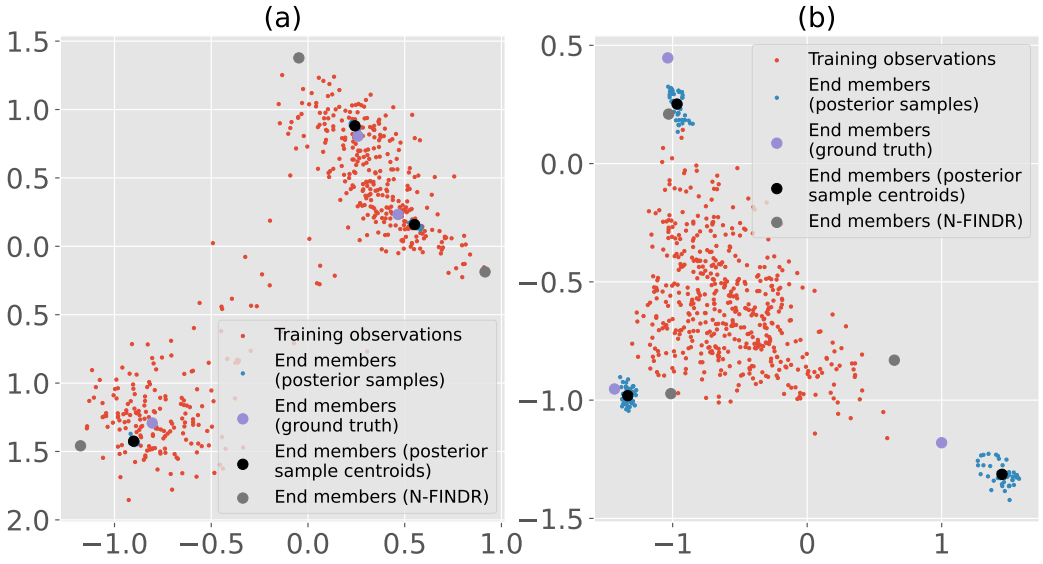

Figure 1: Comparison of end member coordinates extracted using N-FINDER vs. through sampling of the model posterior using HMC and the no-U turn (NUTS) sampler in Pyro Bingham et al. [2019]. The left (a) and right (b) panels show datasets generated under small and large values of the concentration hyperparamter α_1, respectively.

as well be treated as an unordered set of spectra in which other (e.g. spatial) parameters are not necessarily needed for the end member extraction.

Each observation in such a dataset is a spectrum made up of discrete samples along spectral dimension λ. In the typical interpretation, each spectrum is a mixture of individual spectral signatures corresponding to, e.g., electromagnetic emission from different chemical species. An end member is made up of a singular spectral signature and its associated species, and the goal of the analysis is to produce weights over the end members present in each observation along with an approximation of each end member signature.

In prior literature two of the most prevalent types of end member extraction algorithms are geometric/simplex methods (e.g. N-FINDER, the most widespread instance of these) and subspace reduction (e.g. nonnegative matrix factorization) Winter [1999], Miao and Qi [2007]. Rather than give a complete overview of existing end member extraction methods (Bioucas-Dias et al. Bioucas-Dias et al. [2012]), we focus here on the simplex approach and its properties which, when compared to a Bayesian model, draws out the differences between probabilistic and point-estimate approaches and the advantages of the former.

If linear mixing of spectral signatures is assumed, the number of sample points in the spectral dimension λ is D, and up to K end members are present, then fractional abundances correspond to points in the K-simplex and the K end members define a D-volume (in a $D-$ dimensional spectral space) that fully encloses all observed spectra. If we additionally assume no noise or observation artefacts and that for each end member there is at least one "pure" observation in the dataset for which the fractional abundance of that end member is 1, then it is automatic that the end members correspond to a choice of K observations from the dataset that define the maximum-possible N-volume. This maximum-volume condition is the basis for the N-FINDER algorithm and other simplex methods.

The above description is a useful geometric schema for viewing the mixing process, but exploiting it for end member extraction requires two additional assumptions beyond that of linear mixing: zero noise and presence of pure spectra in the data. While these two strong assumptions are not common to all non-Bayesian approaches, those methods that avoid, e.g., the zero noise assumption still restrict to isotropic noise and are opaque with respect to the consequences of noise on the model output (for example, noise adjusted principal components, NAPC Lee et al. [1990]).

We introduce a mixture model for probabilistic estimation of end member parameters. Using the model and simulated observations we approximate a full posterior distribution over end member spectra via Hamiltonian Monte Carlo and Variational Inference. By comparing end member reconstruction via inference under the model to N-FINDER, we find that the former is robust to noise as well as to incomplete observation of pure samples. Overall, it offers significantly more accurate end member extraction, as quantified by the reconstruction error of end member coordinates.

2 Data Generation and Models

The formulation is as a Bayesian hierarchical model with structure given by the plate diagram, Fig. 2. Sampling is performed once per mixture component k to determine end member parameters (for a total of K, the number of endmembers); per dth spectral dimension to scale the variance of observation noise in each dimension (for a total of $D = K-1$, the number of spectral dimensions); and per ith observed data point.

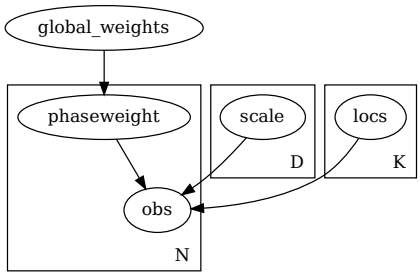

Figure 2: Plate diagram for the Bayesian end member mixture model

Notably, we have simplified the problem by setting the number of spectral dimensions to one less than the number of end members, K. In an experimental application, this would require prior dimensionality reduction of the data. Second and more significantly, we have assumed that K is known. The suitability of this assumption is likely domain-dependent. To take a particular example, in the context of phase identification in materials science the assumption may be reasonable to the extent that K is known (or bounded) by the Gibbs phase rule or similar considerations. The prior distributions for parameters $\vec{global\_weights}$, $\vec{locs}_k$ and $scale_d$ are

$$\begin{aligned} \vec{global\_weights} = \vec{w} &\sim Dir(\alpha_0 \vec{1}), \\ \vec{locs}_k = \vec{c_k} &\sim N(\vec{0}, \vec{I}), \\ scale_d = s_d &\sim U(\sigma/2, 3\sigma/2), \end{aligned} \tag{1}$$

where $Dir(\alpha_0 \vec{1})$ is a symmetric K-dimensional Dirichlet distribution over global end member frequencies; N is a D-dimensional Gaussian over end member coordinates; and U is a uniform distribution used to rescale samples from an LKJ distribution over observation noise correlations matrices. For a single sample, the underlying distribution over phase fractions is a Dirichlet distribution that depends on both the global weights and a second concentration hyperparameter, α_1, a measure of the degree of mixing:

$$phaseweight_i = \vec{w_i} \sim Dir(\alpha_1 \vec{w_k}) \tag{2}$$

Finally, the observation likelihood is a multivariate Gaussian with mean vector composed of the matrix product between the $D \times K$ end member coordinates and the local phase weight vector $\vec{w_i}$, and covariance given by $\Sigma(\vec{s})$ (itself dependent on both the LKJ prior and the scale parameters $\vec{s} = (s_0, ...s_d)$):

$$obs_i = y_i = N(C\vec{w_i}, \Sigma), \tag{3}$$

where C is the $D \times K$ matrix containing the end member coordinate vector $\vec{c_k}$ for all values of k. A single simulated dataset is produced by first drawing $\vec{c}$, $\vec{s}$ and $\vec{w}$ from the respective Bayesian priors and then generating N samples from the likelihood, eq. 3.

3 Results & Discussion

After generating 100 simulated datasets of $N = 500$ observations each in the above manner, we infer end-member coordinates in two ways: first, by using the model distribution function to sample from the posterior over model parameters with the no- U turn (NUTS) Hamiltonian Monte Carlo sampler;

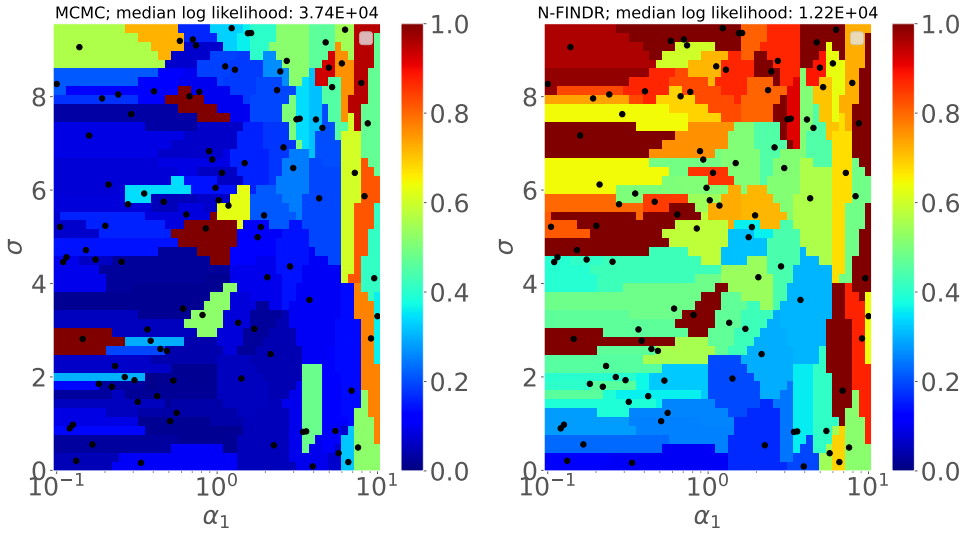

Figure 3: L2 reconstruction error for end member coordinates via HMC posterior samples (left) compared to N-FINDER (right), with varying values of the hyperparameters α_1 and σ.

second, with mean-field stochastic variational inference (SVI) using the ADAM optimizer and a learning rate of 0.005. To mitigate failures to converge (a common issue with mixture models) we run multiple rounds of inference with different random initializations. We select between 3 starts per dataset for HMC and 10 for VI; in each case, we retain only the inference round that produces the highest log likelihood for the dataset under the posterior samples (HMC) or approximation (VI). Per dataset, the sampling and inference ran for a total of 10 minutes on 12 CPU cores.

Figure 3 compares the reconstruction accuracy for 100 instances of this simulation-inference procedure, with 100 distinct combinations of the hyperparameters α_1 and σ. The reconstruction accuracy of HMC (a) is compared to that of a baseline algorithm, N-FINDER (b). We note, first, that high concentration parameter α_1 corresponds to incomplete sampling by the data of the end member simplex. The combination of high noise (σ) and high α_1 is the most difficult case for inference and consequently results in poor reconstruction accuracy for both approaches. Conversely, both methods do well when σ and α_1 are both small. Where σ or α_1 is large, however, HMC posterior sampling performs much better than N-FINDER. The improved reconstruction when pure samples are missing (high α_1) or the noise amplitude is high (large σ) conforms to our expectations of the benefits of this mixture model approach. Finally, we investigate posterior approximation with variational inference as a computationally-cheaper alternative to HMC. Figure 4 presents a three-way comparison of the reconstruction error for end-member coordinates under HMC, VI and N-FINDER. HMC and N-FINDER generate the best and worst reconstructions, respectively, and VI is intermediate.

4 Conclusions & Future Outlook

We have applied a probabilistic mixture model to the extraction of end members from spectral mixtures. Our approach formulates the retrieval of end member spectra and per-observation end member weights as a Bayesian inference problem using relatively uninformative priors for the phase-mixing and observation processes. Using simulated datasets we've sampled the model posterior with HMC and found that the resulting recovered end member parameters are more robust to observation noise and absence, in the data, of pure end member spectra compared to the popular N-FINDER algorithm. The Bayesian approach provides a principled framework for incorporating relevant prior information without introducing undue assumptions and opens the door to other intrinsically probabilistic analyses, such as uncertainty quantification.

The model that we have presented incorporates at least one possibly-onerous assumption, namely that the number of end members is known. This could be relaxed by replacing the fixed number of mixture

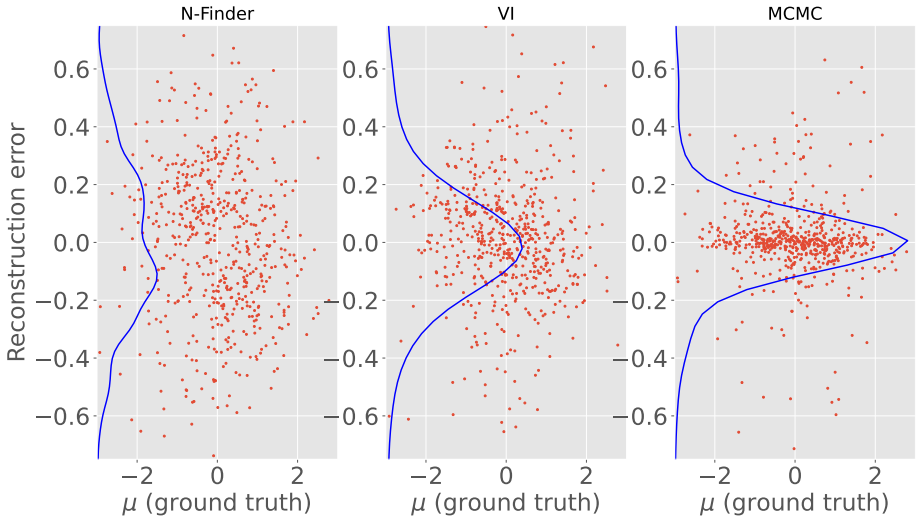

Figure 4: Residuals for reconstructed end member coordinate components under HMC sampling (right) and mean field variational inference posterior estimates (center) under the same model and data, compared to N-FINDER (left).

components and corresponding end member weights by a set of samples from a stick-breaking distribution, as in a Dirichlet process mixture model. Assumptions about K would thus be condensed into a single hyperparameter, the concentration parameter of the stick-breaking process.

In impending work, we intend to consider more flexible, non-linear mappings via the incorporation of Bayesian Neural Networks to infer embeddings. Additionally, we shall apply this analysis framework to real datasets.

Impact Statement

Spectral unmixing is the process of decomposing the spectral signature of a mixed observation into a set of end-members and their corresponding weights. It is used for providing information to monitor different natural resources in agricultural, forest, geological planning; and environmental problems such as erosion, deforestation, forest fires, etc. Spectral unmixing is often made challenging by noisy and incomplete data. To extract the most information possible from such sets of observations, we propose a probabilistic approach that offers a principled framework for incorporating relevant prior information without introducing undue assumptions. In addition this Bayesian reformulation of the problem is a necessary precursor to probabilistic analyses such as uncertainty quantification, which are unavailable under a large majority of existing approaches to spectral unmixing. While we are not the first to apply Bayesian inference to this task, our model encompasses the entire analysis and is less restrictive with respect to assumptions on the noisiness and information content of observations. In addition we've applied black box inference techniques instead of customized sampling. [Zare and Gader, 2011, Bchir et al., 2010]

5 NeurIPS checklist

(a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? *Yes*

(b) Have you read the ethics review guidelines and ensured that your paper conforms to them? *Yes*

(c) Did you discuss any potential negative societal impacts of your work? *No, but we are not aware of any*

(d) Did you describe the limitations of your work? *Yes*

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? *No, but these will be provided through an existing Github repository that we will make public*

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? *Yes*

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? *No, but some of the result figures encapsulate equivalent information*

(d) Did you include the amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? *Yes*

(a) If your work uses existing assets, did you cite the creators? *Yes*

(b) Did you mention the license of the assets? *No, but one may refer to the citation on Pyro* (c) Did you include any new assets either in the supplemental material or as a URL? *No; no such assets are needed, aside from the code (which we will release)*

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? *n/a*

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? *n/a*

References

Ouiem Bchir, Hichem Frigui, Alina Zare, and Paul Gader. Multiple model endmember detection based on spectral and spatial information. In *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pages 1–4. IEEE, 2010.

Eli Bingham, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D Goodman. Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research*, 20(1):973–978, 2019.

José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):354–379, 2012.

James B Lee, A Stephen Woodyatt, and Mark Berman. Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing*, 28(3):295–304, 1990.

Lidan Miao and Hairong Qi. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3):765–777, 2007.

Michael E Winter. N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Imaging Spectrometry V*, volume 3753, pages 266–275. SPIE, 1999.

Alina Zare and Paul Gader. Piece-wise convex spatial-spectral unmixing of hyperspectral imagery using possibilistic and fuzzy clustering. In *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, pages 741–746. IEEE, 2011.