Robust Near-Separable Nonnegative Matrix Factorization Using Linear Optimization

Nicolas Gillis

nicolas.gillis@uclouvain.be
https://sites.google.com/site/nicolasgillis/

ICTEAM Institute, Université catholique de Louvain

Joint work with Robert Luce (T.U. Berlin)

ROKS 2013

July 10, 2013

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?

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

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

→ **Many applications.** Taxt 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?

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

→ **Many applications.** Taxt 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 :

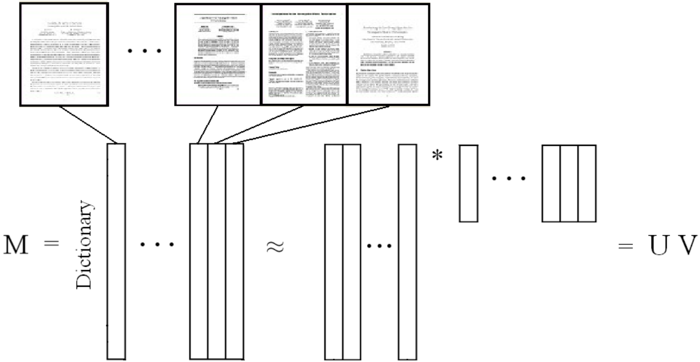
$$\underbrace{M(:,i)}_{\geq 0} \approx \sum_{k=1}^{r} \underbrace{U(:,k)}_{\geq 0} \underbrace{V(k,i)}_{\geq 0} \qquad \text{for all } i.$$

Why nonnegativity?

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

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

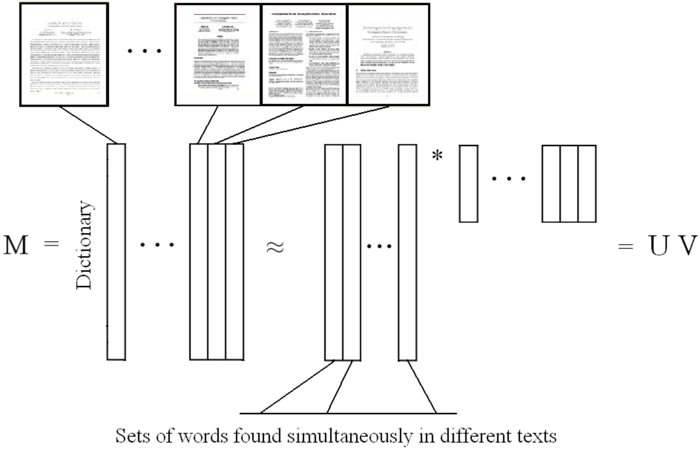
Text Mining



◇ Basis elements allow to recover the different topics;

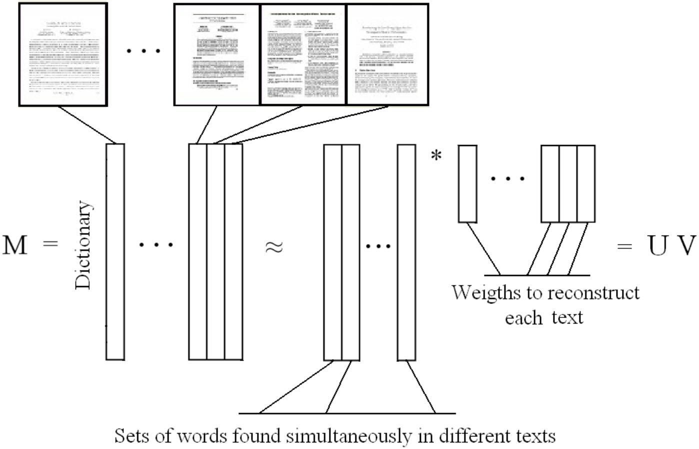
◇ Weights allow to assign each text to its corresponding topics.

Text Mining



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

Text Mining



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

Hyperspectral Unmixing

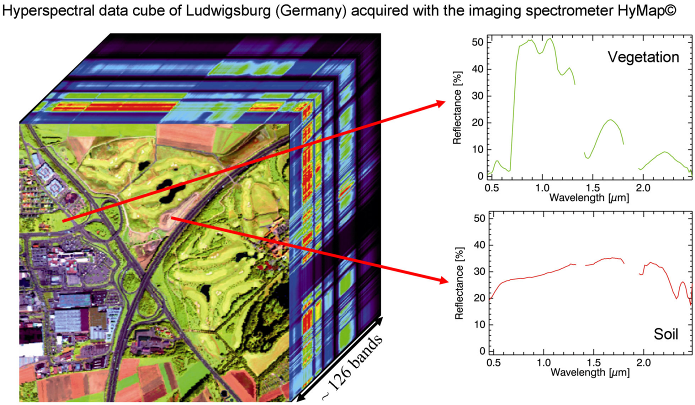


Figure: Hyperspectral image.

Goal. Recover the endmembers and their abundances.

Model. Linear mixing model.

Hyperspectral Unmixing

Hyperspectral data cube of Ludwigsburg (Germany) acquired with the imaging spectrometer HyMap©

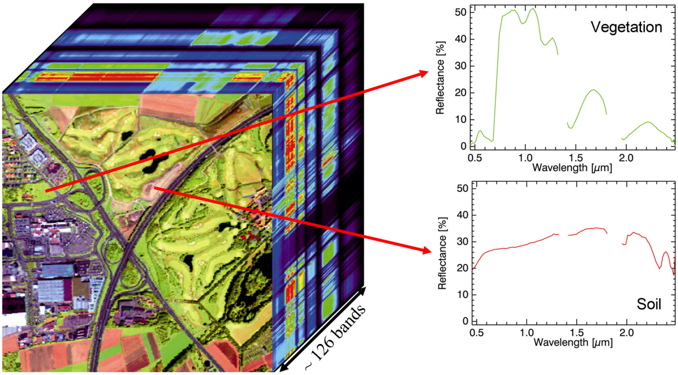
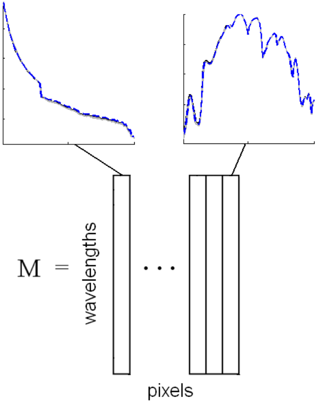


Figure: Hyperspectral image.

Goal. Recover the endmembers and their abundances.

Model. Linear mixing model.

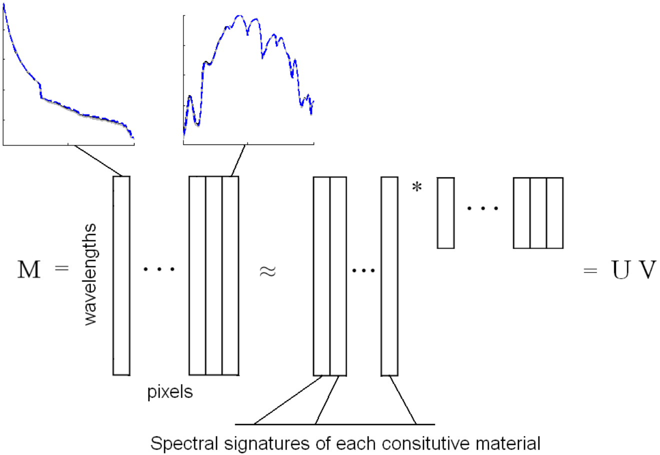
Hyperspectral Unmixing



◇ Basis elements allow to recover the different materials;

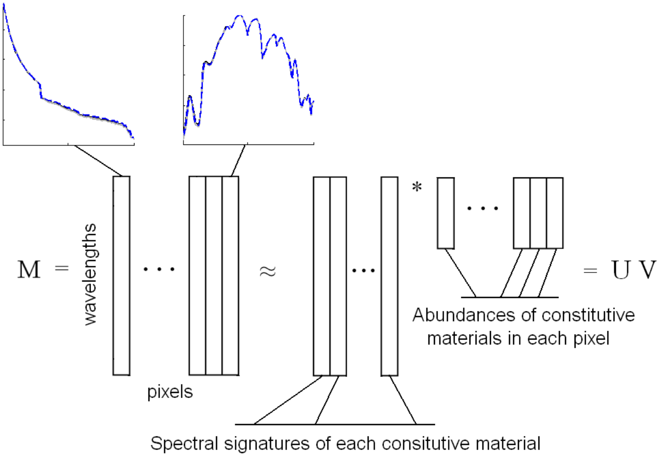
◇ Weights allow to know which pixel contains which material.

Hyperspectral Unmixing



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

Hyperspectral Unmixing



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

Hyperspectral Unmixing

Figure: Urban dataset.

Hyperspectral Unmixing

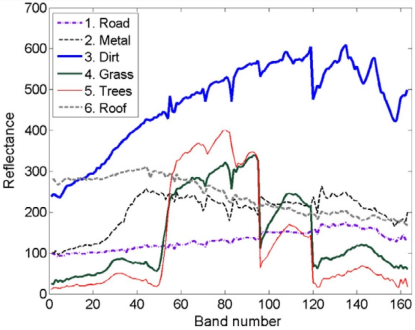
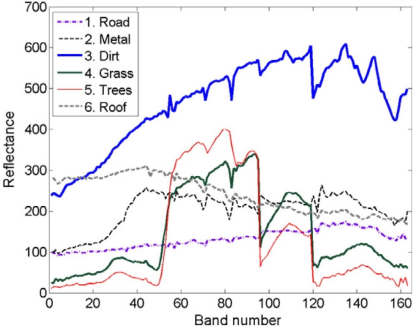


Figure: Urban dataset.

Hyperspectral Unmixing



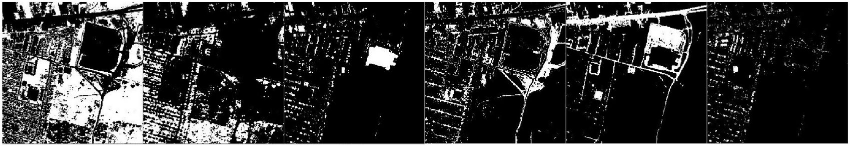


Figure: Urban dataset.

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.
- 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 generative 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

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

- NMF is NP-hard [V09], and highly ill-posed.
- 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 generative 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?

- **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.

- **Text mining**: for each topic, there is a 'pure' document on that topic, or, for each topic, there is a 'pure' word (an anchor 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.

Is separability a reasonable assumption?

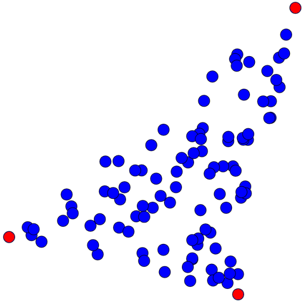
- **Hyperspectral unmixing**: separability is particularly natural: for each constitutive material, there is a 'pure' pixel containing only that material. This is the so called pure-pixel assumption which is widely used in hyperspectral imaging.

- **Text mining**: for each topic, there is a 'pure' document on that topic, or, for each topic, there is a 'pure' word (an anchor 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.

Geometric Interpretation of Separability

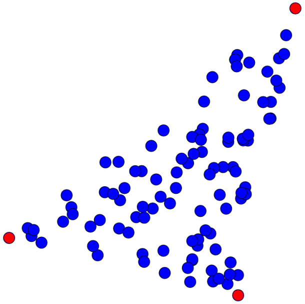
After normalization, the columns of M, U and V sum to one: the columns of U are the vertices of the convex hull of the columns of M.



Separable NMF

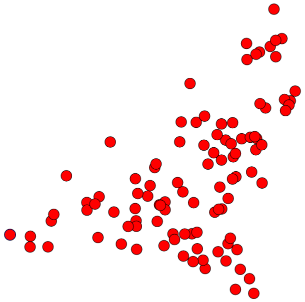
$$M \text{ is } r\text{-separable} \iff M = U[I_r, V']\Pi,$$

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



Near-Separable NMF

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



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:

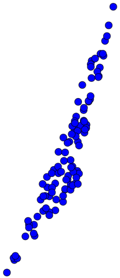
Parameter α = minimum distance of a vertex to the convex hull of other vertices.

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:



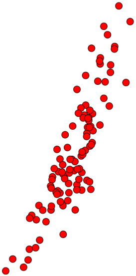
Parameter α = minimum distance of a vertex to the convex hull of other vertices.

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:



Parameter α = minimum distance of a vertex to the convex hull of other vertices.

Hottopixx, a Linear Optimization Model

For a normalized separable matrix M, we have, up to permutation,

$$M = [U, UV'] = M \underbrace{\begin{pmatrix} I_r & V' \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix}}_{X^0 \in \mathbb{R}^{n\times n}} = MX^0.$$

where $V' \leq \mathbf{1}_{r\times(n-r)}$. [BRRT12] proposed the following model:

$$\begin{aligned} \min_{X\in\mathbb{R}^{n\times n}} \quad & p^T \operatorname{diag}(X) \\ \text{such that} \quad & ||\tilde{M} - \tilde{M}X||_1 \leq 2\epsilon, \\ & \operatorname{tr}(X) = r, \\ & 0 \leq X_{ij} \leq X_{ii} \leq 1 \text{ for all } i, j. \end{aligned}$$

where the entries of p are distinct.

[BRRT12] Bittorf, Recht, Ré, Tropp, *Factoring nonnegative matrices with linear programs*, NIPS 2012.

Hottopixx, a Linear Optimization Model

For a normalized separable matrix M, we have, up to permutation,

$$M = [U, UV'] = M \underbrace{\begin{pmatrix} I_r & V' \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix}}_{X^0 \in \mathbb{R}^{n\times n}} = MX^0.$$

where $V' \leq \mathbf{1}_{r\times(n-r)}$. [BRRT12] proposed the following model:

$$\begin{aligned} \min_{X\in\mathbb{R}^{n\times n}} \quad & p^T \operatorname{diag}(X) \\ \text{such that} \quad & \|\tilde{M} - \tilde{M}X\|_1 \leq 2\epsilon, \\ & \operatorname{tr}(X) = r, \\ & 0 \leq X_{ij} \leq X_{ii} \leq 1 \text{ for all } i,j. \end{aligned}$$

where the entries of p are distinct.

[BRRT12] Bittorf, Recht, Ré, Tropp, *Factoring nonnegative matrices with linear programs*, NIPS 2012.

Hottopixx, a Linear Optimization Model

For a normalized separable matrix M, we have, up to permutation,

$$M = [U, UV'] = M \underbrace{\begin{pmatrix} I_r & V' \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix}}_{X^0 \in \mathbb{R}^{n\times n}} = MX^0.$$

where $V' \leq \mathbf{1}_{r\times(n-r)}$. [BRRT12] proposed the following model:

$$\begin{aligned} \min_{X\in\mathbb{R}^{n\times n}} \quad & p^T \operatorname{diag}(X) \\ \text{such that} \quad & \|\tilde{M} - \tilde{M}X\|_1 \leq 2\epsilon, \\ & \operatorname{tr}(X) = r, \\ & 0 \leq X_{ij} \leq X_{ii} \leq 1 \text{ for all } i, j. \end{aligned}$$

where the entries of p are distinct.

[BRRT12] Bittorf, Recht, Ré, Tropp, *Factoring nonnegative matrices with linear programs*, NIPS 2012.

Hottopixx, a Linear Optimization Model

For a normalized separable matrix M, we have, up to permutation,

$$M = [U, UV'] = M \underbrace{\begin{pmatrix} I_r & V' \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix}}_{X^0 \in \mathbb{R}^{n\times n}} = MX^0.$$

where $V' \leq \mathbf{1}_{r\times(n-r)}$. [BRRT12] proposed the following model:

$$\begin{aligned} \min_{X\in\mathbb{R}^{n\times n}} \quad & p^T \operatorname{diag}(X) \\ \text{such that} \quad & \|\tilde{M} - \tilde{M}X\|_1 \leq 2\epsilon, \\ & \operatorname{tr}(X) = r, \\ & 0 \leq X_{ij} \leq X_{ii} \leq 1 \text{ for all } i, j. \end{aligned}$$

where the entries of p are distinct.

[BRRT12] Bittorf, Recht, Ré, Tropp, *Factoring nonnegative matrices with linear programs*, NIPS 2012.

Hottopixx, a Linear Optimization Model

For a normalized separable matrix M, we have, up to permutation,

$$M = [U, UV'] = M \underbrace{\begin{pmatrix} I_r & V' \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix}}_{X^0 \in \mathbb{R}^{n\times n}} = MX^0.$$

where $V' \leq \mathbf{1}_{r\times(n-r)}$. [BRRT12] proposed the following model:

$$\begin{aligned} \min_{X \in \mathbb{R}^{n\times n}} \quad & p^T \operatorname{diag}(X) \\ \text{such that} \quad & \|\tilde{M} - \tilde{M}X\|_1 \leq 2\epsilon, \\ & \operatorname{tr}(X) = r, \\ & 0 \leq X_{ij} \leq X_{ii} \leq 1 \text{ for all } i, j. \end{aligned}$$

where the entries of p are distinct.

[BRRT12] Bittorf, Recht, Ré, Tropp, *Factoring nonnegative matrices with linear programs*, NIPS 2012.

Hottopixx, a Linear Optimization Model

$$\min_{X \in \mathbb{R}^{n \times n}} \quad p^T \operatorname{diag}(X)$$

$$\text{such that} \quad ||\tilde{M} - \tilde{M}X||_1 \leq 2\epsilon,$$

$$\operatorname{tr}(X) = r,$$

$$0 \leq X_{ij} \leq X_{ii} \leq 1 \text{ for all } i, j.$$

Theorem ([G12]). If $\epsilon \leq \mathcal{O}\left(\frac{\alpha^2}{r}\right)$, their algorithm leads to an NMF (W, H) s.t.

$$||\tilde{M} - UV||_1 \leq \mathcal{O}\left(\frac{r\epsilon}{\alpha}\right).$$

Drawbacks. Requires to solve a LP in $\mathcal{O}(n^2)$ variables, the parameters ϵ and r have to be estimated, not very robust in practice, normalization is necessary.

[G12] G., *Robustness Analysis of Hottopixx, a Linear Programming Model for Factoring Nonnegative Matrices*, SIMAX, 2013.

Hottopixx, a Linear Optimization Model

$$\begin{aligned} \min_{X \in \mathbb{R}^{n \times n}} \quad & p^T \operatorname{diag}(X) \\ \text{such that} \quad & ||\tilde{M} - \tilde{M}X||_1 \leq 2\epsilon, \\ & \operatorname{tr}(X) = r, \\ & 0 \leq X_{ij} \leq X_{ii} \leq 1 \text{ for all } i, j. \end{aligned}$$

Theorem ([G12]). If $\epsilon \leq \mathcal{O}\left(\frac{\alpha^2}{r}\right)$, their algorithm leads to an NMF (W, H) s.t.

$$||\tilde{M} - UV||_1 \leq \mathcal{O}\left(\frac{r\epsilon}{\alpha}\right).$$

Drawbacks. Requires to solve a LP in $\mathcal{O}(n^2)$ variables, the parameters ϵ **and** r have to be estimated, not very robust in practice, normalization is necessary.

[G12] G., *Robustness Analysis of Hottopixx, a Linear Programming Model for Factoring Nonnegative Matrices*, SIMAX, 2013.

An Improved Linear Optimization Model

$$\min_{X \in \mathbb{R}_+^{n \times n}} \quad p^T \operatorname{diag}(X)$$

such that

$$||\tilde{M} - \tilde{M}X||_1 \leq 2\epsilon,$$
$$\operatorname{tr}(X) = r,$$
$$X_{ij} \leq X_{ii} \text{ for all } i, j,$$
$$X_{ii} \leq 1 \text{ for all } i,$$

where p is a vector with positive entries.

The new model detects the factorization rank r automatically.

Same robustness analysis as Hottopixx applies for any $\rho > 0$.

Does not require column normalization.

If the columns of U are isolated: $\epsilon \leq \mathcal{O}(\alpha) \Rightarrow ||\tilde{M} - UV||_1 \leq \mathcal{O}(\epsilon)$, which is provably more robust than Hottopixx for which $\epsilon \leq \mathcal{O}\left(\frac{\alpha}{r}\right)$.

[GL13] G., Luce, *Robust Near-Separable Nonnegative Matrix Factorization Using Linear Optimization*, February 2013.

An Improved Linear Optimization Model

$$\begin{aligned} \min_{X \in \mathbb{R}^{n \times n}_+} \quad & p^T \operatorname{diag}(X) \\ \text{such that} \quad & ||\tilde{M} - \tilde{M}X||_1 \leq 2\epsilon, \\ & \text{tr}(X) = r, \\ & X_{ij} \leq X_{ii} \text{ for all } i, j, \\ & X_{ii} \leq 1 \text{ for all } i, \end{aligned}$$

where p is a vector with positive entries.

The new model detects the factorization rank r automatically.

Same robustness analysis as Hottopixx applies for any $\rho > 0$.

Does not require column normalization.

If the columns of U are isolated: $\epsilon \leq \mathcal{O}(\alpha) \Rightarrow ||\tilde{M} - UV||_1 \leq \mathcal{O}(\epsilon)$, which is provably more robust than Hottopixx for which $\epsilon \leq \mathcal{O}\left(\frac{\alpha}{r}\right)$.

[GL13] G., Luce, *Robust Near-Separable Nonnegative Matrix Factorization Using Linear Optimization*, February 2013.

An Improved Linear Optimization Model

$$\min_{X \in \mathbb{R}^{n \times n}_+} \quad p^T \operatorname{diag}(X)$$

such that $||\tilde{M} - \tilde{M}X||_1 \leq \rho\epsilon,$

~~$\operatorname{tr}(X) = r$~~,

$X_{ij} \leq X_{ii}$ for all i, j,

$X_{ii} \leq 1$ for all i,

where p is a vector with positive entries.

The new model detects the factorization rank r automatically.

Same robustness analysis as Hottopixx applies for any $\rho > 0$.

Does not require column normalization.

If the columns of U are isolated: $\epsilon \leq \mathcal{O}(\alpha) \Rightarrow ||\tilde{M} - UV||_1 \leq \mathcal{O}(\epsilon)$, which is provably more robust than Hottopixx for which $\epsilon \leq \mathcal{O}\left(\frac{\alpha}{r}\right)$.

[GL13] G., Luce, *Robust Near-Separable Nonnegative Matrix Factorization Using Linear Optimization*, February 2013.

An Improved Linear Optimization Model

$$\min_{X \in \mathbb{R}^{n \times n}_{+}} \quad p^T \operatorname{diag}(X)$$

such that $\quad ||\tilde{M} - \tilde{M}X||_1 \leq \rho\epsilon,$

~~$\operatorname{tr}(X) = r$~~,

$||\tilde{M}(:,i)||_1 X_{ij} \leq ||\tilde{M}(:,j)||_1 X_{ii}$ for all i, j,

$X_{ii} \leq 1$ for all i,

where p is a vector with positive entries.

The new model detects the factorization rank r automatically.

Same robustness analysis as Hottopixx applies for any $\rho > 0$.

Does not require column normalization.

If the columns of U are isolated: $\epsilon \leq \mathcal{O}(\alpha) \Rightarrow ||\tilde{M} - UV||_1 \leq \mathcal{O}(\epsilon)$, which is provably more robust than Hottopixx for which $\epsilon \leq \mathcal{O}\left(\frac{\alpha}{r}\right)$.

[GL13] G., Luce, *Robust Near-Separable Nonnegative Matrix Factorization Using Linear Optimization*, February 2013.

An Improved Linear Optimization Model

$$\min_{X \in \mathbb{R}^{n \times n}_+} \quad p^T \operatorname{diag}(X)$$

such that $||\tilde{M} - \tilde{M}X||_1 \leq \rho\epsilon$,

~~$\operatorname{tr}(X) = r$~~,

$||\tilde{M}(:,i)||_1 X_{ij} \leq ||\tilde{M}(:,j)||_1 X_{ii}$ for all i, j,

$X_{ii} \leq 1$ for all i,

where p is a vector with positive entries.

The new model detects the factorization rank r automatically.

Same robustness analysis as Hottopixx applies for any $\rho > 0$.

Does not require column normalization.

If the columns of U are isolated: $\epsilon \leq \mathcal{O}(\alpha) \Rightarrow ||\tilde{M} - UV||_1 \leq \mathcal{O}(\epsilon)$, which is provably more robust than Hottopixx for which $\epsilon \leq \mathcal{O}\left(\frac{\alpha}{r}\right)$.

[GL13] G., Luce, *Robust Near-Separable Nonnegative Matrix Factorization Using Linear Optimization*, February 2013.

Numerical Experiments

- ◇ Each entry of $U \in \mathbb{R}_+^{50\times 10}$ uniform in $[0, 1]$; each column normalized.
- ◇ Each of the 90 columns of $V' \in \mathbb{R}_+^{10}$, Dirichlet.

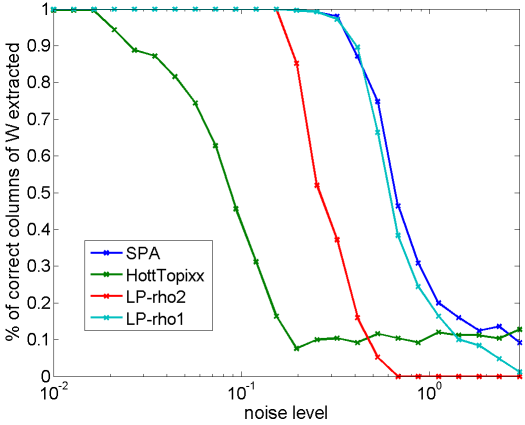


Figure: Noise is Gaussian.

Numerical Experiments

- ◇ Each entry of $U \in \mathbb{R}_+^{50\times 10}$ uniform in $[0,1]$; each column normalized.
- ◇ Each of the 90 columns of $V' \in \mathbb{R}_+^{10}$, Dirichlet.

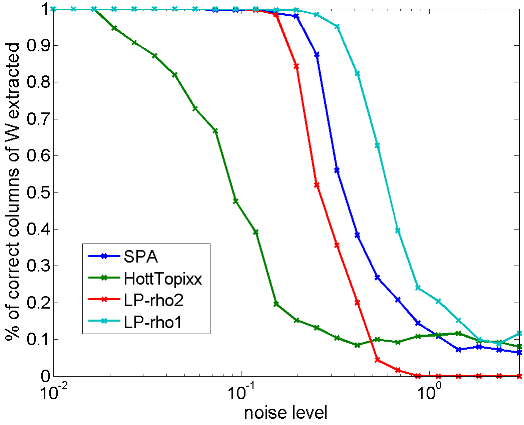


Figure: Noise is sparse (75%), non-zero entries are Gaussian.

Numerical Experiments

- Each entry of $U \in \mathbb{R}_+^{50\times 10}$ uniform in $[0,1]$; each column normalized.
- Each of the 90 columns of $V' \in \mathbb{R}_+^{10}$, Dirichlet.

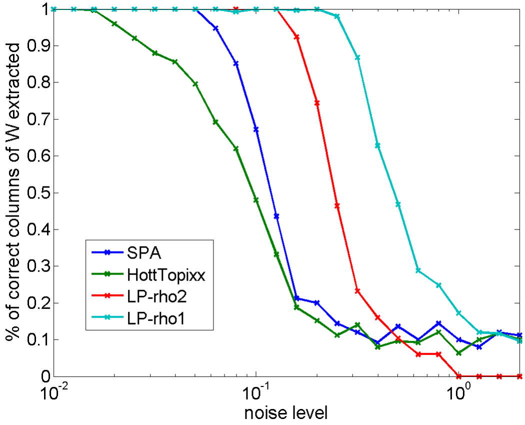


Figure: Noise is very sparse: one non-zero entry per column.

Conclusion

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

2. Separable NMF
   - Separability makes NMF problems efficiently solvable
   - Need for fast, practical and robust algorithms

3. A new LP model for near-separable NMF
   - More robust, more flexible, always feasible, no normalization
   - but ... computationally expensive.
     (Possible fix: preselect a 'good' subset of columns.)

Conclusion

1. Nonnegative matrix factorization (NMF)
   - Easily interpretable linear dimensionality reduction technique for nonnegative data, with *many* applications
2. Separable NMF
   - Separability makes NMF problems efficiently solvable
   - Need for fast, practical and robust algorithms
3. A new LP model for near-separable NMF
   - More robust, more flexible, always feasible, no normalization
   - but . . . computationally expensive.
     (Possible fix: preselect a 'good' subset of columns.)

Conclusion

1. Nonnegative matrix factorization (NMF)
    - Easily interpretable linear dimensionality reduction technique for nonnegative data, with *many* applications
2. Separable NMF
    - Separability makes NMF problems efficiently solvable
    - Need for fast, practical and robust algorithms
3. A new LP model for near-separable NMF
    - More robust, more flexible, always feasible, no normalization
    - but . . . computationally expensive.
      (Possible fix: preselect a 'good' subset of columns.)

Reference.

G., R. Luce, *Robust Near-Separable Nonnegative Matrix Factorization Using Linear Optimization*, `arXiv:1302.4385`.

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

Thank you for your attention!