Bayesian Non-Negative Matrix Factorization with Volume Prior

Morten Arngren[†∘], Mikkel N. Schmidt[‡] and Jan Larsen[†]

DTU Informatics[†] / University of Cambridge[‡] / FOSS Analytical A/S[∘]

ISP, Aug. 2009

Outline

- Introduction to hyperspectral image analysis.
  - The linear model and the constraint.
  - The convex geometry model.
  -
- The NMF model with volume regularizations.
  - Simulation results.
- Baysian NMF with volume priors.
  - Simulation results.

Hyperspectral Image Analysis

Introduction

- Classic image analysis is usually conducted on photographes having up to 3 RGB colors, sufficient for visualization.
- *Hyperspectral images* includes multiple color bands, typically > 50 bands and thus offers a more detailed analysis.

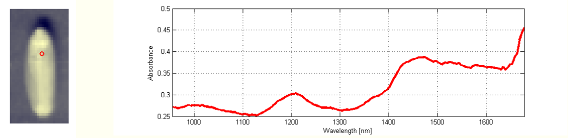


Figure 1 : Example of hyperspectral image. Each pixel consist of a 150 band spectra.

- The
- The observed spectra is a mix a set of pure components in the subject, dominated by a linear mixture.
- The objective is hence to decompose the image to into these pure spectral signatures.

Hyperspectral Image Analysis

Introduction

- Classic image analysis is usually conducted on photographes having up to 3 RGB colors, sufficient for visualization.
- *Hyperspectral images* includes multiple color bands, typically > 50 bands and thus offers a more detailed analysis.

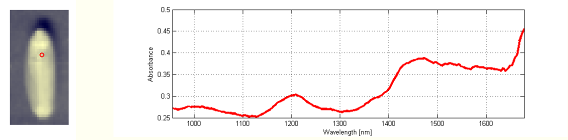


Figure 1 : Example of hyperspectral image. Each pixel consist of a 150 band spectra.

- The
- The observed spectra is a mix a set of pure components in the subject, dominated by a linear mixture.
- The objective is hence to decompose the image to into these pure spectral signatures.

Linear Modeling

The mixing of the constituents can be approximated linearly as

$$\mathbf{X} \approx \mathbf{WH} + \epsilon, \quad (1)$$

where **W** are the spectral signatures, **H** denote the fractional abundances (concentrations) and ϵ is the residual white Gaussian noise.

Constraints

- Non-Negativity, intensities can not be negative :

$$x_{i,j} > 0 \qquad \wedge \qquad w_{i,j} > 0 \qquad \wedge \qquad h_{i,j} > 0 \quad (2)$$

- Additivity, concentrations must sum to one : $\sum_i h_{i,j} = 1$

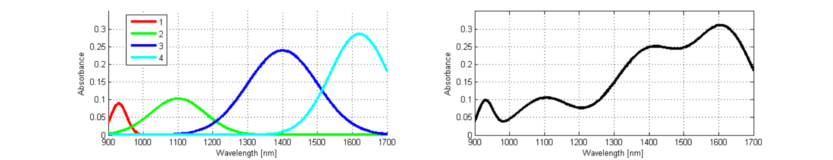


Figure 2 : Linear mixing of spectral signature into observed pixel spectra.

Convex Geometry Model

- These constraints lead to the following geometry for the 2 component case.

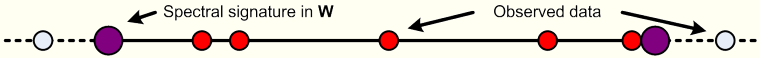


Figure 3 : Simple mixing with only 2 vertices.

- Purple vertices are the basis vertices denoted *endmembers*.
- Red points inside designate valid observed samples.
- Gray points outside are invalid due to noise (constraints violated).

For multiple endmembers the structure becomes an N-simplex.

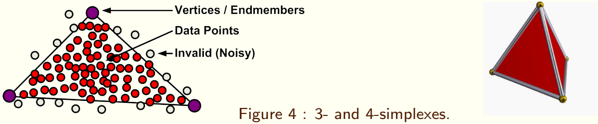


Figure 4 : 3- and 4-simplexes.

Objective is to locate the vertices as the basis spectral signatures based only on the observed data (unsupervised learning).

Data Acquisition

- Hyperspectral image acquired of wheat kernels, front and backside.
- Image is 320 pixels wide with a spectral range from 900-1700nm in 165 bands.
- Light preprocessing incl. spectral correction and kernel segmentation.
- Dataset becomes 8 small images of wheat kernels of size $44 \times 26 \times 152$.

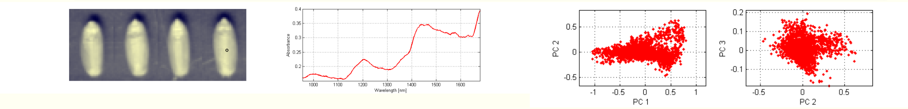


Figure 5 : Hyperspectral images of wheat kernels.

Synthetic Dataset

- True endmember spectra produced from hyperspectral image of pure constituents, protein, starch and oil from 950-1650nm (145 bands).
- Synthetic dataset includes 2000 samples, i.e. $\mathbf{X} \in \mathbb{R}^{145 \times 2000}$.

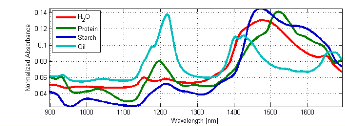


Figure 6 : Spectra of food constituents.

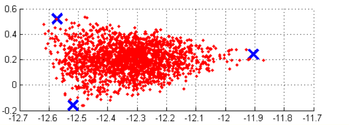


Figure 7 : Example of synthetic dataset.

Non-negative Matrix Factorization

Linear model $\mathbf{X} = \mathbf{WH} + \epsilon$ subject to non-negativity constraints (2).

- Regular NMF is not sufficient to capture vertices as likelihood term dominates. Hence additional constraint are required.

Cost function

$$E = \frac{1}{2}||\mathbf{X} - \mathbf{WH}||^2 + J_W(\mathbf{W}) \quad (3)$$

Regularizations

- Endmembers $\mathbf{W}$ should encourage tight solutions to data simplex.
- Additivity constraint is incorporated as by the *normalization invariance* approach.

Volume Regularization

- Sparsity regularization is not optimal as $\mathbf{W}$ are attracted toward origo.
- Volume-based regularization by minimizing a *simplex* volume for K endmembers

$$V_{simplex} = \frac{1}{K!}\,|\det(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})|^{1/2}, \quad (4)$$

where $\tilde{\mathbf{W}}$ are the vectors spanning the simplex.

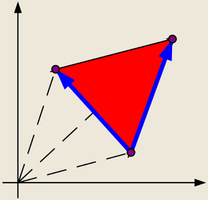


Figure 8 : The simplex volume.

Volume Regularization

- Subspace volume, [Miao (2007)]

$$J_{\mathbf{W}}(\mathbf{W}) = \det{}^2(\mathbf{C} + \mathbf{B}\mathbf{U_x}^T(\mathbf{X} - \mu)), \tag{5}$$

  where $\mathbf{U}_x$ holds the $K-1$ most significant PC's. Captures subspace parallelepiped and suppresses noisy directions, but depend on observations $\mathbf{X}$.

- Parallelepiped volume [Schachtner (2009)]

$$J_{\mathbf{W}}(\mathbf{W}) = \det(\mathbf{W}^T\mathbf{W}) \tag{6}$$

  Captures non-simplex shape and is sensitive to mean offset.

Our NMF algorithm is implemented in a *normalization invariant projected gradient* framework. A gradient descent with projections to the non-negative orthant i.e. no multiplicative updates.
All three types of volume regularizations are implemented

- Simplex volume (4)
- Subspace volume (5) [Miao (2007)]
- Parallelepiped volume (6) [Schachtner (2009)]

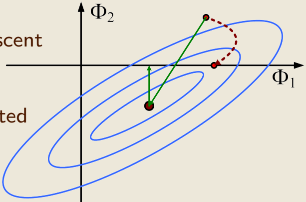


Figure 9 : Projected gradient approach.

NMF with volume regularizations

Synthetic Dataset, $\mathbf{X} \in \mathbb{R}^{145 \times 2000}$.
Employed with different noise levels and all types of volume regularizations.

- Low noise case and simplex volume regularization:

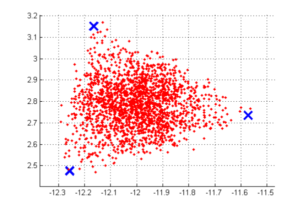


Figure 10 : Synthetic dataset.

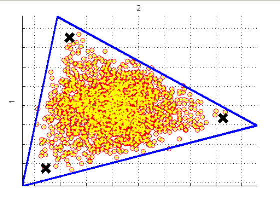


Figure 11 : No regularization

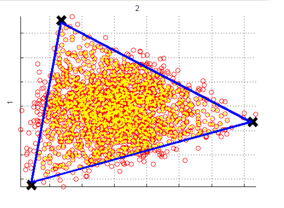


Figure 12 : Optimal regularization.

- MAP solution captures the endmembers (vertices) with success.
- Optimal regularization also depends on the mixing profile of each endmember.
- Parallelepiped [Schachtner (2009)] and simplex volume approaches superior to subspace volume [Miao (2007)] in terms of convergence speed.

Wheat Kernel Dataset
Employed with all types of volume regularizations.

- Expect 3 components from PCA plot. Parallelepiped volume regularization:

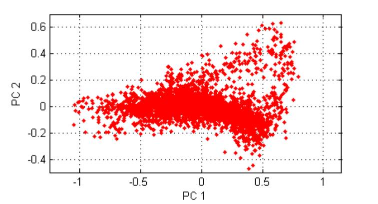


Figure 13 : PCA scatterplot.

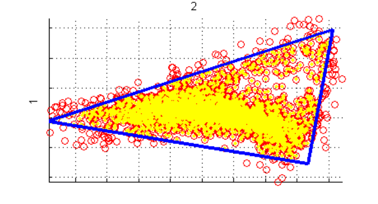


Figure 14 : Captured endmembers.

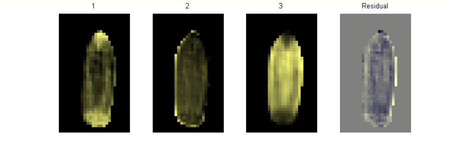


Figure 15 : Image components.

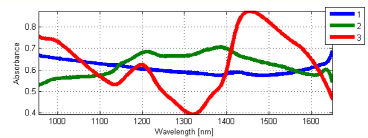


Figure 16 : Pure spectral signatures.

- Constituents successfully identified: *starch*, *oil* and *background*.
- High correlation between constituents leads to difficult identification.

Bayesian NMF

Probabilistic modeling of the bilinear NMF model

$$p(\mathbf{W},\mathbf{H}|\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{W},\mathbf{H},\mathbf{\Theta})p(\mathbf{W}|\mathbf{\Theta})p(\mathbf{H})p(\sigma^2|\mathbf{\Theta})}{p(\mathbf{X})}, \qquad \mathbf{\Theta} = \{\alpha, \sigma^2\} \tag{7}$$

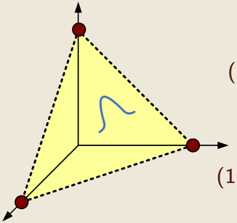
The likelihood becomes a Gaussian distribution due to the bilinear NMF model.

$$p(\mathbf{X}|\mathbf{W},\mathbf{H},\mathbf{\Theta}) = \prod_{n=1}^{N}\prod_{m=1}^{M}\mathcal{N}\left(x_{mn}|\mathbf{W}_{m:}\mathbf{H}_{:n},\sigma^2\right), \tag{8}$$

The non-negative priors are given by

$$p(\mathbf{W}|\mathbf{\Theta}) \propto \begin{cases} \exp\left(-\gamma\det(\mathbf{W}^\top\mathbf{W})\right) & w_{mk} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

$$p(\mathbf{H}|\mathbf{\Theta}) \propto \begin{cases} 1 & h_{kn} \geq 0, \sum_{k=1}^{K} h_{kn} = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$



Our Bayesian NMF model implemented in a *Gibbs sampling* framework.

Bayesian NMF with volume prior

Synthetic Dataset, $\mathbf{X} \in \mathbb{R}^{3\times 3000}$.
Employed with low noise level.

- Comparing with regular NMF volume regularized:

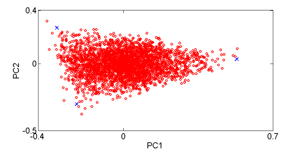


Figure 18 : Dataset.

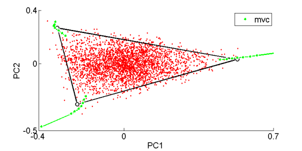


Figure 19 : Subspace projection regularization.

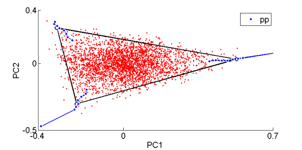


Figure 20 : Parallelepiped regularization.

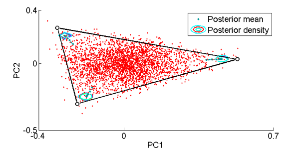


Figure 21 : Bayesian volume NMF.

- Bayes modeling offers posterior mode solution with variance estimations.
- Use MAP solution from NMF as input to Bayesian NMF to reduce burn-in.

Bayesian NMF with volume prior

Wheat kernel dataset

- Expect 3 components from PCA plot.

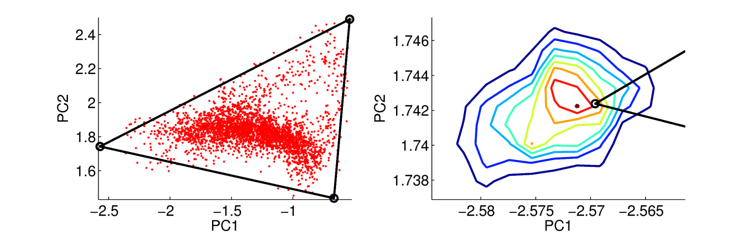


Figure 22 : Bayesian volume NMF applied on wheat kernel dataset.

- Posterior mode is not MAP solution.
- Achieve variance estimation not acquired by regular NMF.
- Non-symmetric posterior distribution - due to attraction to data points within simplex?

Summary

Conclusions

- Hyperspectral image can be decomposed efficiently using NMF algorithm with *volume* regularizations.
- Bayesian framework provide variance estimation leading to confidence intervals of solution.

Current limitations and future work

- Bayesian NMF requires further testing with more sophisticated synthetical data structures.
- Incorporate spatial information.

References

Thank you for your attention....:-)