Some Recent Advances in Nonnegative Matrix Factorization and their Applications to Hyperspectral Unmixing

Nicolas Gillis

https://sites.google.com/site/nicolasgillis/

Université de Mons
Department of Mathematics and Operational Research

Joint work with Robert Plemmons (Wake Forest U.) and Stephen Vavasis (U. of Waterloo)

International Workshop on Numerical Linear Algebra with Applications in honor of the 75th birthday of Prof. Robert Plemmons

First...

Figure: Bob and I in Lisbon (Third workshop on hyperspectral image and signal processing: evolution in remote sensing –WHISPERS, 2011)

Outline

1. Nonnegative Matrix Factorization (NMF)
   - Definition, motivations & applications
2. Using Underapproximations for NMF
   - Solving NMF recursively with underapproximations
   - Sparse and spatial underapproximations for hyperspectral unmixing
3. Separable and Near-Separable NMF
   - A subclass of efficiently solvable NMF problems
   - Robust algorithms for Near-Separable NMF
   - Application to hyperspectral unmixing

Nonnegative Matrix Factorization (NMF)

Given a matrix $M \in \mathbb{R}_{+}^{m \times n}$ and a factorization rank $r \in \mathbb{N}$, find $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{r \times n}$ such that

$$\min_{U \geq 0, V \geq 0} ||M - UV||_F^2 = \sum_{i,j} (M - UV)_{ij}^2. \qquad \text{(NMF)}$$

NMF is a linear dimensionality reduction technique for nonnegative data :

$$\underbrace{M(:,i)}_{\geq 0} \approx \sum_{k=1}^{r} \underbrace{U(:,k)}_{\geq 0} \underbrace{V(k,i)}_{\geq 0} \qquad \text{for all } i.$$

Why nonnegativity?

$\rightarrow$ **Interpretability**: Nonnegativity constraints lead to a sparse and part-based representation.

$\rightarrow$ **Many applications.** Text mining, hyperspectral unmixing, image processing, community detection, clustering, etc.

Nonnegative Matrix Factorization (NMF)

Given a matrix $M \in \mathbb{R}^{m \times n}_{+}$ and a factorization rank $r \in \mathbb{N}$, find $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{r \times n}$ such that

$$\min_{U \geq 0, V \geq 0} \ ||M - UV||_F^2 = \sum_{i,j} (M - UV)_{ij}^2. \qquad \text{(NMF)}$$

NMF is a linear dimensionality reduction technique for nonnegative data :

$$\underbrace{M(:,i)}_{\geq 0} \approx \sum_{k=1}^{r} \underbrace{U(:,k)}_{\geq 0} \underbrace{V(k,i)}_{\geq 0} \qquad \text{for all } i.$$

Why nonnegativity?

$\rightarrow$ **Interpretability**: Nonnegativity constraints lead to a sparse and part-based representation.

$\rightarrow$ **Many applications.** Text mining, hyperspectral unmixing, image processing, community detection, clustering, etc.

Nonnegative Matrix Factorization (NMF)

Given a matrix $M \in \mathbb{R}^{m\times n}_{+}$ and a factorization rank $r \in \mathbb{N}$, find $U \in \mathbb{R}^{m\times r}$and $V \in \mathbb{R}^{r\times n}$ such that

$$\min_{U\geq 0, V\geq 0} \; ||M - UV||_F^2 = \sum_{i,j} (M - UV)^2_{ij}. \qquad \text{(NMF)}$$

NMF is a linear dimensionality reduction technique for nonnegative data :

$$\underbrace{M(:,i)}_{\geq 0} \approx \sum_{k=1}^{r} \underbrace{U(:,k)}_{\geq 0} \underbrace{V(k,i)}_{\geq 0} \qquad \text{for all } i.$$

Why nonnegativity?

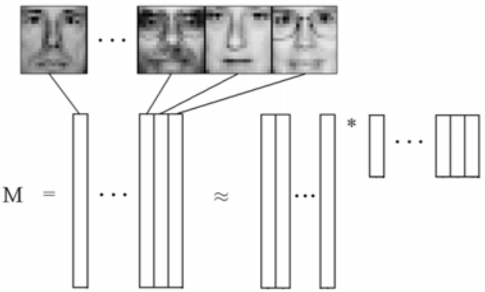
→ **Interpretability**: Nonnegativity constraints lead to a sparse and part-based representation.

→ **Many applications.** Text mining, hyperspectral unmixing, image processing, community detection, clustering, etc.

Application 1: image processing

$U \geq 0$ constraints the basis elements to be nonnegative.

Moreover $V \geq 0$ imposes an additive reconstruction.

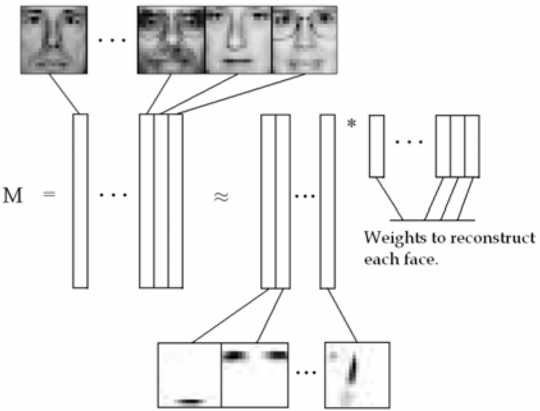


The basis elements extract facial features such as eyes, nose and lips.

Application 1: image processing

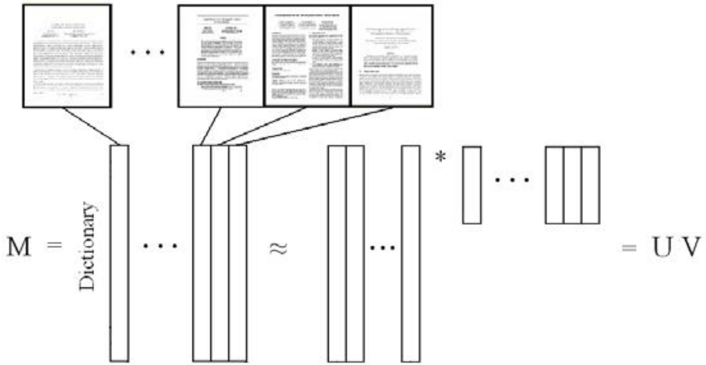
$U \geq 0$ constraints the basis elements to be nonnegative.

Moreover $V \geq 0$ imposes an additive reconstruction.



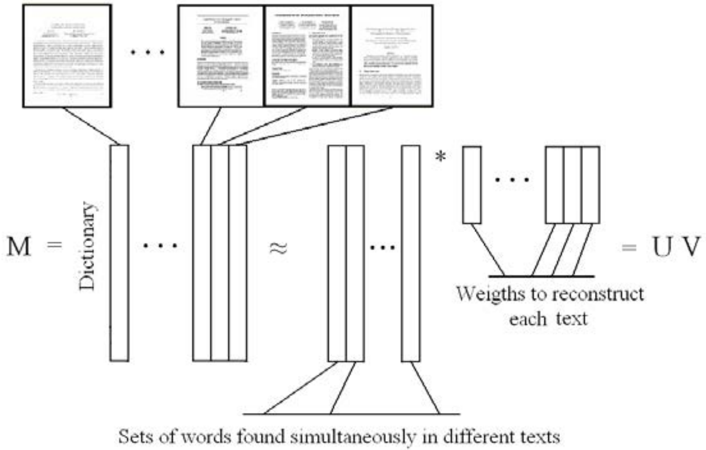
The basis elements extract facial features such as eyes, nose and lips.

Application 2: text mining



- Basis elements allow to recover the different topics;
- Weights allow to assign each text to its corresponding topics.

Application 2: text mining



- ◇ Basis elements allow to recover the different topics;
- ◇ Weights allow to assign each text to its corresponding topics.

Application 3: hyperspectral unmixing

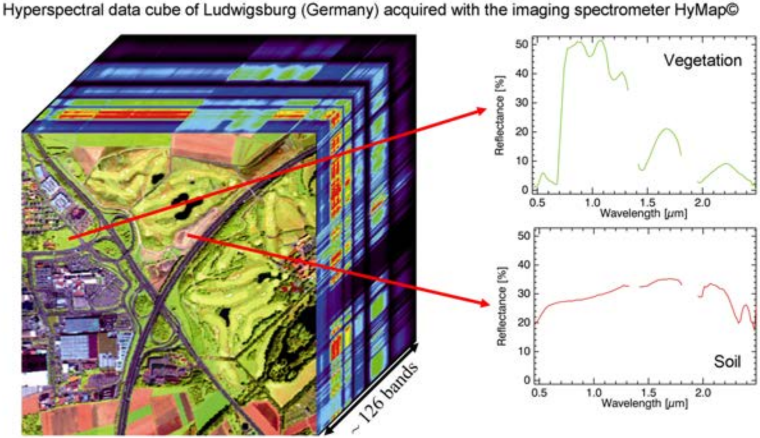
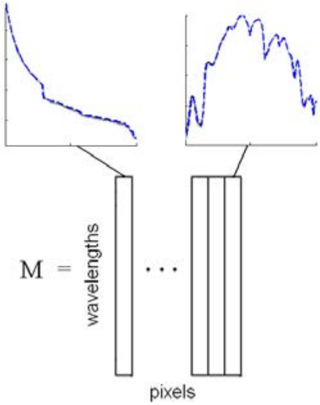


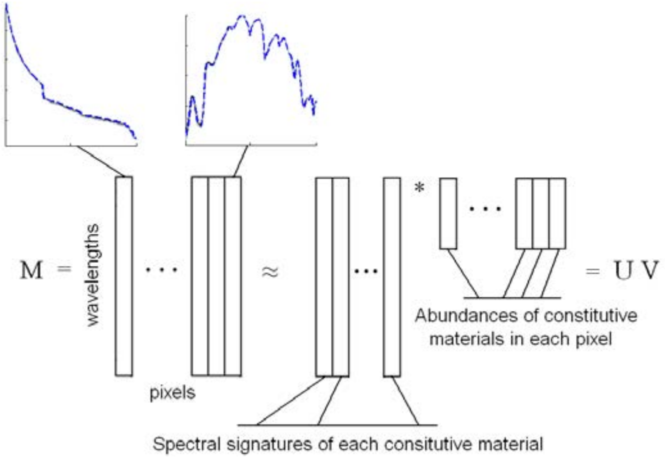
Figure: Hyperspectral image.

Application 3: hyperspectral unmixing



- ◇ Basis elements allow to recover the different materials;
- ◇ Weights allow to know which pixel contains which material.

Application 3: hyperspectral unmixing



- Basis elements allow to recover the different materials;
- Weights allow to know which pixel contains which material.

Application 3: hyperspectral unmixing

Figure: Urban dataset.

Application 3: hyperspectral unmixing

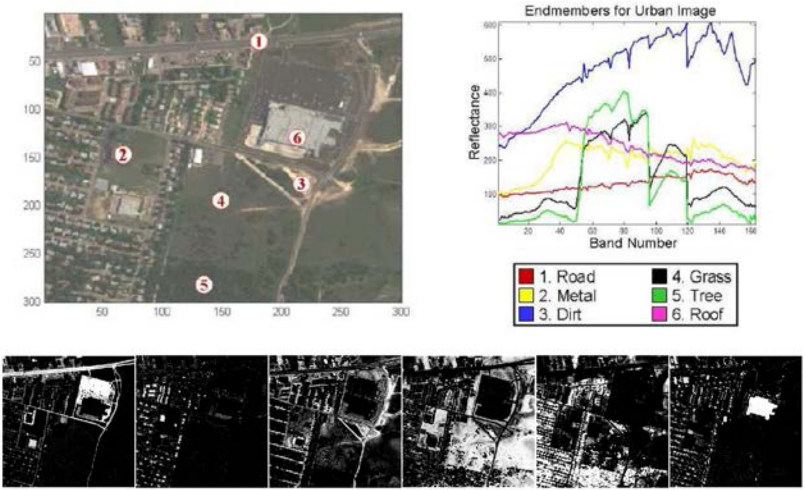


Figure: Urban dataset.

Using Underapproximations for NMF

Issues of using NMF

1. NMF is NP-hard [V09].

2. The optimal solution is, in most cases, non-unique and the problem is ill-posed [G12]. Many variants of NMF impose additional constraints (e.g., sparsity on U, smoothness of V, etc.).

3. In practice, it is difficult to choose the factorization rank (in general, trial and error approach or estimation using the SVD).

A possible way to overcome drawbacks 2. and 3. is to use underapproximation constraints to solve NMF recursively.

[V09] Vavasis, *On the Complexity of Nonnegative Matrix Factorization*, SIAM J. on Optimization, 2009.
[G12] G., *Sparse and Unique Nonnegative Matrix Factorization Through Data Preprocessing*, J. of Machine Learning Research, 2012.

Issues of using NMF

1. NMF is NP-hard [V09].
2. The optimal solution is, in most cases, non-unique and the problem is ill-posed [G12]. Many variants of NMF impose additional constraints (e.g., sparsity on U, smoothness of V, etc.).
3. In practice, it is difficult to choose the factorization rank (in general, trial and error approach or estimation using the SVD).

A possible way to overcome drawbacks 2. and 3. is to use underapproximation constraints to solve NMF recursively.

[V09] Vavasis, *On the Complexity of Nonnegative Matrix Factorization*, SIAM J. on Optimization, 2009.
[G12] G., *Sparse and Unique Nonnegative Matrix Factorization Through Data Preprocessing*, J. of Machine Learning Research, 2012.

Issues of using NMF

1. NMF is NP-hard [V09].
2. The optimal solution is, in most cases, non-unique and the problem is ill-posed [G12]. Many variants of NMF impose additional constraints (e.g., sparsity on U, smoothness of V, etc.).
3. In practice, it is difficult to choose the factorization rank (in general, trial and error approach or estimation using the SVD).

A possible way to overcome drawbacks 2. and 3. is to use underapproximation constraints to solve NMF recursively.

[V09] Vavasis, *On the Complexity of Nonnegative Matrix Factorization*, SIAM J. on Optimization, 2009.
[G12] G., *Sparse and Unique Nonnegative Matrix Factorization Through Data Preprocessing*, J. of Machine Learning Research, 2012.

Issues of using NMF

1. NMF is NP-hard [V09].
2. The optimal solution is, in most cases, non-unique and the problem is ill-posed [G12]. Many variants of NMF impose additional constraints (e.g., sparsity on U, smoothness of V, etc.).
3. In practice, it is difficult to choose the factorization rank (in general, trial and error approach or estimation using the SVD).

A possible way to overcome drawbacks 2. and 3. is to use underapproximation constraints to solve NMF recursively.

[V09] Vavasis, *On the Complexity of Nonnegative Matrix Factorization*, SIAM J. on Optimization, 2009.
[G12] G., *Sparse and Unique Nonnegative Matrix Factorization Through Data Preprocessing*, J. of Machine Learning Research, 2012.

Nonnegative Matrix Underapproximation (NMU)

It is possible to solve NMF recursively, solving at each step

$$\min_{u \geq 0, v \geq 0} ||M - uv^T||_F^2 \quad \text{such that} \quad uv^T \leq M \iff M - uv^T \geq 0.$$

NMU is yet another linear dimensionality reduction technique. However,

- As PCA, it is computed recursively and is well-posed [GG10].
- As NMF, it leads to a separation by parts. Moreover the additional underapproximation constraints enhance this property.
- In the presence of pure-pixels, the NMU recursion is able to detect materials individually [GP11].

[GG10] G., Glineur, *Using Underapproximations for Sparse Nonnegative Matrix Factorization*, Pattern Recognition, 2010.
[GP11] G., Plemmons, *Dimensionality Reduction, Classification, and Spectral Mixture Analysis using Nonnegative Underapproximation*, Optical Engineering, 2011.

Nonnegative Matrix Underapproximation (NMU)

It is possible to solve NMF recursively, solving at each step

$$\min_{u \geq 0, v \geq 0} ||M - uv^T||_F^2 \quad \text{such that} \quad uv^T \leq M \iff M - uv^T \geq 0.$$

NMU is yet another linear dimensionality reduction technique. However,

- As PCA, it is computed recursively and is well-posed [GG10].
- As NMF, it leads to a separation by parts. Moreover the additional underapproximation constraints enhance this property.
- In the presence of pure-pixels, the NMU recursion is able to detect materials individually [GP11].

[GG10] G., Glineur, *Using Underapproximations for Sparse Nonnegative Matrix Factorization*, Pattern Recognition, 2010.
[GP11] G., Plemmons, *Dimensionality Reduction, Classification, and Spectral Mixture Analysis using Nonnegative Underapproximation*, Optical Engineering, 2011.

Nonnegative Matrix Underapproximation (NMU)

It is possible to solve NMF recursively, solving at each step

$$\min_{u \geq 0, v \geq 0} ||M - uv^T||_F^2 \quad \text{such that} \quad uv^T \leq M \iff M - uv^T \geq 0.$$

NMU is yet another linear dimensionality reduction technique. However,

- ◇ As PCA, it is computed recursively and is well-posed [GG10].
- ◇ As NMF, it leads to a separation by parts. Moreover the additional underapproximation constraints enhance this property.
- ◇ In the presence of pure-pixels, the NMU recursion is able to detect materials individually [GP11].

[GG10] G., Glineur, *Using Underapproximations for Sparse Nonnegative Matrix Factorization*, Pattern Recognition, 2010.

[GP11] G., Plemmons, *Dimensionality Reduction, Classification, and Spectral Mixture Analysis using Nonnegative Underapproximation*, Optical Engineering, 2011.

Nonnegative Matrix Underapproximation (NMU)

It is possible to solve NMF recursively, solving at each step

$$\min_{u \geq 0, v \geq 0} ||M - uv^T||_F^2 \quad \text{such that} \quad uv^T \leq M \iff M - uv^T \geq 0.$$

NMU is yet another linear dimensionality reduction technique. However,

- ◇ As PCA, it is computed recursively and is well-posed [GG10].
- ◇ As NMF, it leads to a separation by parts. Moreover the additional underapproximation constraints enhance this property.
- ◇ In the presence of pure-pixels, the NMU recursion is able to detect materials individually [GP11].

[GG10] G., Glineur, *Using Underapproximations for Sparse Nonnegative Matrix Factorization*, Pattern Recognition, 2010.
[GP11] G., Plemmons, *Dimensionality Reduction, Classification, and Spectral Mixture Analysis using Nonnegative Underapproximation*, Optical Engineering, 2011.

Example of NMU on the Urban dataset

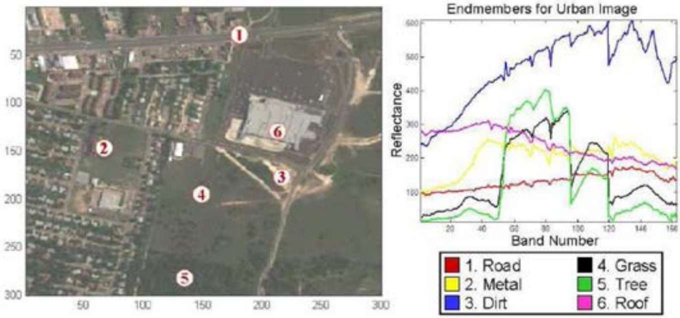
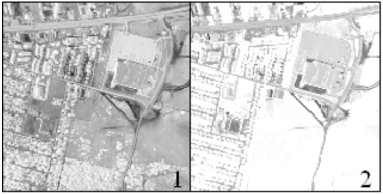


Figure: Hyperspectral data from aircraft - Army Geospatial Center.

Figure: Hyperspectral image from aircraft - Army Geospatial Center - $307 \times 307 \times 162$.

Example of NMU on the Urban dataset

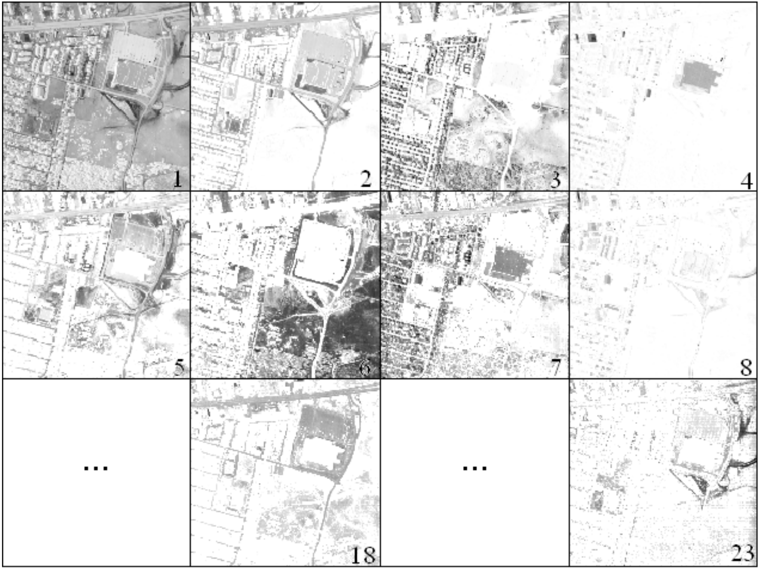
Example of NMU on the Urban dataset



Example of NMU on the Urban dataset

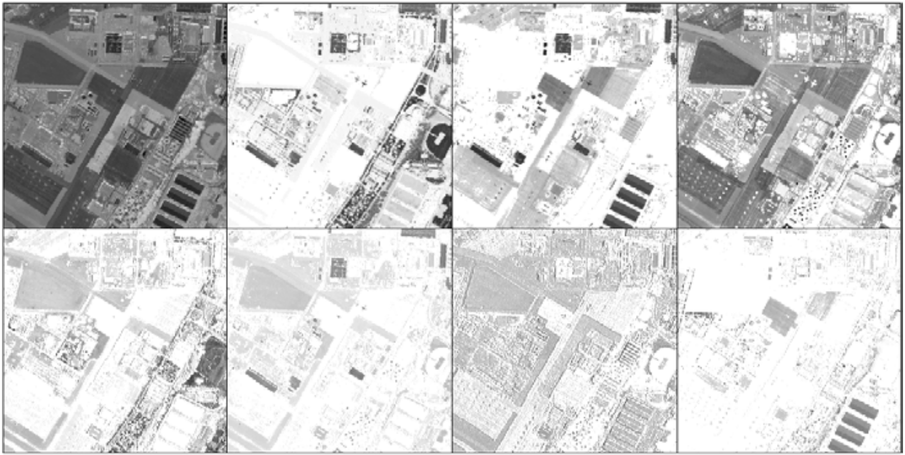
Example of NMU on the Urban dataset

Example of NMU on the Urban dataset



Example on the San Diego Airport dataset

Example on the San Diego Airport dataset



Additional Sparsity Constraints

- With more blur and noise, NMU typically fails to detect materials individually.
- However, it can still be a useful dimensionality reduction technique (e.g., combined with nearest neighbor or k-means).
- In order to detect materials, it is possible to enhance sparsity of the abundance matrix V to extract more localized features. We solve [GP13]

$$\min_{u \geq 0, v \geq 0} ||M - uv^T||_F^2 + \mu ||v||_1$$

$$uv^T \leq M. \qquad \text{(sparse NMU)}$$

[GP13] G., Plemmons, *Nonnegative Matrix Underapproximation and its Application for Hyperspectral Image Analysis*, Lin. Alg. Appl., 2013.

Additional Sparsity Constraints

- With more blur and noise, NMU typically fails to detect materials individually.
- However, it can still be a useful dimensionality reduction technique (e.g., combined with nearest neighbor or k-means).
- In order to detect materials, it is possible to enhance sparsity of the abundance matrix V to extract more localized features. We solve [GP13]

$$\min_{u \geq 0, v \geq 0} ||M - uv^T||_F^2 + \mu ||v||_1$$

$$uv^T \leq M. \qquad \text{(sparse NMU)}$$

[GP13] G., Plemmons, *Nonnegative Matrix Underapproximation and its Application for Hyperspectral Image Analysis*, Lin. Alg. Appl., 2013.

San Diego Airport

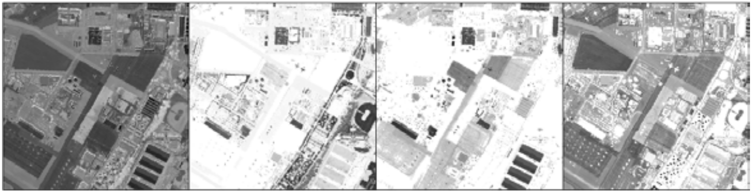


Figure: First four basis elements of NMU

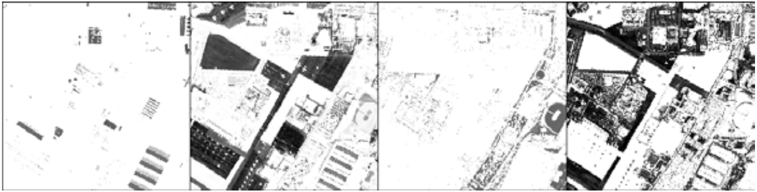


Figure: First four basis elements of sparse NMU

Additional Spatial Constraints

It is also possible to take into account spatial constraints [GPZ12]: neighbor pixels are more likely to contain the same materials.

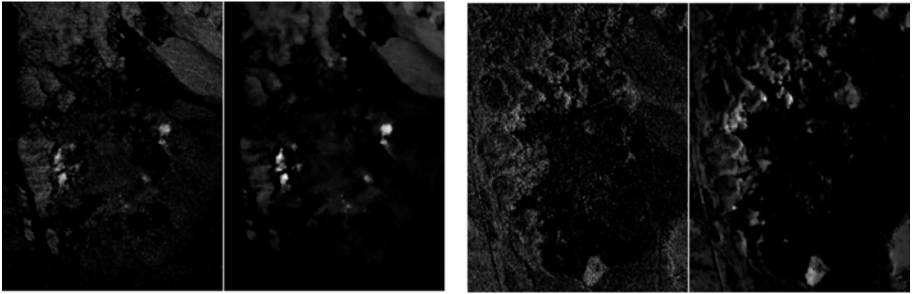


Figure: NMU vs. spatial NMU on the Cuprite data set.

[GPZ12] G., Plemmons, Zhang, *Priors in Sparse Recursive Decompositions of Hyperspectral Images*, Proc. of SPIE, 2012.

Separable and Near-Separable NMF

Can we only solve NMF problems?

Given a matrix $M \in \mathbb{R}_{+}^{m \times n}$ and a factorization rank $r \in \mathbb{N}$, find $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{r \times n}$ such that

$$\min_{U \geq 0, V \geq 0} ||M - UV||_F^2 = \sum_{i,j} (M - UV)_{ij}^2. \tag{NMF}$$

- NMF is NP-hard [V09].
- In practice, it is often satisfactory to use locally optimal solutions for further analysis of the data. In other words, heuristics often solve the problem efficiently with acceptable answers.
- Try to analyze this state of affairs by considering enerative models and algorithms that can recover hidden data.

[V09] Vavasis, *On the Complexity of Nonnegative Matrix Factorization*, SIAM J. on Optimization, 2009.

Can we only solve NMF problems?

Given a matrix $M \in \mathbb{R}_{+}^{m\times n}$ and a factorization rank $r \in \mathbb{N}$, find $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{r\times n}$ such that

$$\min_{U\geq 0, V\geq 0} ||M - UV||_F^2 = \sum_{i,j} (M - UV)_{ij}^2. \tag{NMF}$$

- NMF is NP-hard [V09].
- In practice, it is often satisfactory to use locally optimal solutions for further analysis of the data. In other words, heuristics often solve the problem efficiently with acceptable answers.
- Try to analyze this state of affairs by considering enerative models and algorithms that can recover hidden data.

[V09] Vavasis, *On the Complexity of Nonnegative Matrix Factorization*, SIAM J. on Optimization, 2009.

Separability Assumption

For NMF, it is possible to compute optimal solutions in polynomial time, given that the input data matrix M satisfies a (rather strong) condition: separability [AGKM12].

The nonnegative matrix M is r-separable if and only if

> *there exists an NMF $(U, V) \geq 0$ of rank r with $M = UV$ where each column of U is equal to a column of M.*

[AGKM12] Arora, Ge, Kannan, Moitra, *Computing a Nonnegative Matrix Factorization – Provably*, STOC 2012.

Separability Assumption

For NMF, it is possible to compute optimal solutions in polynomial time, given that the input data matrix M satisfies a (rather strong) condition: separability [AGKM12].

The nonnegative matrix M is r-separable if and only if

there exists an NMF $(U, V) \geq 0$ of rank r with $M = UV$ where each column of U is equal to a column of M.

[AGKM12] Arora, Ge, Kannan, Moitra, *Computing a Nonnegative Matrix Factorization – Provably*, STOC 2012.

Is separability a reasonable assumption?

- **Text mining**: for each topic, there is a 'pure' document on that topic, or, for each topic, there is a 'pure' word used only by that topic.
  [KSK13] Kumar, Sindhwani, Kambadur, *Fast Conical Hull Algorithms for Near-separable Non-Negative Matrix Factorization*, ICML 2013.
  [AG+13] Arora, Ge, Halpern, Mimno, Moitra, Sontag, Wu, Zhu, *A Practical Algorithm for Topic Modeling with Provable Guarantees*, ICML 2013.
  [DRIS13] Ding, Rohban, Ishwar, Saligrama, *Topic Discovery through Data Dependent and Random Projections*, ICML 2013.

- **Hyperspectral unmixing**: separability is particularly natural: for each constitutive material, there is a 'pure' pixel containing only that material. This is the so called pure-pixel assumption which is widely used in hyperspectral imaging.

- **General image processing**: No.

Is separability a reasonable assumption?

◇ **Text mining**: for each topic, there is a 'pure' document on that topic, or, for each topic, there is a 'pure' word used only by that topic.

[KSK13] Kumar, Sindhwani, Kambadur, *Fast Conical Hull Algorithms for Near-separable Non-Negative Matrix Factorization*, ICML 2013.
[AG+13] Arora, Ge, Halpern, Mimno, Moitra, Sontag, Wu, Zhu, *A Practical Algorithm for Topic Modeling with Provable Guarantees*, ICML 2013.
[DRIS13] Ding, Rohban, Ishwar, Saligrama, *Topic Discovery through Data Dependent and Random Projections*, ICML 2013.

◇ **Hyperspectral unmixing**: separability is particularly natural: for each constitutive material, there is a 'pure' pixel containing only that material. This is the so called pure-pixel assumption which is widely used in hyperspectral imaging.

◇ **General image processing**: No.

Is separability a reasonable assumption?

- **Text mining**: for each topic, there is a 'pure' document on that topic, or, for each topic, there is a 'pure' word used only by that topic.

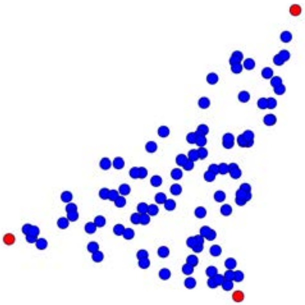
  [KSK13] Kumar, Sindhwani, Kambadur, *Fast Conical Hull Algorithms for Near-separable Non-Negative Matrix Factorization*, ICML 2013.
  [AG+13] Arora, Ge, Halpern, Mimno, Moitra, Sontag, Wu, Zhu, *A Practical Algorithm for Topic Modeling with Provable Guarantees*, ICML 2013.
  [DRIS13] Ding, Rohban, Ishwar, Saligrama, *Topic Discovery through Data Dependent and Random Projections*, ICML 2013.

- **Hyperspectral unmixing**: separability is particularly natural: for each constitutive material, there is a 'pure' pixel containing only that material. This is the so called pure-pixel assumption which is widely used in hyperspectral imaging.

- **General image processing**: No.

Geometric Interpretation of Separable NMF

After normalization, the columns of M, U and V sum to one: the columns of U are the vertices of the convex hull of the columns of M.



$$M \text{ is } r\text{-separable} \quad \iff \quad M = U[I_r, V']\Pi,$$

for some $V' \geq 0$ and some permutation Π.

Separable NMF with Noise

$$\tilde{M} = U[I_r, V']\Pi + N, \qquad \text{where } N \text{ is the noise.}$$

Near-Separable NMF: Noise and Conditioning

We will assume that the noise is bounded (but otherwise arbitrary):

$$||N(:,i)||_1 \leq \epsilon, \quad \text{for all } i,$$

and some dependence on some condition number is unavoidable:

Near-Separable NMF: Noise and Conditioning

We will assume that the noise is bounded (but otherwise arbitrary):

$$||N(:,i)||_1 \leq \epsilon, \qquad \text{for all } i,$$

and some dependence on some condition number is unavoidable:

Near-Separable NMF: Noise and Conditioning

We will assume that the noise is bounded (but otherwise arbitrary):

$$||N(:,i)||_1 \leq \epsilon, \quad \text{for all } i,$$

and some dependence on some condition number is unavoidable:

Fast and Robust Algorithm for Separable NMF

$$M = UV = U[I_r, V'] = [U, UV']\,\Pi,$$

where $V' \geq 0$ and its column sum to one.

Observation 1. The maximum of a strongly convex function f over a polytope is attained at a vertex:

$$\max_{1 \leq i \leq n} f(M(:,i)) = \max_{1 \leq i \leq r} f(U(:,i)).$$

Observation 2. This property is robust: for $\tilde{M} = M + N$, if $\tilde{M}(:,i)$ is the column of $\tilde{M}$ maximizing f, then there exists p such that

$$||\tilde{M}(:,i) - U(:,p)||_2 \leq \mathcal{O}\left(\epsilon\, \kappa^2(U)\right).$$

Observation 3. Pre-multiplying M preserves separability:

$$PM = (PU)\,[I_r, V']\,\Pi.$$

Fast and Robust Algorithm for Separable NMF

$$M = UV = U[I_r, V'] = [U, UV']\,\Pi,$$

where $V' \geq 0$ and its column sum to one.

Observation 1. The maximum of a strongly convex function f over a polytope is attained at a vertex:

$$\max_{1 \leq i \leq n} f(M(:,i)) = \max_{1 \leq i \leq r} f(U(:,i)).$$

Observation 2. This property is robust: for $\tilde{M} = M + N$, if $\tilde{M}(:,i)$ is the column of $\tilde{M}$ maximizing f, then there exists p such that

$$||\tilde{M}(:,i) - U(:,p)||_2 \leq \mathcal{O}\left(\epsilon\, \kappa^2(U)\right).$$

Observation 3. Pre-multiplying M preserves separability:

$$PM = (PU)\,[I_r, V']\,\Pi.$$

Fast and Robust Algorithm for Separable NMF

$$M = UV = U[I_r, V'] = [U, UV']\,\Pi,$$

where $V' \geq 0$ and its column sum to one.

Observation 1. The maximum of a strongly convex function f over a polytope is attained at a vertex:

$$\max_{1 \leq i \leq n} f(M(:,i)) = \max_{1 \leq i \leq r} f(U(:,i)).$$

Observation 2. This property is robust: for $\tilde{M} = M + N$, if $\tilde{M}(:,i)$ is the column of $\tilde{M}$ maximizing f, then there exists p such that

$$||\tilde{M}(:,i) - U(:,p)||_2 \leq \mathcal{O}\left(\epsilon\, \kappa^2(U)\right).$$

Observation 3. Pre-multiplying M preserves separability:

$$PM = (PU)\,[I_r, V']\,\Pi.$$

Successive Projection Algorithm (SPA)

0: $R = \tilde{M}$.

For $i = 1 : r$

% Identify the column of R maximizing $||.||_2$.

1: $j^* = \mathrm{argmax}_j \, ||R(:,j)||_2$ and $U(:,i) = \tilde{M}(:,j^*)$.

% Project all columns of R onto its orthogonal complement.

2: $R \leftarrow \left(I - \frac{R(:,j^*)R(:,j^*)^T}{||R(:,j^*)||_2^2} \right) R$.

end

It is essentially modified Gram-Schmidt with column pivoting.

Theorem ([GV12]). If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\kappa^2(U)}\right)$, SPA leads to an NMF (U, V) s.t.

$$||\tilde{M} - UV||_2 \leq \mathcal{O}\left(\epsilon\kappa^2(U)\right).$$

Advantages. Extremely fast, no parameter.

Drawbacks. Requires U to be full rank; bound is weak.

[GV12] G., Vavasis, *Fast and Robust Recursive Algorithms for Separable Nonnegative Matrix Factorization*, to appear in IEEE Trans. Patt. Anal. Mach. Intell.

Successive Projection Algorithm (SPA)

0: $R = \tilde{M}$.

For $i = 1 : r$

% Identify the column of R maximizing $||.||_2$.

1: $j^* = \text{argmax}_j \, ||R(:,j)||_2$ and $U(:,i) = \tilde{M}(:,j^*)$.

% Project all columns of R onto its orthogonal complement.

2: $R \leftarrow \left(I - \frac{R(:,j^*)R(:,j^*)^T}{||R(:,j^*)||_2^2} \right) R$.

end

It is essentially modified Gram-Schmidt with column pivoting.

Theorem ([GV12]). If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\kappa^2(U)}\right)$, SPA leads to an NMF (U, V) s.t.

$$||\tilde{M} - UV||_2 \leq \mathcal{O}\left(\epsilon \kappa^2(U)\right).$$

Advantages. Extremely fast, no parameter.

Drawbacks. Requires U to be full rank; bound is weak.

[GV12] G., Vavasis, *Fast and Robust Recursive Algorithms for Separable Nonnegative Matrix Factorization*, to appear in IEEE Trans. Patt. Anal. Mach. Intell.

Pre-conditioning for More Robust SPA

Observation 3. Pre-multiplying M preserves separability:

$$P\tilde{M} \;=\; P\left(U[I_r, V'] + N\right) \;=\; (PU)\,[I, V'] \,+\, PN.$$

Ideally, $P = U^{-1}$ so that $\kappa(PU) = 1$.

Solving the minimum volume ellipsoid centered at the origin and containing all the columns of $\tilde{M}$ (which is SDP representable)

$$\min_{A \in \mathbb{S}^r_+} \;\log\det(A)^{-1} \;\text{ s.t. }\; \tilde{m}_i{}^T A \tilde{m}_i \leq 1 \quad \forall i,$$

allows to approximate U^{-1}: in fact, $A^* \approx U^{-T}U^{-1}$.

Theorem ([GV13]). If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{r\sqrt{r}}\right)$, preconditioned SPA leads to an NMF (U, V) s.t. $||\tilde{M} - UV||_2 \leq \mathcal{O}\left(\epsilon\kappa(U)\right)$.

[GV13] G., Vavasis, *SDP-based Preconditioning for More Robust Near-Separable NMF*, arXiv:1310.2273.

Pre-conditioning for More Robust SPA

Observation 3. Pre-multiplying M preserves separability:

$$P\tilde{M} \;=\; P\left(U[I_r, V'] + N\right) \;=\; (PU)\,[I, V'] \;+\; PN.$$

Ideally, $P = U^{-1}$ so that $\kappa(PU) = 1$.

Solving the minimum volume ellipsoid centered at the origin and containing all the columns of $\tilde{M}$ (which is SDP representable)

$$\min_{A \in \mathbb{S}^r_+} \; \log\det(A)^{-1} \;\text{ s.t. }\; \tilde{m}_i{}^T A \tilde{m}_i \leq 1 \quad \forall\, i,$$

allows to approximate U^{-1}: in fact, $A^* \approx U^{-T}U^{-1}$.

Theorem ([GV13]). If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{r\sqrt{r}}\right)$, preconditioned SPA leads to an NMF (U, V) s.t. $||\tilde{M} - UV||_2 \leq \mathcal{O}\left(\epsilon\kappa(U)\right)$.

[GV13] G., Vavasis, *SDP-based Preconditioning for More Robust Near-Separable NMF*, arXiv:1310.2273.

Pre-conditioning for More Robust SPA

Observation 3. Pre-multiplying M preserves separability:

$$P\,\tilde{M} \;=\; P\,(U[I_r, V'] + N) \;=\; (PU)\,[I, V'] \,+\, PN.$$

Ideally, $P = U^{-1}$ so that $\kappa(PU) = 1$.

Solving the minimum volume ellipsoid centered at the origin and containing all the columns of $\tilde{M}$ (which is SDP representable)

$$\min_{A \in \mathbb{S}^r_+} \; \log\det(A)^{-1} \;\text{ s.t. }\; \tilde{m}_i{}^T A \tilde{m}_i \leq 1 \quad \forall\, i,$$

allows to approximate U^{-1}: in fact, $A^* \approx U^{-T}U^{-1}$.

Theorem ([GV13]). If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{r\sqrt{r}}\right)$, preconditioned SPA leads to an NMF (U, V) s.t. $||\tilde{M} - UV||_2 \leq \mathcal{O}\left(\epsilon\kappa(U)\right)$.

[GV13] G., Vavasis, *SDP-based Preconditioning for More Robust Near-Separable NMF*, arXiv:1310.2273.

Pre-conditioning for More Robust SPA

Observation 3. Pre-multiplying M preserves separability:

$$P\tilde{M} \; = \; P\,(U[I_r, V'] + N) \; = \; (PU)\,[I, V'] \, + \, PN.$$

Ideally, $P = U^{-1}$ so that $\kappa(PU) = 1$.
Solving the minimum volume ellipsoid centered at the origin and containing all the columns of $\tilde{M}$ (which is SDP representable)

$$\min_{A \in \mathbb{S}^r_+} \;\; \log\det(A)^{-1} \;\; \text{s.t.} \;\; \tilde{m}_i{}^T A \tilde{m}_i \leq 1 \quad \forall\, i,$$

allows to approximate U^{-1}: in fact, $A^* \approx U^{-T} U^{-1}$.

Theorem ([GV13]). If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{r\sqrt{r}}\right)$, preconditioned SPA leads to an NMF (U, V) s.t. $||\tilde{M} - UV||_2 \leq \mathcal{O}\left(\epsilon\kappa(U)\right)$.

[GV13] G., Vavasis, *SDP-based Preconditioning for More Robust Near-Separable NMF*, arXiv:1310.2273.

Synthetic data sets

- Each entry of $U \in \mathbb{R}_+^{20\times 20}$ uniform in $[0,1]$; each column normalized.
- The other columns of M are the middle points of the columns of U (hence there are $\binom{20}{2} = 190$).
- The noise moves the middle points toward the outside of the convex hull of the column of U.

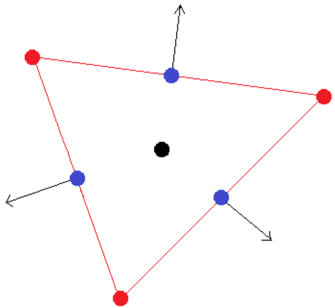


Figure: Example for $r = 3$.

Results for the synthetic data sets

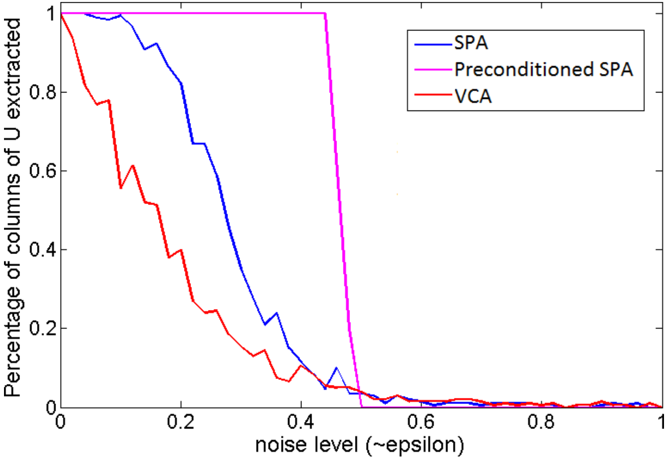


Figure: Average of the percentage of columns correctly extracted depending on the noise level (for each noise level, 10 matrices are generated).

Hubble telescope hyperspectral image

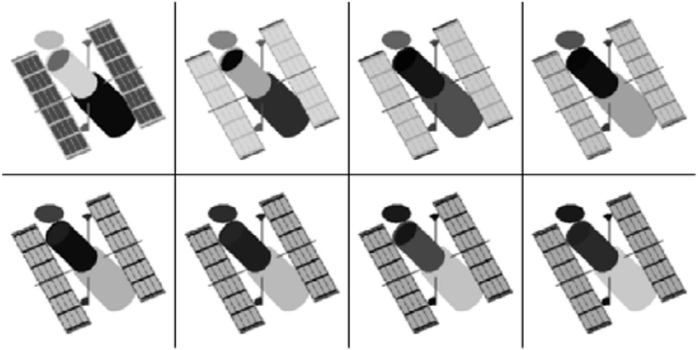


Figure: Sample of images for the Hubble telescope hyperspectral image with 100 spectral bands and 128×128 pixels.

Hubble telescope hyperspectral image

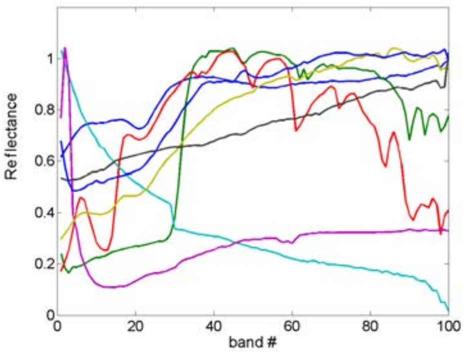


Figure: Spectral signatures extracted by SPA, corresponding to constitutive materials (matrix U with $\kappa(U) = 115$).

Hubble telescope hyperspectral image

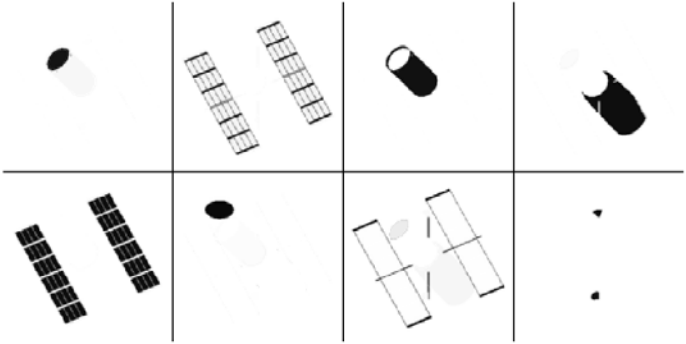


Figure: Reconstructed abundance maps (matrix H).

Hubble telescope with blur and noise

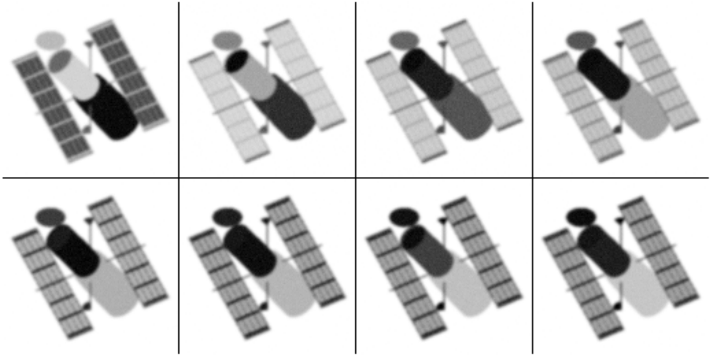


Figure: Sample of images for the Hubble telescope.

Hubble telescope with blur and noise

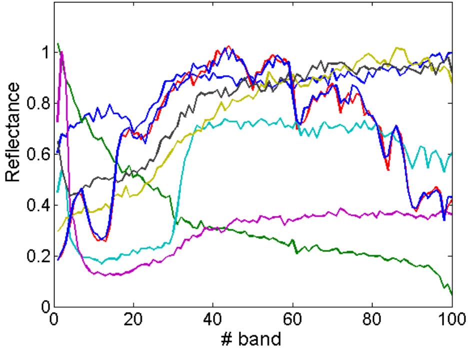


Figure: Spectral signatures extracted by SPA, corresponding to constitutive materials (matrix U).

Hubble telescope with blur and noise

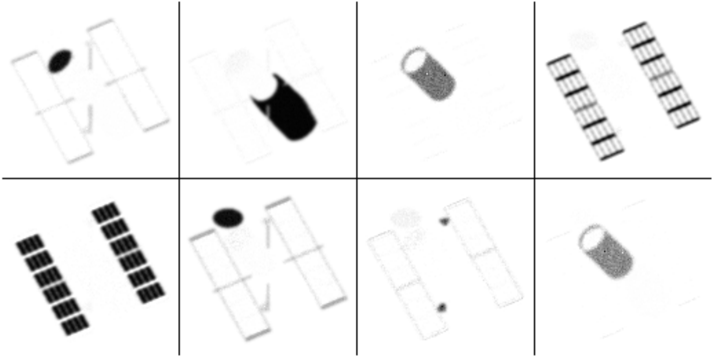


Figure: Reconstructed abundance maps (matrix V). With the blur and noise, SPA fails to identify good columns.

Hubble telescope with blur and noise

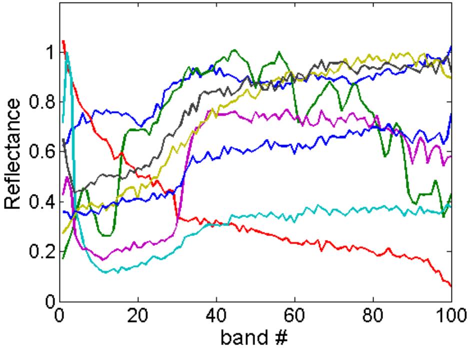


Figure: Spectral signatures extracted by preconditioned SPA, corresponding to constitutive materials (matrix U).

Hubble telescope with blur and noise

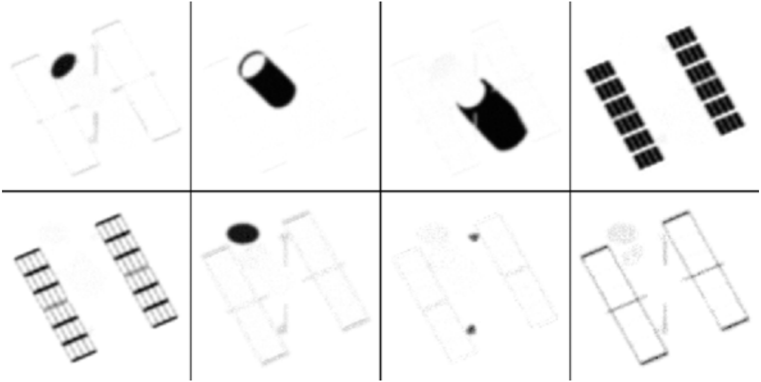


Figure: Reconstructed abundance maps (matrix V). With the blur and noise, preconditioned SPA is able to identify the right columns.

Conclusion

1. Nonnegative matrix factorization (NMF)
   - Easily interpretable linear dimensionality reduction technique for nonnegative data, with *many* applications

2. Nonnegative matrix underapproximation (NMU)
   - Underapproximations allow to solve NMF recursively
   - additional sparsity and regularity constraints leads to better separation by parts

3. Separable NMF
   - Separability makes NMF problems efficiently solvable
   - Need for fast, practical and robust algorithms
   - SPA, a recursive algorithm for separable NMF
   - SDP preconditionning can be used to make SPA significantly more robust to noise

Conclusion

1. Nonnegative matrix factorization (NMF)
   - Easily interpretable linear dimensionality reduction technique for nonnegative data, with *many* applications
2. Nonnegative matrix underapproximation (NMU)
   - Underapproximations allow to solve NMF recursively
   - additional sparsity and regularity constraints leads to better separation by parts
3. Separable NMF
   - Separability makes NMF problems efficiently solvable
   - Need for fast, practical and robust algorithms
   - SPA, a recursive algorithm for separable NMF
   - SDP preconditionning can be used to make SPA significantly more robust to noise

Conclusion

1. Nonnegative matrix factorization (NMF)
   - Easily interpretable linear dimensionality reduction technique for nonnegative data, with *many* applications
2. Nonnegative matrix underapproximation (NMU)
   - Underapproximations allow to solve NMF recursively
   - additional sparsity and regularity constraints leads to better separation by parts
3. Separable NMF
   - Separability makes NMF problems efficiently solvable
   - Need for fast, practical and robust algorithms
   - SPA, a recursive algorithm for separable NMF
   - SDP preconditionning can be used to make SPA significantly more robust to noise

Thank you for your attention

and Thanks to Bob!

Code and papers available on
`https://sites.google.com/site/nicolasgillis/`.