Enhancing Pure-Pixel Identification Performance via Preconditioning

Nicolas Gillis

Department of Mathematics and Operational Research
University of Mons, Belgium

Joint work with

Stephen A. Vavasis, University of Waterloo
Wing-Kin Ma, The Chinese University of Hong Kong

MAHI 2014

Outline

1. Nonnegative Matrix Factorization (NMF)
2. NMF under the Pure-Pixel Assumption (a.k.a. Separability)
3. The Successive Projection Algorithm (SPA)
4. Preconditioning

Nonnegative Matrix Factorization (NMF)

Given a matrix $M \in \mathbb{R}^{p \times n}_{+}$ and a factorization rank $r \ll \min(p,n)$, find $U \in \mathbb{R}^{p \times r}$ 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?

→ **Interpretability**: Nonnegativity constraints lead to easily interpretable factors (and a sparse and part-based representation).
→ **Many applications.** image processing, text mining, hyperspectral unmixing, community detection, clustering, etc.

Nonnegative Matrix Factorization (NMF)

Given a matrix $M \in \mathbb{R}^{p \times n}_{+}$ and a factorization rank $r \ll \min(p,n)$, find $U \in \mathbb{R}^{p \times r}$ 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?

→ **Interpretability**: Nonnegativity constraints lead to easily interpretable factors (and a sparse and part-based representation).
→ **Many applications.** image processing, text mining, hyperspectral unmixing, community detection, clustering, etc.

Nonnegative Matrix Factorization (NMF)

Given a matrix $M \in \mathbb{R}^{p \times n}_{+}$ and a factorization rank $r \ll \min(p, n)$, find $U \in \mathbb{R}^{p \times r}$ 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?

→ **Interpretability**: Nonnegativity constraints lead to easily interpretable factors (and a sparse and part-based representation).
→ **Many applications.** image processing, text mining, hyperspectral unmixing, community detection, clustering, etc.

Example 1: Blind hyperspectral unmixing

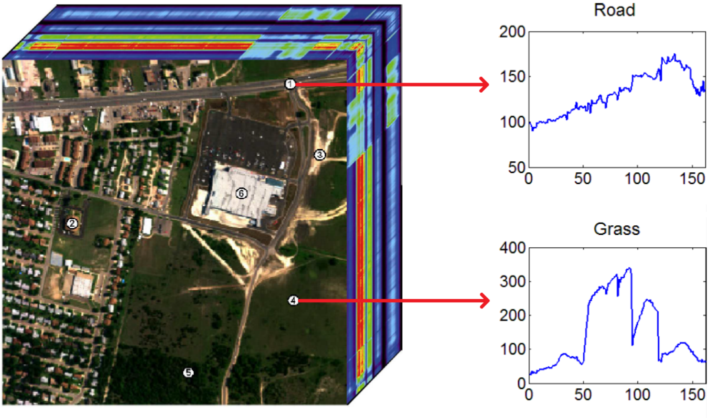
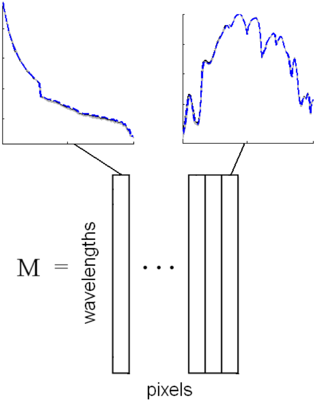


Figure : Urban hyperspectral image with 162 spectral bands and 307-by-307 pixels.

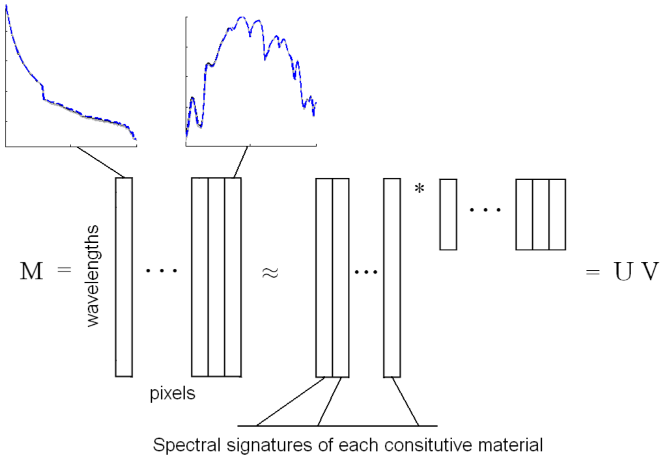
Example 1: Blind hyperspectral unmixing with NMF



◇ Basis elements allow to recover the different endmembers: $U \geq 0$;

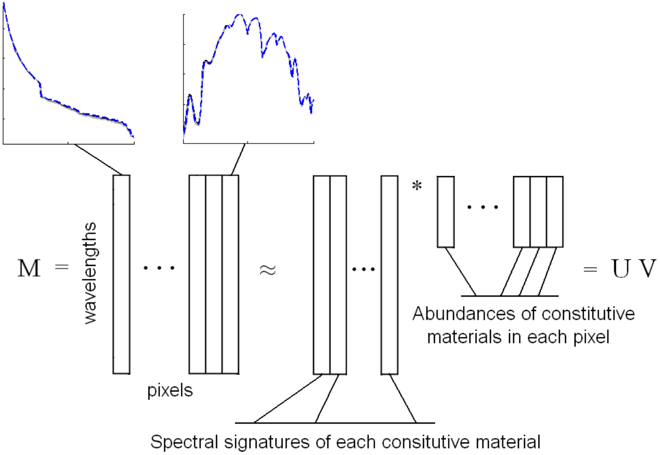
◇ Abundances of the endmembers in each pixel: $V \geq 0$.

Example 1: Blind hyperspectral unmixing with NMF



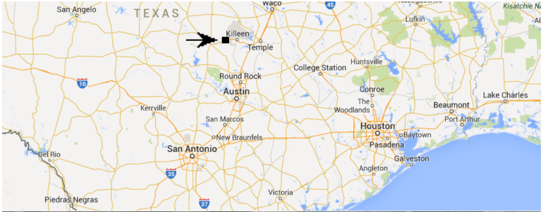
- ◇ Basis elements allow to recover the different endmembers: $U \geq 0$;
- ◇ Abundances of the endmembers in each pixel: $V \geq 0$.

Example 1: Blind hyperspectral unmixing with NMF



- Basis elements allow to recover the different endmembers: $U \geq 0$;
- Abundances of the endmembers in each pixel: $V \geq 0$.

Urban hyperspectral image



Urban hyperspectral image

$$\underbrace{M(:,j)}_{\substack{\text{spectral signature}\\ \text{of } j\text{th pixel}}} \approx \sum_{k=1} \underbrace{U(:,k)} \qquad \underbrace{V(k,j)} \quad .$$

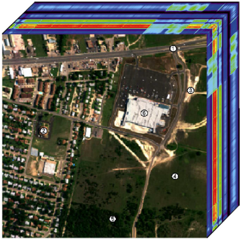
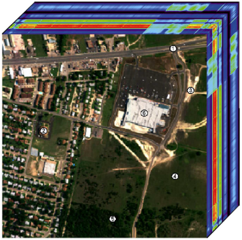


Figure : Decomposition of the Urban dataset.

Urban hyperspectral image

$$\underbrace{M(:,j)}_{\substack{\text{spectral signature}\\ \text{of } j\text{th pixel}}} \approx \sum_{k=1} \underbrace{U(:,k)}_{\substack{\text{spectral signature}\\ \text{of } k\text{th endmember}}} \quad \underbrace{V(k,j)} \quad .$$



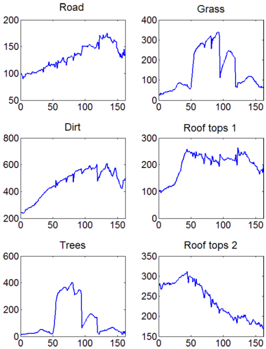
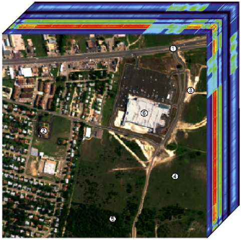
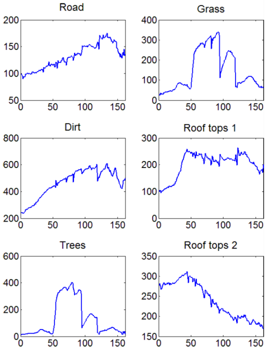


Figure : Decomposition of the Urban dataset.

Urban hyperspectral image

$$\underbrace{M(:,j)}_{\substack{\text{spectral signature}\\ \text{of } j\text{th pixel}}} \approx \sum_{k=1} \underbrace{U(:,k)}_{\substack{\text{spectral signature}\\ \text{of } k\text{th endmember}}} \underbrace{V(k,j)}_{\substack{\text{abundance of } k\text{th endmember}\\ \text{in } j\text{th pixel}}} .$$





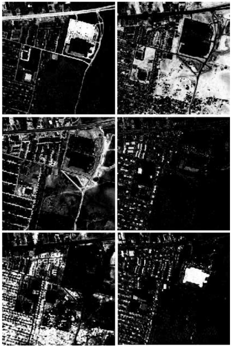
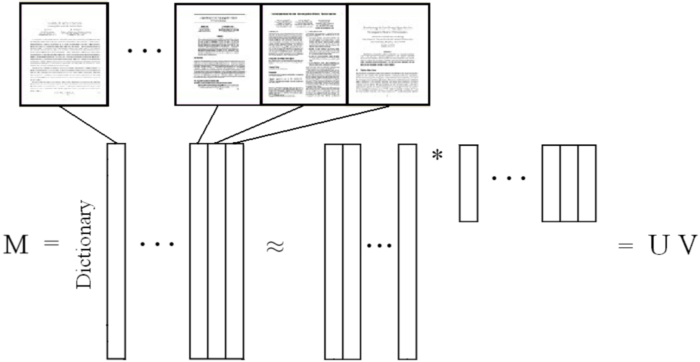


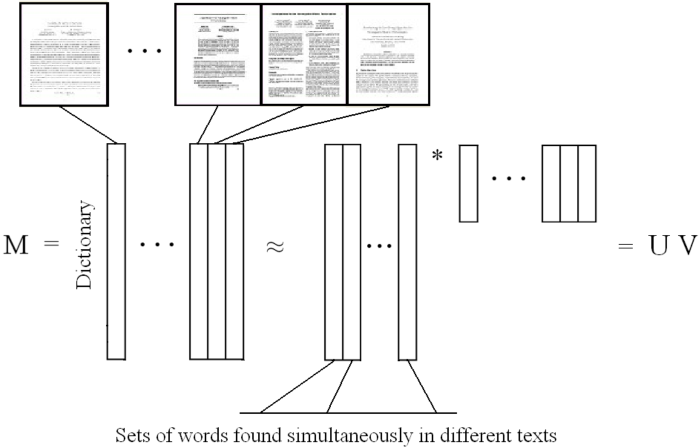
Figure : Decomposition of the Urban dataset.

Example 2: topic recovery and document classification



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

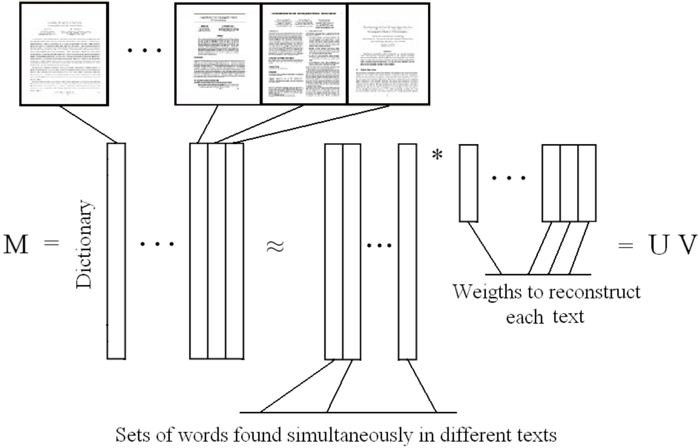
Example 2: topic recovery and document classification



◇ Basis elements allow to recover the different topics;

◇ Weights allow to assign each text to its corresponding topics.

Example 2: topic recovery and document classification

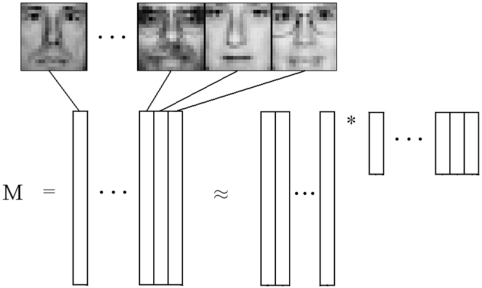


- ◇ Basis elements allow to recover the different topics;
- ◇ Weights allow to assign each text to its corresponding topics.

Exemple 3: feature extraction and classification

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

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

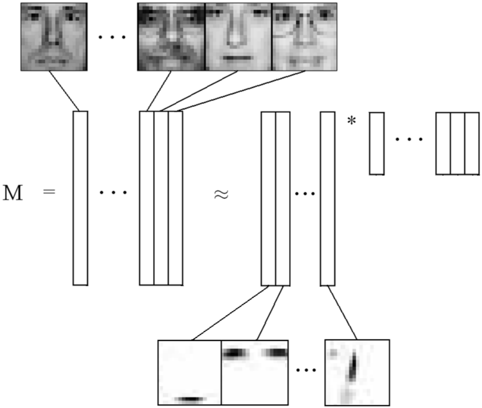


The basis elements extract facial features such as eyes, nose and lips.

Exemple 3: feature extraction and classification

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

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

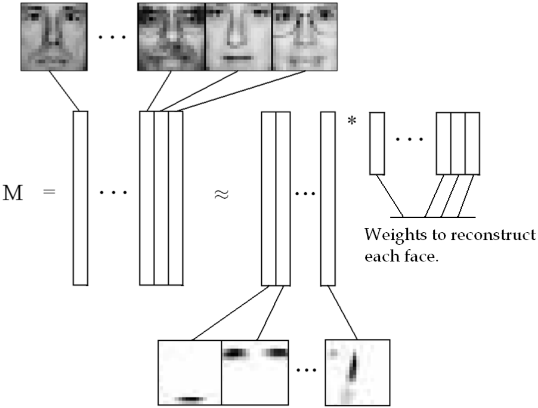


The basis elements extract facial features such as eyes, nose and lips.

Exemple 3: feature extraction and classification

$U \geq 0$ 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.

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

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

- ◇ NMF is NP-hard [V09], and highly ill-posed [G12].
- ◇ In practice, it is often satisfactory to use regularization along with locally optimal solutions for further analysis of the data.
- ◇ However, under the pure-pixel assumption, the problem becomes tractable.

[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.

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

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

- NMF is NP-hard [V09], and highly ill-posed [G12].
- In practice, it is often satisfactory to use regularization along with locally optimal solutions for further analysis of the data.
- However, under the pure-pixel assumption, the problem becomes tractable.

[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.

Pure-Pixel Assumption a.k.a. Separability

Separability of M = there exists an NMF $(U, V) \geq 0$ with $M = UV$ where each column of U is equal to a column of M. [AGKM12]

This is the pure-pixel assumption: columns of U are the spectral signatures of the endmembers present in the hyperspectral image.

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

Another important application. In document classification: for each topic, there is a 'pure' word used only by that topic (an 'anchor' word).

[KSK13] Kumar, Sindhwani, Kambadur, *Fast Conical Hull Algorithms for Near-separable Non-Negative Matrix Factorization*, ICML 2013.
[A+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.

Pure-Pixel Assumption a.k.a. Separability

Separability of M = there exists an NMF $(U, V) \geq 0$ with $M = UV$ where each column of U is equal to a column of M. [AGKM12]

This is the pure-pixel assumption: columns of U are the spectral signatures of the endmembers present in the hyperspectral image.

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

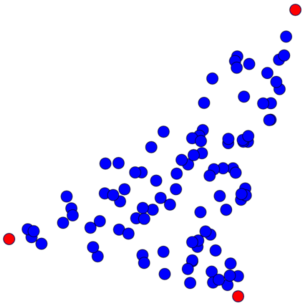
Another important application. In document classification: for each topic, there is a 'pure' word used only by that topic (an 'anchor' word).

[KSK13] Kumar, Sindhwani, Kambadur, *Fast Conical Hull Algorithms for Near-separable Non-Negative Matrix Factorization*, ICML 2013.
[A+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.

Geometric Interpretation

The columns of U are the vertices of the convex hull of the columns of M:

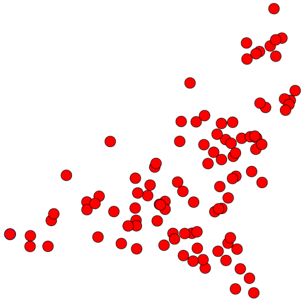
$$M(:,j) \ = \sum_{k=1}^{r} U(:,k)V(k,j)\ \forall j, \quad \text{where} \sum_{k=1}^{r} V(k,j) = 1, V \geq 0.$$



Geometric Interpretation with Noise

The columns of U are the vertices of the convex hull of the columns of M:

$$M(:,j) \approx \sum_{k=1}^{r} U(:,k)V(k,j) \; \forall j, \quad \text{where} \sum_{k=1}^{r} V(k,j) = 1, V \geq 0.$$



Goal: theoretical analysis of the robustness to noise of pure-pixel search algorithms

Key Parameters: Noise and Conditioning

We assume

$$M \; = \; U[I_r, \; V']\Pi + N,$$

where Π is a permutation and N is the noise.

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

$$||N(:,j)||_2 \leq \epsilon, \qquad \text{for all } j,$$

and some dependence on the conditioning $\kappa(U) = \frac{\sigma_{\max}(U)}{\sigma_{\min}(U)}$ is unavoidable:

Key Parameters: Noise and Conditioning

We assume

$$M \;=\; U[I_r,\; V']\Pi + N,$$

where Π is a permutation and N is the noise.

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

$$||N(:,j)||_2 \leq \epsilon, \quad \text{for all } j,$$

and some dependence on the conditioning $\kappa(U) = \frac{\sigma_{\max}(U)}{\sigma_{\min}(U)}$ is unavoidable:

Key Parameters: Noise and Conditioning

We assume

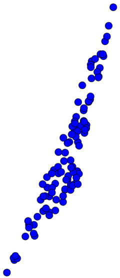
$$M = U[I_r, V']\Pi + N,$$

where Π is a permutation and N is the noise.

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

$$||N(:,j)||_2 \leq \epsilon, \quad \text{for all } j,$$

and some dependence on the conditioning $\kappa(U) = \frac{\sigma_{\max}(U)}{\sigma_{\min}(U)}$ is unavoidable:



Key Parameters: Noise and Conditioning

We assume

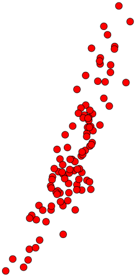
$$M = U[I_r,\ V']\Pi + N,$$

where Π is a permutation and N is the noise.

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

$$||N(:,j)||_2 \leq \epsilon, \quad \text{for all } j,$$

and some dependence on the conditioning $\kappa(U) = \frac{\sigma_{\max}(U)}{\sigma_{\min}(U)}$ is unavoidable:



Successive Projection Algorithm (SPA)

0: Initially $\mathcal{K} = \emptyset$.

For $i = 1 : r$

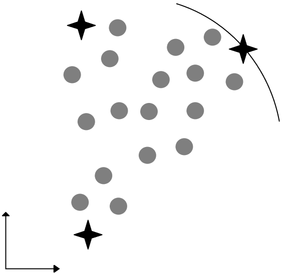
1: Find $j^* = \text{argmax}_j ||M(:,j)||$.

2: $\mathcal{K} = \mathcal{K} \cup \{j^*\}$.

3: $M \leftarrow \left(I - uu^T\right) M$ where $u = \frac{M(:,j^*)}{||M(:,j^*)||_2}$.

end

$\sim$modified Gram-Schmidt with column pivoting.



Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\kappa^2(U)}\right)$, SPA satisfies

$$||U - M(:,\mathcal{K})|| = \max_{1 \leq k \leq r} ||U(:,k) - M(:,\mathcal{K}(k))|| \leq \mathcal{O}\left(\epsilon \kappa^2(U)\right).$$

Advantages. Extremely fast, no parameter.

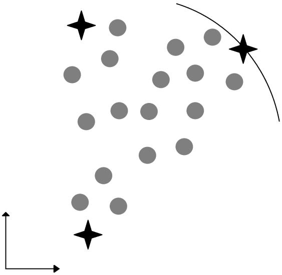
Drawbacks. Requires U to be full rank; bound is weak.

[GV14] G., Vavasis, *Fast and Robust Recursive Algorithms for Separable Nonnegative Matrix Factorization*, IEEE Trans. Patt. Anal. Mach. Intell. 36 (4), pp. 698-714, 2014.

Successive Projection Algorithm (SPA)

0: Initially $\mathcal{K} = \emptyset$.
For $i = 1 : r$
1: Find $j^* = \text{argmax}_j ||M(:,j)||$.
2: $\mathcal{K} = \mathcal{K} \cup \{j^*\}$.
3: $M \leftarrow \left(I - uu^T\right) M$ where $u = \frac{M(:,j^*)}{||M(:,j^*)||_2}$.
end

$\sim$modified Gram-Schmidt with column pivoting.



Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\kappa^2(U)}\right)$, SPA satisfies

$$||U - M(:,\mathcal{K})|| = \max_{1 \leq k \leq r} ||U(:,k) - M(:,\mathcal{K}(k))|| \leq \mathcal{O}\left(\epsilon\kappa^2(U)\right).$$

Advantages. Extremely fast, no parameter.

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

[GV14] G., Vavasis, *Fast and Robust Recursive Algorithms for Separable Nonnegative Matrix Factorization*, IEEE Trans. Patt. Anal. Mach. Intell. 36 (4), pp. 698-714, 2014.

Pre-conditioning for More Robust SPA

Observation. Pre-multiplying M preserves separability:

$$PM = P(U[I_r, V']\Pi + N) = (PU)[I_r, V']\Pi + PN.$$

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

$$\min_{A \in \mathbb{S}^r_+} \log\det(A)^{-1} \text{ s.t. } {m_j}^T A m_j \leq 1 \quad \forall j,$$

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

Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{r\sqrt{r}}\right)$, preconditioned SPA satisfies

$$||U - M(:,\mathcal{K})|| \leq \mathcal{O}(\epsilon\kappa(U)).$$

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

Pre-conditioning for More Robust SPA

Observation. Pre-multiplying M preserves separability:

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

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

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

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

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

Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{r\sqrt{r}}\right)$, preconditioned SPA satisfies

$$||U - M(:, \mathcal{K})|| \leq \mathcal{O}\left(\epsilon\kappa(U)\right).$$

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

Pre-conditioning for More Robust SPA

Observation. Pre-multiplying M preserves separability:

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

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

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

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

Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{r\sqrt{r}}\right)$, preconditioned SPA satisfies

$$||U - M(:,\mathcal{K})|| \leq \mathcal{O}\left(\epsilon\kappa(U)\right).$$

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

Pre-conditioning for More Robust SPA

Observation. Pre-multiplying M preserves separability:

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

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

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

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

Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{r\sqrt{r}}\right)$, preconditioned SPA satisfies

$$||U - M(:, \mathcal{K})|| \leq \mathcal{O}\left(\epsilon\kappa(U)\right).$$

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

Geometric Interpretation

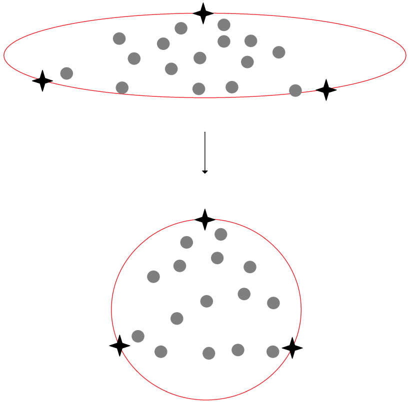


Figure : Geometric Interpretation of the SDP-based Preconditioning.

See also Mizutani, *Ellipsoidal Rounding for Nonnegative Matrix Factorization Under Noisy Separability*, JMLR, 2014.

Computational Cost

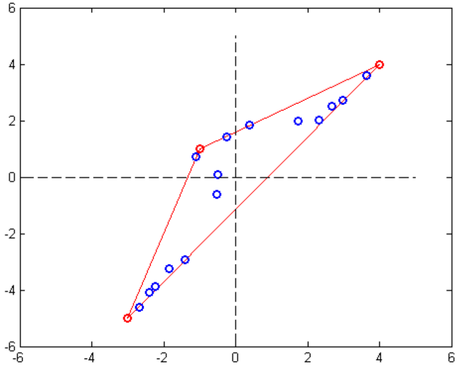
Solving the minimum volume ellipsoid centered at the origin and containing all the columns of M is SDP representable:

$$\min_{A \in \mathbb{S}^r_+} \log\det(A)^{-1} \text{ s.t. } {m_j}^T A m_j \leq 1 \quad \forall j.$$

There are $\mathcal{O}(r^2)$ variables and n constraints. Using an active-set method (only $\mathcal{O}(r^2)$ constraints active at optimality), one can solve the problem for $r \sim 50$ on a standard laptop.

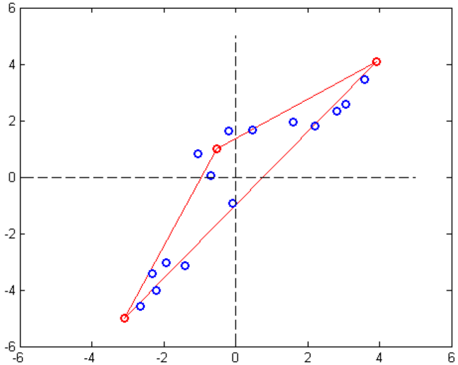
Can we do something faster?

Faster Preconditioning 1: Pre-whitening



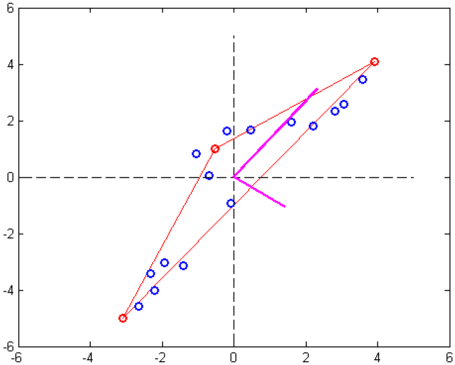
$$M = U \, [I_r \, V'] \Pi$$

Faster Preconditioning 1: Pre-whitening



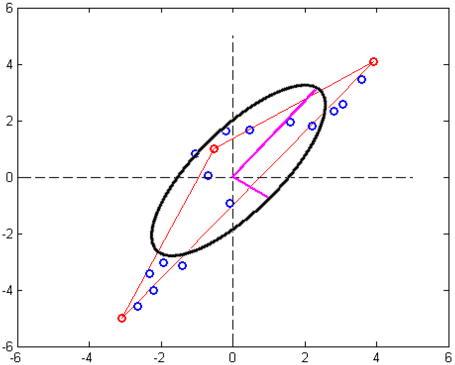
$$M = U\,[I_r\,V']\Pi + N$$

Faster Preconditioning 1: Pre-whitening



$$M = U\,[I_r\; V']\Pi + N \approx U_r \Sigma_r V_r^T$$

Faster Preconditioning 1: Pre-whitening



$$M = U\,[I_r\,V']\Pi + N \approx U_r \Sigma_r V_r^T$$

Prewhitening: $P = \Sigma_r^{-1} U_r^T$ (noise filtering + whitening).

Pre-whitening - Robustness

Computational cost. Compute the truncated SVD of M in $O(mnr^2)$ + SPA in $O(nr^2)$. This is much cheaper.

Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\mathbf{n^{3/2}}}\right)$, SPA satisfies

$$||U - M(:,\mathcal{K})|| \leq \mathcal{O}\left(\epsilon \mathbf{n^{3/2}} \kappa(U)\right).$$

Worst-case scenario is extremely bad as, in practice, $n \sim 10^6$.

The model does not rule out one endmember being present in a very small proportion (e.g., only present in one pixel, the pure one).

Pre-whitening - Robustness

Computational cost. Compute the truncated SVD of M in $O(mnr^2)$ + SPA in $O(nr^2)$. This is much cheaper.

Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\mathbf{n^{3/2}}}\right)$, SPA satisfies

$$||U - M(:,\mathcal{K})|| \leq \mathcal{O}\left(\epsilon \mathbf{n^{3/2}} \kappa(U)\right).$$

Worst-case scenario is extremely bad as, in practice, $n \sim 10^6$.

The model does not rule out one endmember being present in a very small proportion (e.g., only present in one pixel, the pure one).

Pre-whitening - Robustness

Computational cost. Compute the truncated SVD of M in $O(mnr^2)$ + SPA in $O(nr^2)$. This is much cheaper.

Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\mathbf{n^{3/2}}}\right)$, SPA satisfies

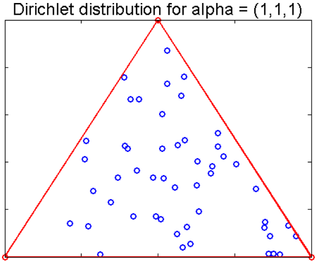
$$||U - M(:, \mathcal{K})|| \leq \mathcal{O}\left(\epsilon \mathbf{n^{3/2}} \kappa(U)\right).$$

Worst-case scenario is extremely bad as, in practice, $n \sim 10^6$.

The model does not rule out one endmember being present in a very small proportion (e.g., only present in one pixel, the pure one).

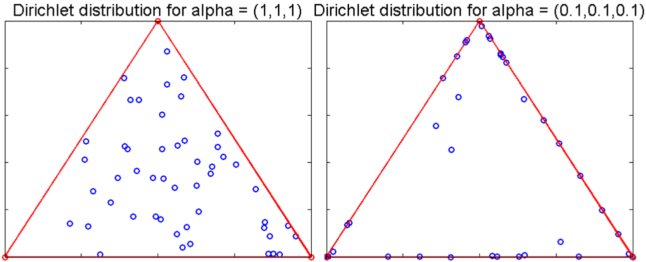
Pre-whitening - Generative Model

Model: U full rank, $V(:,j)$ Dirichlet ($\alpha \in \mathbb{R}^r_+$), Gaussian Noise, $n \to \infty$.



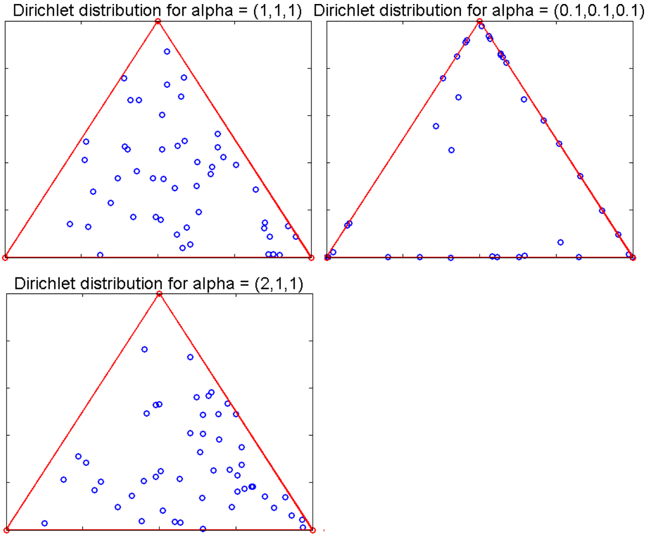
Pre-whitening - Generative Model

Model: U full rank, $V(:,j)$ Dirichlet ($\alpha \in \mathbb{R}^r_+$), Gaussian Noise, $n \to \infty$.



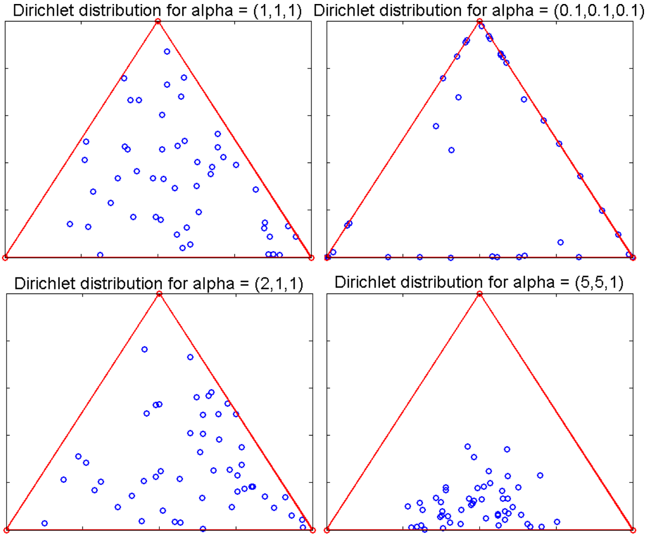
Pre-whitening - Generative Model

Model: U full rank, $V(:,j)$ Dirichlet ($\alpha \in \mathbb{R}^r_+$), Gaussian Noise, $n \to \infty$.



Pre-whitening - Generative Model

Model: U full rank, $V(:,j)$ Dirichlet ($\alpha \in \mathbb{R}^r_+$), Gaussian Noise, $n \to \infty$.



Pre-whitening - Generative Model

Let $\beta = \left(\frac{\alpha_{\max} + ||\alpha||_2^2}{\alpha_{\min}} \right)^{3/2}$. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\beta} \right)$, SPA satisfies

$$||U - M(:, \mathcal{K})|| \leq \mathcal{O}\left(\epsilon \kappa(U) \beta \right).$$

The parameter β is large when $\alpha_{\max} \gg \alpha_{\min}$.

If endmembers are present in similar proportions: $\gamma \approx \alpha_i \; \forall i$, we have

$$\beta \approx (1 + r\gamma)^{3/2}.$$

$\rightarrow$ This makes sense: robustness decreases (β increases) when r and γ increase.

$\rightarrow$ If $\gamma \approx \mathcal{O}(1)$: slightly less robust than SDP (r^2 vs. $r^{3/2}$).

$\rightarrow$ If $\gamma \approx \frac{1}{r}$: slightly more robust than SDP (r vs. $r^{3/2}$).

Conclusion: pre-whitening is effective when endmembers are present in similar proportions.

Pre-whitening - Generative Model

Let $\beta = \left(\frac{\alpha_{\max}+||\alpha||_2^2}{\alpha_{\min}}\right)^{3/2}$. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\beta}\right)$, SPA satisfies

$$||U - M(:,\mathcal{K})|| \leq \mathcal{O}\left(\epsilon\kappa(U)\beta\right).$$

The parameter β is large when $\alpha_{\max} \gg \alpha_{\min}$.

If endmembers are present in similar proportions: $\gamma \approx \alpha_i \ \forall i$, we have

$$\beta \approx (1 + r\gamma)^{3/2}.$$

→ This makes sense: robustness decreases (β increases) when r and γ increase.

→ If $\gamma \approx \mathcal{O}(1)$: slightly less robust than SDP (r^2 vs. $r^{3/2}$).

→ If $\gamma \approx \frac{1}{r}$: slightly more robust than SDP (r vs. $r^{3/2}$).

Conclusion: pre-whitening is effective when endmembers are present in similar proportions.

Pre-whitening - Generative Model

Let $\beta = \left(\frac{\alpha_{\max}+||\alpha||_2^2}{\alpha_{\min}}\right)^{3/2}$. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\beta}\right)$, SPA satisfies

$$||U - M(:,\mathcal{K})|| \leq \mathcal{O}\left(\epsilon\kappa(U)\beta\right).$$

The parameter β is large when $\alpha_{\max} \gg \alpha_{\min}$.

If endmembers are present in similar proportions: $\gamma \approx \alpha_i \; \forall i$, we have

$$\beta \approx (1 + r\gamma)^{3/2}.$$

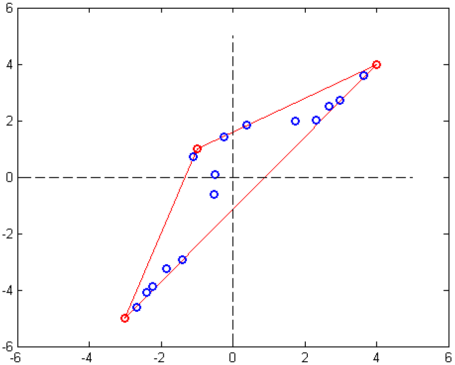
$\rightarrow$ This makes sense: robustness decreases (β increases) when r and γ increase.

$\rightarrow$ If $\gamma \approx \mathcal{O}(1)$: slightly less robust than SDP (r^2 vs. $r^{3/2}$).

$\rightarrow$ If $\gamma \approx \frac{1}{r}$: slightly more robust than SDP (r vs. $r^{3/2}$).

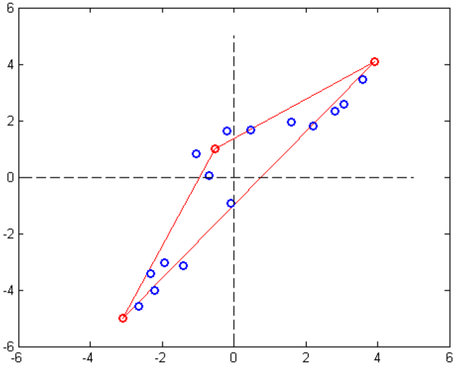
Conclusion: pre-whitening is effective when endmembers are present in similar proportions.

Faster Preconditioning 2: SPA-based



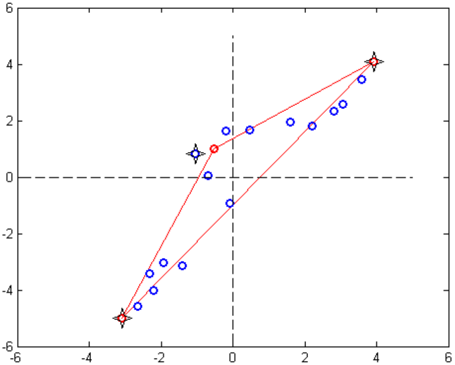
$$M = U\,[I_r\, V']\Pi$$

Faster Preconditioning 2: SPA-based



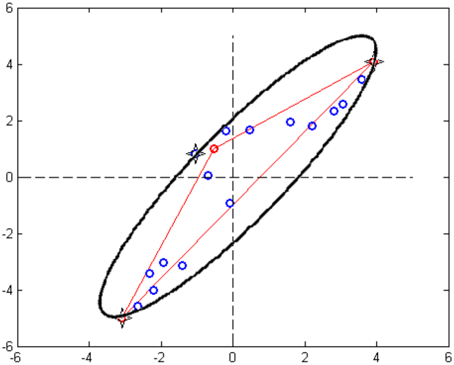
$$M = U\,[I_r\,V']\Pi + N$$

Faster Preconditioning 2: SPA-based



$$M = U\,[I_r\,V']\Pi + N \approx M(:,\mathcal{K})\tilde{V}$$

Faster Preconditioning 2: SPA-based



$$M = U\,[I_r\,V']\Pi + N \approx M(:,\mathcal{K})\tilde{V}$$

SPA-based: $P = M(:,\mathcal{K})^{\dagger}$ (pinv), where $\mathcal{K} = \text{SPA}(M,r)$.

SPA-based Preconditioning: Robustness

Computational cost. Compute SPA twice in $O(mnr)$ + SVD of an r-by-r matrix in $O(r^3)$. This is extremely cheap, like SPA.

Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\kappa^2(U)}\right)$, SPA satisfies

$$||U - M(:,\mathcal{K})|| \leq \mathcal{O}\left(\epsilon\kappa(U)\right).$$

The noise level ϵ allowed is the same as SPA.

The error on U is decreased by a factor $\kappa(U)$ and matches the SDP preconditioning.

SPA-based Preconditioning: Robustness

Computational cost. Compute SPA twice in $O(mnr)$ + SVD of an r-by-r matrix in $O(r^3)$. This is extremely cheap, like SPA.

Theorem. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}\kappa^2(U)}\right)$, SPA satisfies

$$||U - M(:,\mathcal{K})|| \leq \mathcal{O}\left(\epsilon\kappa(U)\right).$$

The noise level ϵ allowed is the same as SPA.

The error on U is decreased by a factor $\kappa(U)$ and matches the SDP preconditioning.

Summary – Faster Preconditionings

1. **Pre-whitening**. Given the truncated SVD $M \approx U_r \Sigma_r V_r^T$, premultiply with U_r^T (noise filtering) and then Σ_r^{-1} (whitening) to keep V_r^T:
   - sensitive to important differences between abundances of endmembers,
   - provably behaves similarly as the SDP under a standard generative model [GM14].

2. **SPA-based Preconditioning**. Identify r columns of M using SPA: $\tilde{U} = M(:, \mathcal{K})$ and then premultiply M with $\tilde{U}^\dagger$ (left-inverse of $\tilde{U}$):
   - not sensitive to important differences between abundances of endmembers,
   - is provably more robust than SPA [GM14],
   - is extremely fast (essentially, cost of running SPA twice).

[GM14] G., Ma, *Enhancing Pure-Pixel Identification Performance via Preconditioning*, arXiv:1406.5286.

Summary – Faster Preconditionings

1. **Pre-whitening**. Given the truncated SVD $M \approx U_r \Sigma_r V_r^T$, premultiply with U_r^T (noise filtering) and then Σ_r^{-1} (whitening) to keep V_r^T:
   - sensitive to important differences between abundances of endmembers,
   - provably behaves similarly as the SDP under a standard generative model [GM14].

2. **SPA-based Preconditioning**. Identify r columns of M using SPA: $\tilde{U} = M(:, \mathcal{K})$ and then premultiply M with $\tilde{U}^\dagger$ (left-inverse of $\tilde{U}$):
   - not sensitive to important differences between abundances of endmembers,
   - is provably more robust than SPA [GM14],
   - is extremely fast (essentially, cost of running SPA twice).

[GM14] G., Ma, *Enhancing Pure-Pixel Identification Performance via Preconditioning*, arXiv:1406.5286.

Summary – Preconditionings

	Noise level (ϵ) $= \mathcal{O}\left(\frac{\sigma_{\min}(U)}{\sqrt{r}}\right) \cdot$	Error $= \mathcal{O}\left(\epsilon\kappa(U)\right) \cdot$	Computational Cost
SPA	$\kappa^{-2}(U)$	$\kappa(U)$	$\mathcal{O}(mnr)$
SDP-SPA	r^{-1}	1	$\Omega(mnr^2)$
PW-SPA	$n^{-3/2}$	$n^{3/2}$	$\mathcal{O}(mnr^2)$
PW-SPA + Model	β^{-1}	β	$\mathcal{O}(mnr^2)$
SPA-SPA	$\kappa^{-2}(U)$	1	$\mathcal{O}(mnr)$

Table : Robustness of Preconditionings

Two-by-Three Toy Example

Let $k \geq 0$ and

$$U = \begin{pmatrix} k+1 & k \\ k & k+1 \end{pmatrix}.$$

We have $\sigma_{\min}(U) = 1$ and $\sigma_{\max}(U) = 2k+1$ hence $\kappa(U) = 2k+1$. Let

$$V = \begin{pmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0.5 \end{pmatrix}, \; M = UV + N,$$

with $N = \delta \, [-UV(:,1), \; -UV(:,2), \; UV(:,3)]$ for some small $\delta > 0$.
Hence

$$M = (1-\delta) \begin{pmatrix} k+1 & k & \left(\frac{1+\delta}{1-\delta}\right)(k+\frac{1}{2}) \\ k & k+1 & \left(\frac{1+\delta}{1-\delta}\right)(k+\frac{1}{2}) \end{pmatrix}.$$

Two-by-Three Toy Example

Let $k \geq 0$ and

$$U = \begin{pmatrix} k+1 & k \\ k & k+1 \end{pmatrix}.$$

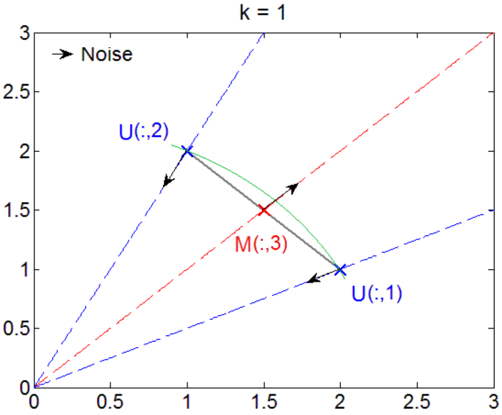
We have $\sigma_{\min}(U) = 1$ and $\sigma_{\max}(U) = 2k+1$ hence $\kappa(U) = 2k+1$. Let

$$V = \begin{pmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0.5 \end{pmatrix}, \; M = UV + N,$$

with $N = \delta\, [-UV(:,1), \; -UV(:,2), \; UV(:,3)]$ for some small $\delta > 0$.
Hence

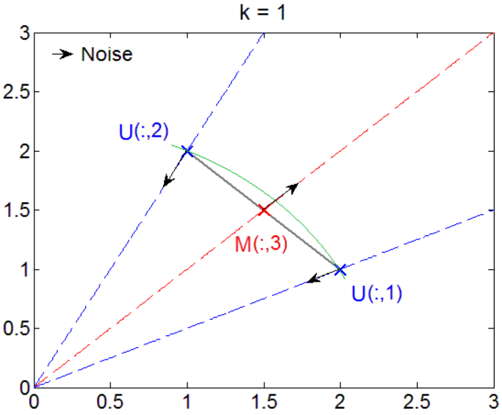
$$M = (1-\delta) \begin{pmatrix} k+1 & k & \left(\frac{1+\delta}{1-\delta}\right)(k+\frac{1}{2}) \\ k & k+1 & \left(\frac{1+\delta}{1-\delta}\right)(k+\frac{1}{2}) \end{pmatrix}.$$

Two-by-Three Toy Example



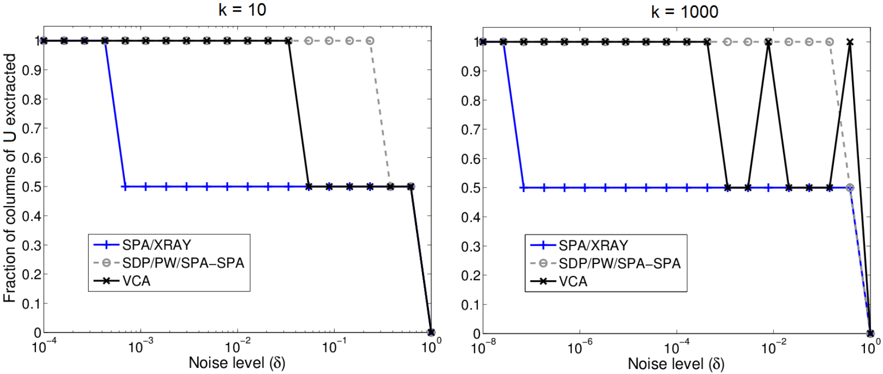
For $\delta \geq \frac{1}{8k^2}$, SPA fails: first extracted column is $M(:,3)$.

Two-by-Three Toy Example



For $\delta \geq \frac{1}{8k^2}$, SPA fails: first extracted column is $M(:,3)$.

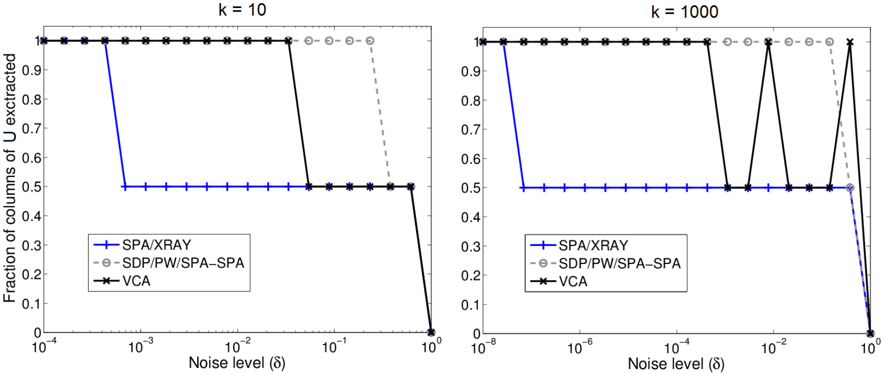
Two-by-Three Toy Example



For $\delta \geq \frac{1}{8k^2}$, SPA fails: first extracted column is $M(:,3)$.

For any k and $\delta \leq \mathcal{O}(1)$, all preconditioned variants succeed.

Two-by-Three Toy Example

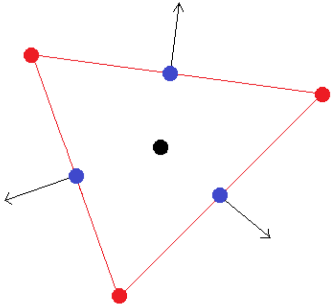


For $\delta \geq \frac{1}{8k^2}$, SPA fails: first extracted column is $M(:,3)$.

For any k and $\delta \leq \mathcal{O}(1)$, all preconditioned variants succeed.

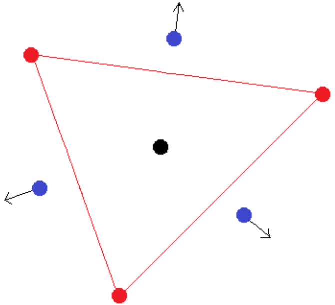
Synthetic data sets

- Each entry of $U \in \mathbb{R}_+^{40 \times 20}$ 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.



Synthetic data sets

- Each entry of $U \in \mathbb{R}_{+}^{40 \times 20}$ 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.



Results for the synthetic data sets

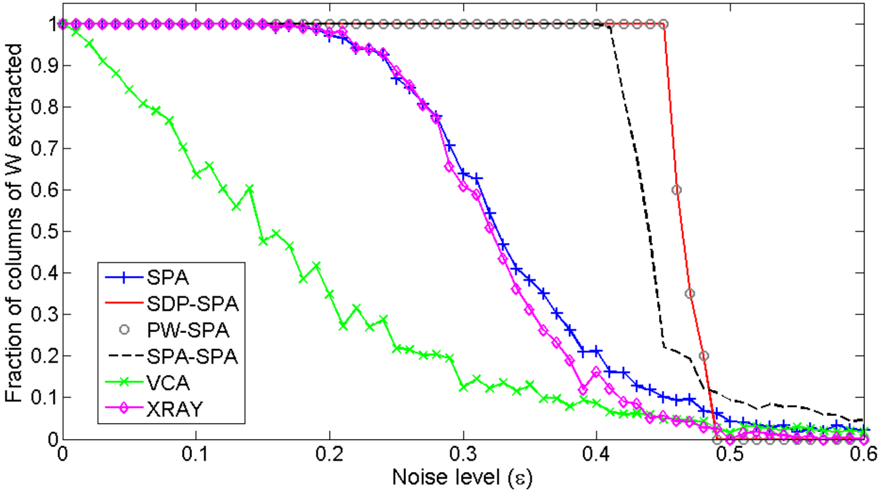


Figure : Average of the fraction of columns correctly extracted depending on the noise level (for each noise level, 25 matrices are generated).

Hubble telescope hyperspectral image

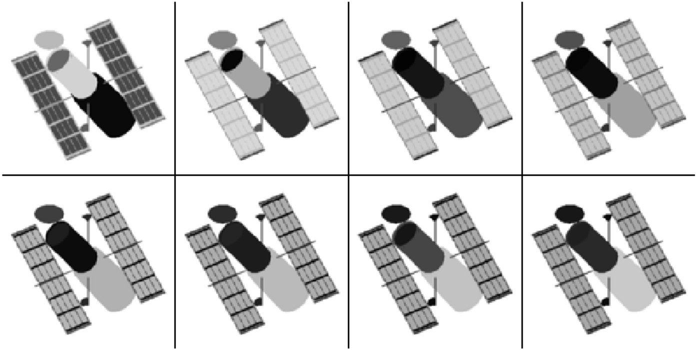


Figure : Sample of images for the Hubble telescope hyperspectral image with 100 spectral bands and 128×128 pixels.

Hubble telescope hyperspectral image

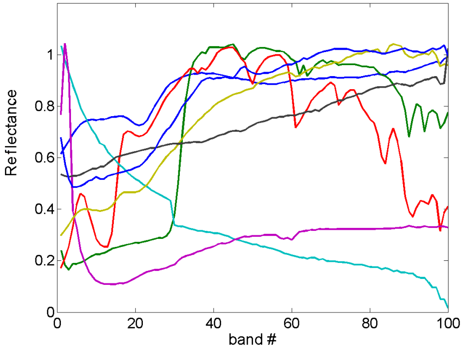


Figure : Spectral signatures extracted by SPA, corresponding to constitutive materials (matrix U with $\kappa(U) = 115$).

Hubble telescope hyperspectral image

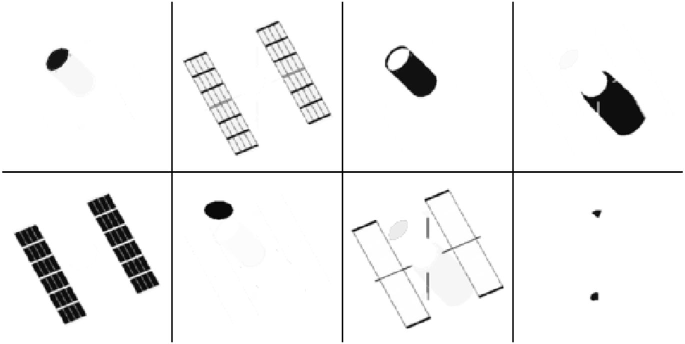


Figure : Reconstructed abundance maps (matrix V).

Hubble telescope with blur and noise

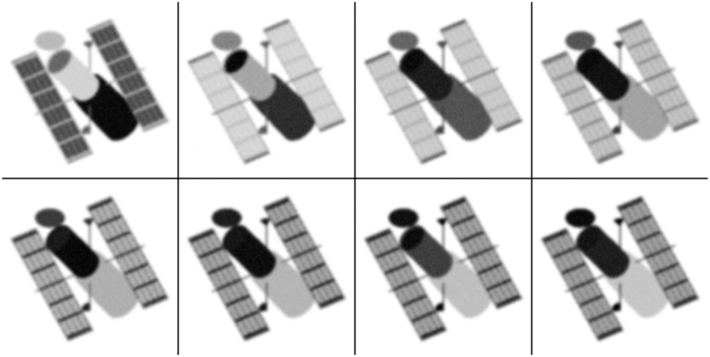


Figure : Sample of images for the Hubble telescope.

Hubble telescope with blur and noise

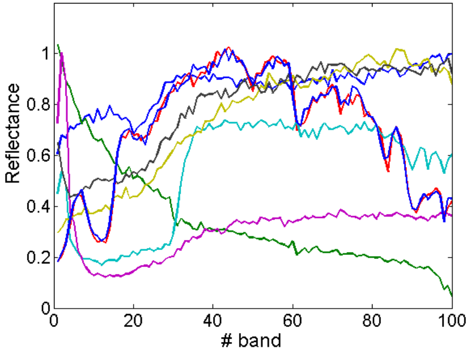


Figure : Spectral signatures extracted by SPA, corresponding to constitutive materials (matrix U).

Hubble telescope with blur and noise

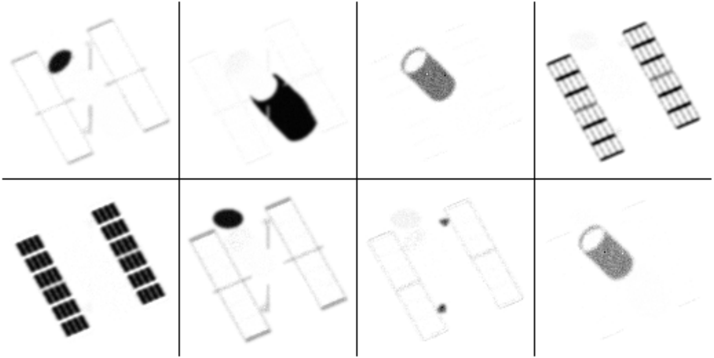


Figure : Reconstructed abundance maps (matrix V). With the blur and noise, SPA fails to identify good columns (two correspond to the same material).

Hubble telescope with blur and noise

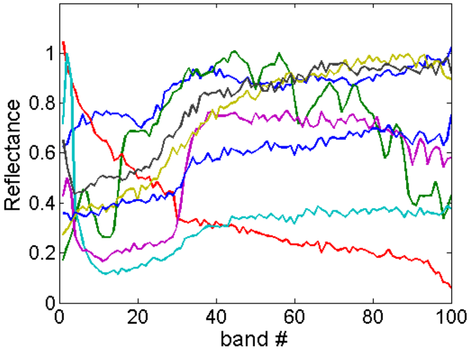


Figure : Spectral signatures extracted by preconditioned SPA, corresponding to constitutive materials (matrix U).

Hubble telescope with blur and noise

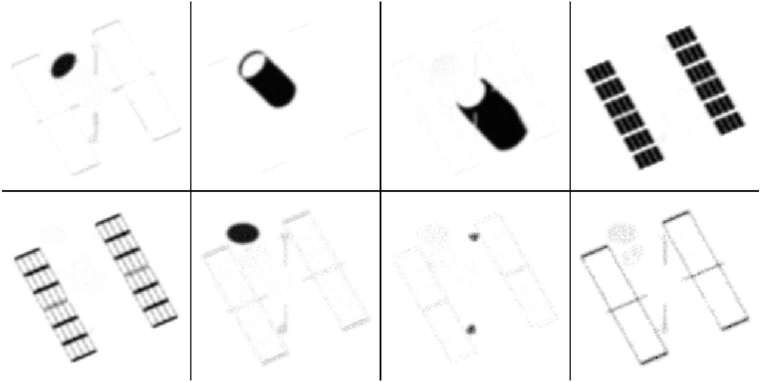


Figure : Reconstructed abundance maps (matrix V). With the blur and noise, preconditioned SPA is able to identify the right columns.

Hubble telescope with blur and noise

	SPA	SDP-SPA	PW-SPA	SPA-SPA
Hon. side	6.51	6.94	6.94	**6.15**
Cop. Strip.	26.83	7.46	**7.44**	**7.44**
Green glue	2.09	**2.03**	**2.03**	**2.03**
Aluminum	**1.71**	1.80	1.80	1.80
Solar cell	**4.96**	5.48	5.48	**4.96**
Hon. top	2.34	**2.30**	**2.30**	**2.30**
Black edge	27.09	13.16	13.16	**13.13**
Bolts	**2.65**	**2.65**	**2.65**	2.70
Average	9.27	5.23	5.23	**5.06**
Time (s.)	0.05	4.74	2.18	0.37

Table : MRSA of the identified endmembers with the true endmembers, and running time in seconds of the different preconditioned SPA algorithms.

Conclusion

1. Blind hyperspectral unmixing
   - Challenging but important problem
   - Under the pure-pixel assumption, it is tractable even in the presence of noise
2. Pure-Pixel Identification
   - SPA is simple, fast and robust to noise
   - Its robustness can be significantly improved using preconditioning
     - SDP: ideal preconditioning but computationally rather expensive
     - Pre-whitening: faster but sensitive to different proportions of endmembers.
     - SPA-based: extremely fast, provably more robust than SPA
3. Future Work
   - Apply and analyze the effect of preconditionings on other algorithms, and on real-world data sets.

Conclusion

1. Blind hyperspectral unmixing
   - Challenging but important problem
   - Under the pure-pixel assumption, it is tractable even in the presence of noise
2. Pure-Pixel Identification
   - SPA is simple, fast and robust to noise
   - Its robustness can be significantly improved using preconditioning
     - SDP: ideal preconditioning but computationally rather expensive
     - Pre-whitening: faster but sensitive to different proportions of endmembers.
     - SPA-based: extremely fast, provably more robust than SPA.
3. Future Work
   - Apply and analyze the effect of preconditionings on other algorithms, and on real-world data sets.

Conclusion

1. Blind hyperspectral unmixing
   - Challenging but important problem
   - Under the pure-pixel assumption, it is tractable even in the presence of noise
2. Pure-Pixel Identification
   - SPA is simple, fast and robust to noise
   - Its robustness can be significantly improved using preconditioning
     - SDP: ideal preconditioning but computationally rather expensive
     - Pre-whitening: faster but sensitive to different proportions of endmembers.
     - SPA-based: extremely fast, provably more robust than SPA.
3. Future Work
   - Apply and analyze the effect of preconditionings on other algorithms, and on real-world data sets.

Thank you for your attention!

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