Estimation of the regularization parameter of an on-line NMF with minimum volume constraint

Ludivine NUS, Sebastian MIRON, David BRIE

CRAN, Université de Lorraine, CNRS Vandoeuvre-lès-Nancy, France
firstname.lastname@univ-lorraine.fr

IEEE SENSOR ARRAY AND MULTICHANNEL SIGNAL PROCESSING WORKSHOP

11th July 2018

Table of contents

Table of contents

Introduction

Developing **real-time** hyperspectral image unmixing methods for industrial applications for controlling and sorting pieces of wood.

→ **On-line** Non-negative Matrix Factorization (NMF):

$$\mathbf{X} \approx \mathbf{S}\mathbf{A}.$$

Sequentially updating the parameters $\mathbf{S}$ and $\mathbf{A}$ as the size of the data matrix $\mathbf{X}$ increases.

→ **On-line MVS-NMF (Minimum Volume Simplex-NMF)** [Nus et al., *SSP '18*].

Introduction

$$\mathscr{J}^{(k+1)}\left(\mathbf{S}^{(k+1)},\mathbf{A}^{(k+1)}\right)=\alpha\sum_{\ell=1}^{k}\left\|\tilde{\mathbf{X}}^{(\ell)}-\tilde{\mathbf{S}}^{(\ell)}\tilde{\mathbf{A}}^{(\ell)}\right\|_F^2+(1-\alpha)\left\|\tilde{\mathbf{X}}^{(k+1)}-\tilde{\mathbf{S}}^{(k+1)}\tilde{\mathbf{A}}^{(k+1)}\right\|_F^2$$
$$+\mu\ln\det\left(\tilde{\mathbf{S}}^{(k+1)T}\tilde{\mathbf{S}}^{(k+1)}\right).$$

- The weighting coefficient α $(0\leq\alpha\leq 1)$ adds some tracking capacity to the algorithm.
- The endmembers vary only slightly between consecutive samples *i.e.* $\tilde{\mathbf{S}}^{(k+1)}\approx\tilde{\mathbf{S}}^{(k)}$ [Bucak and Gunsel, 2009].

→ Estimation of **S** and **A**: minimizing the cost function using a gradient descent technique.
→ The strength of the MVS constraint is controlled by the hyperparameter μ; the main goal is to provide a method to determine it automatically.

Some classical approaches for regularization parameter estimation

- The generalized cross validation method [Golub et al., 1979]. As the NMF problem is bilinear, it is not applicable.
- The L-curve [Hansen, 1992].
  → The optimal value of the hyperparameter: maximum curvature of the L-curve.
  → Multi-objective optimization [Kaufman, 1997].
  → Non-convex and the maximum curvature is not guaranteed to be unique.
- [Belge et al., 2002] proposed the Minimum Distance Criterion (MDC) applied to the L-curve (bi-objective case).
  → Limiting the range of the hyperparameter.
- To overcome these two main drawbacks: MDC on response curve (bi-objective case) defined as the linear plot of regularization cost versus the data fitting [Song et al., 2016].

Table of contents

Response curve

Reformulate the MVS-NMF algorithm as a bi-objective problem:

- Hyperspectral image (slices $\tilde{\mathbf{X}}^{(k)}$, $k = 1, ..., K$), used to learn the value of μ.
- The value of α and the endmembers number are fixed.
- $\mathbf{S}_\mu^{(K)}$ and $\mathbf{A}_\mu^{(K)}$ denote the estimated endmembers and abundances for a given value of μ.

$$\mathscr{J}_1(\mu) = \mathscr{J}_1(\mathbf{S}_\mu^{(K)}, \mathbf{A}_\mu^{(K)}) = \frac{1}{K}\sum_{k=1}^{K} \left\| \tilde{\mathbf{X}}^{(k)} - \tilde{\mathbf{S}}_\mu^{(k)} \tilde{\mathbf{A}}_\mu^{(k)} \right\|_F^2.$$

$$\mathscr{J}_2(\mu) = \mathscr{J}_2(\mathbf{S}_\mu^{(K)}) = \frac{1}{K}\sum_{k=1}^{K} \det\left(\tilde{\mathbf{S}}_\mu^{(k)T} \tilde{\mathbf{S}}_\mu^{(k)} \right).$$

Response curve

Difficulties of the approach:

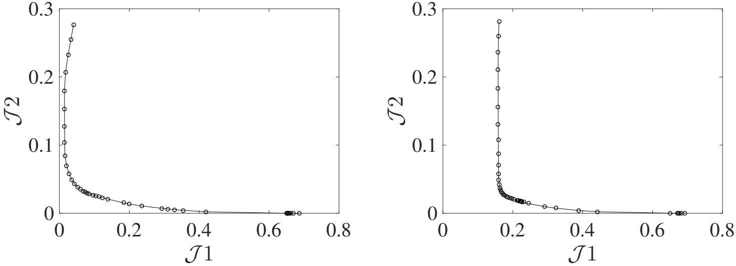
- NMF problems (including MVS-NMF) are bilinear and thus non-convex.
- The obtained solution depends on the initial values of the endmembers and abundances.

What is the shape of a response curve? What is the variability of the response curves for different initial values?

Shape of response curve on a simulated image

- Simulated hyperspectral image consisting in the non-negative mixture of three endmembers, not varying over time.
- None of the three endmembers has any zero value.
  → Decomposition NMF non-unique.
- The three abundance maps are randomly drawn from a continuous uniform distribution on the interval $[0, 1]$.
- The pure pixel condition is approximately fulfilled.
- Noise was added up to have an $SNR = 26dB$.
- $r = 3$ and $\alpha = 0.99$.
- 44 values of $\mu \in [0.0001 \;\; 0.0028]$.

Shape of response curve on a simulated image

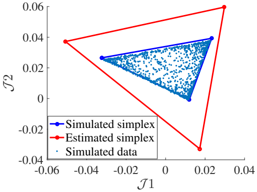


(a) Response curve (noise-free case) (b) Response curve (noisy case)

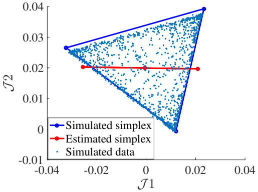
→ The noise is right shifting the response curve by a value which increases with the noise level.

Shape of response curve on a simulated image

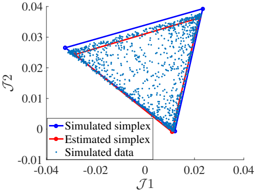
Noisy case



(c) $\mathscr{D} \subset \mathscr{S} (\mu = 0.0001)$

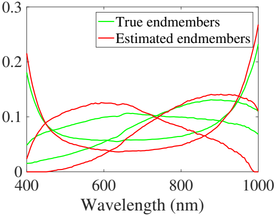


(d) $\mathscr{D} \not\subset \mathscr{S} (\mu = 0.002)$

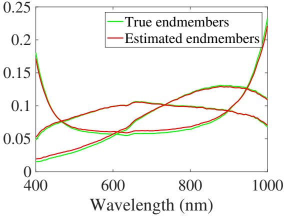


(e) $\mathscr{D} \approx \mathscr{S} (\mu = 0.001)$

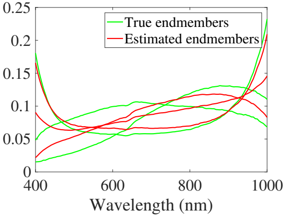
Estimated endmembers for different values of μ



(f) $\mu = 0.0001$



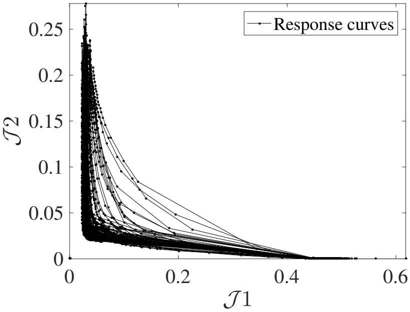
(g) $\mu = 0.001$



(h) $\mu = 0.002$

Influence of the initial conditions

Response curves obtained for 100 different initial values drawn randomly from uniform distribution on the interval $[0, 1]$.



→ Most of the response curves are very similar, but some are deviating from the "mean response curve".

Table of contents

Three strategies

Estimation of the optimal value of μ. Three strategies are considered:

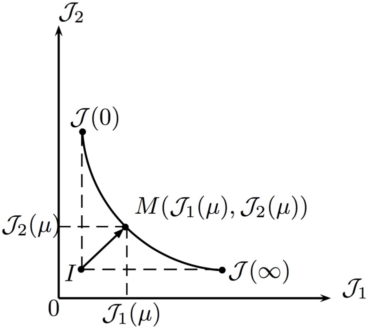
- MDC to the **Pareto Front** estimated from the set of response curves.
- MDC to the **average response curve** to assess how the variability induced by the different initialization is affecting the result.
- MDC to a **single realization** of the response curve.

Pareto Front

- The definition of the Pareto front relies on the notion of domination defined in [Deb, 2001].
- **Non-dominated** or **Pareto optimal** solution: if all other solution in the feasible set has a higher value in at least one of the objectives $\mathscr{J}_i$, with $i \in 1, 2$.
- The image of all the non-dominated solutions is called **Pareto Front**.

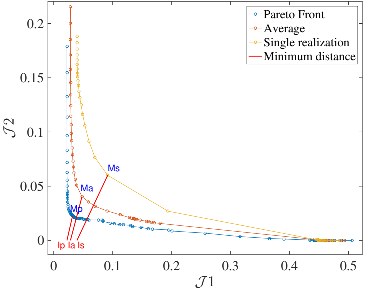
The shape of the Pareto front represents the set of the best achievable tradeoffs between conflicting objectives.

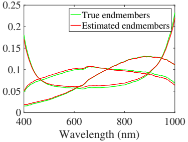
Minimum Distance Criterion



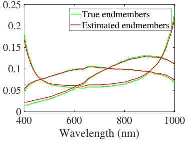
- **The ideal point** I: the point whose coordinates are the minima of the two objective functions.
- **The optimal point** M: the point having the minimum distance to this ideal point.

Results on simulated hyperspectral image



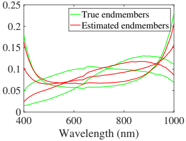


(i) $\mu = 0.00095$



(j) $\mu = 0.0015$

L-curve.

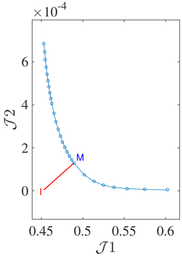


- $\mu = 0.0024$.
- Non-uniqueness.

- Pareto Front: $\mu = 0.00095$.
- Average: $\mu = 0.0011$.
- Single realization: values of μ between 0.001 and 0.0015.

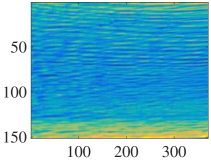
Results on real hyperspectral image

- Single realization strategy.
- Initialization **warm start**.
- 27 values of $\mu \in [0.001 \;\; 0.1]$.

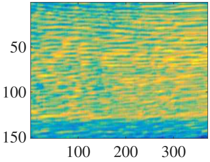


- Value of μ estimated by the MDC: 0.03.

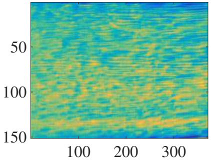
Results on real hyperspectral image



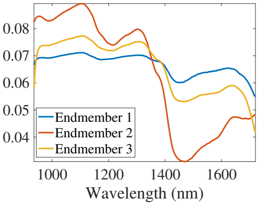
(m) Abundance 1



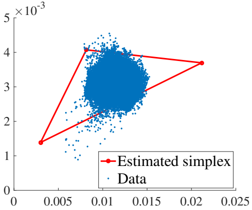
(n) Abundance 2



(o) Abundance 3



(p) Estimated endmembers



(q) Estimated simplex

Limits of the approach for a real hyperspectral image

- Same wood species: oak.
- From one sample to another: the extracted endmembers are very similar and therefore highly correlated.
- Having a "good" initialization: endmembers from an upstream learning.
- Random initialization is complicated: the algorithm converges slowly.

Table of contents

Conclusions:

- Estimation of the regularization hyperparameter for the on-line MVS-NMF.
- Three different MDC-based strategies: all yielding similar results for a simulated image.
- The single realization response curve approach is the most attractive since it presents the lowest computational cost.
- Validation on a real image using a single realization.
- Possibility to learn the regularization parameter on-line (required for industrial applications).

Perspectives:

- Other volume constraints such as the minimum distance between endmembers. Easier to implement and less sensitive.
- Prove the convexity of response curve that guarantees the uniqueness of the MDC.

Table of contents

Belge, M., Kilmer, M. E., and Miller, E. L. (2002).
Efficient determination of multiple regularization parameters in a generalized l-curve framework.
Inverse Problems, 18(4):1161.

Bucak, S. S. and Gunsel, B. (2009).
Incremental subspace learning via non-negative matrix factorization.
Pattern recognition, 42(5):788–797.

Deb, K. (2001).
Multiobjective optimization using evolutionary algorithms., volume 16.
John Wiley & Sons.

Golub, G. H., Heath, M., and Wahba, G. (1979).
Generalized cross-validation as a method for choosing a good ridge parameter.
Technometrics, 21(2):215–223.

Hansen, P. C. (1992).
Analysis of discrete ill-posed problems by means of the l-curve.
SIAM review, 34(4):561–580.

Hoyer, P. O. (2004).
Non-negative matrix factorization with sparseness constraints.
Journal of machine learning research, 5(Nov):1457–1469.

Kaufman, L. (1997).
Regularization of ill-posed problems by envelope guided conjugate gradients.
Journal of Computational and Graphical Statistics, 6(4):451–463.

Lee, D. D. and Seung, H. S. (2001).
Algorithms for non-negative matrix factorization.
In *Advances in neural information processing systems*, pages 556–562.

Miao, L. and Qi, H. (2007).
Endmember extraction from highly mixed data using minimum volume constrained non-negative matrix factorization.
IEEE Transactions on Geoscience and Remote Sensing, 45(3):765–777.

Nus, L., Miron, S., and Brie, D. (2018).
On-line blind unmixing for hyperspectral pushbroom imaging systems.
IEEE Statistical Signal Processing Workshop.

Song, Y., Brie, D., Djermoune, E.-H., and Henrot, S. (2016).
Regularization parameter estimation for non-negative hyperspectral image deconvolution.
IEEE Transactions on Image Processing, 25(11):5316–5330.

Yu, Y. and Sun, W. (2007).
Minimum distance constrained non-negative matrix factorization for the endmember extraction of hyperspectral images.
In *MIPPR 2007: Remote Sensing and GIS Data Processing and Applications; and Innovative Multispectral Technology and Applications*, volume 6790, page 679015. International Society for Optics and Photonics.