Factorisation en matrices (semi-)non-négatives : problématique, applications et tendances

Matthieu PUIGT

Travail collaboratif incluant au LISIC

G. ROUSSEL G. DELMAIRE M. OMIDVAR
A. LIMEM C. DORFFER R. CHREIKY F. YAHAYA

Univ. Littoral Côte d'Opale, EA 4491 – LISIC, F-62228 Calais, France

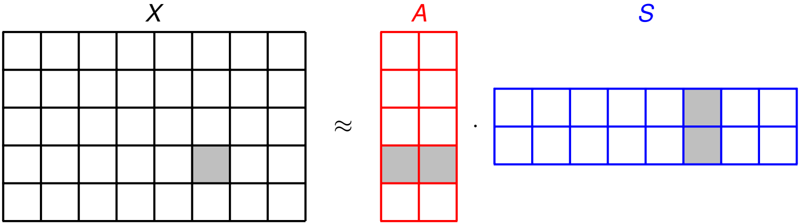
firstname.LASTNAME@univ-littoral.fr

http://www-lisic.univ-littoral.fr/~puigt/

18 juin 2018

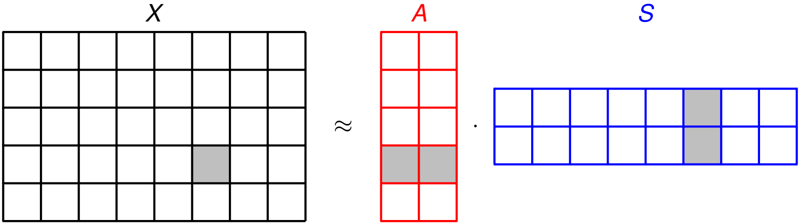
Foreword: matrix factorization

Many problems in *machine learning* and in *signal/image processing* can be rewritten as a system of equations $X \approx A \cdot S$:



Foreword: matrix factorization

Many problems in *machine learning* and in *signal/image processing* can be rewritten as a system of equations $X \approx A \cdot S$:



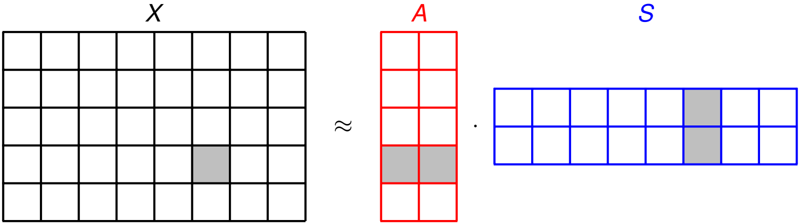
source localization/separation

X observed data matrix, A mixing matrix, S source matrix

- If A estimated (and sensor array geometry is known)
- ➪ mixture estimation (source localization)
- If S estimated (and sensor array geometry is known)
- ➪ source separation (beamforming)

Foreword: matrix factorization

Many problems in *machine learning* and in *signal/image processing* can be rewritten as a system of equations $X \approx A \cdot S$:



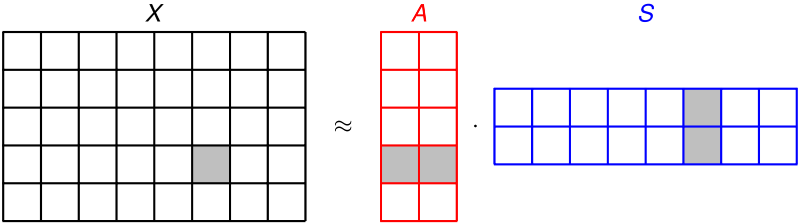
source localization/separation

X observed data matrix, A mixing matrix, S source matrix

- Is source separation a Signal Processing or an AI topic? Ask Google (Google AI blog submitted to SIGGRAPH'18).

Foreword: matrix factorization

Many problems in *machine learning* and in *signal/image processing* can be rewritten as a system of equations $X \approx A \cdot S$:



Low-rank matrix approximation

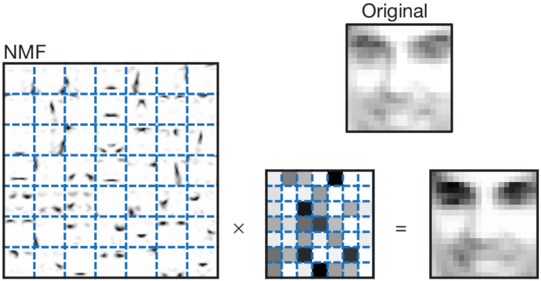
X noisy matrix (possibly with missing entries), A weight matrix, S latent variable matrix

- Topic modeling (e.g., Google News)
- Collaborative filtering (e.g., Netflix Prize)
- Graph analysis (after a transformation into a matrix – e.g., yellow cab / velib network)
- Mobile sensor calibration (Ph.D. theses of Clément Dorffer & Farouk Yahaya)

NMF: why is it so popular?

- In many problems, matrices $X \approx A \cdot S$ are non-negative
  - chemical source separation
  - hyperspectral imagery
  - mobile sensor calibration

➩ Non-negativity on A and/or S yield **better interpretability**

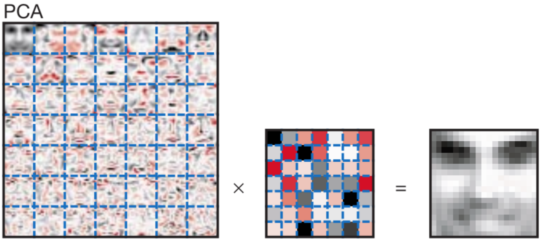


NMF applied to face dataset (source: Lee & Seung, 1999)

NMF: why is it so popular?

- In many problems, matrices $X \approx A \cdot S$ are non-negative
  - chemical source separation
  - hyperspectral imagery
  - mobile sensor calibration

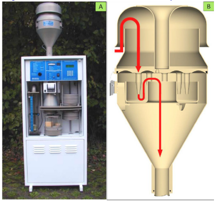
➪ Non-negativity on A and/or S yield **better interpretability**



Principal Component Analysis applied to face dataset (source: Lee & Seung, 1999)

Application (1)

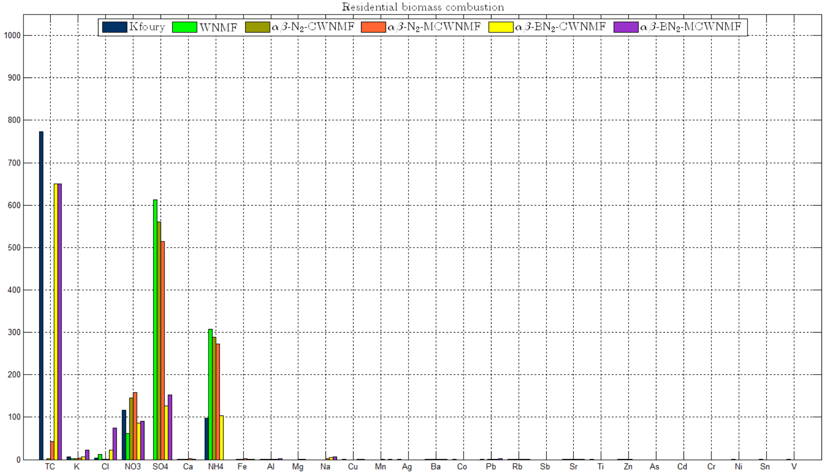
Separation of pollutant sources (Limem *et al.*, 2012–2014, Chreiky *et al.*, 2015–2017)



- Collaborative work with UCEIV (D. Courcot, F. Ledoux)
- Observation of air-suspended fine particulate matter over time
- Particles are analyzed by chemists who provide chemical concentrations of each species (e.g., iron, zinc, calcium, sodium, etc.) in ng/m^3
- ➪ Information stored in $X \simeq A \cdot S$
- Observed data are mixtures of "chemical profiles" (=chemical signatures which are specific to each emission source – factory chimney, exhaust, marine particles, etc – lines of S)
- We proposed **Informed** NMF : using partial expert's (in S) and meteorological knowledge (informing A) to improve the separation enhancement
- Application on Dunkirk air quality

Application (1)

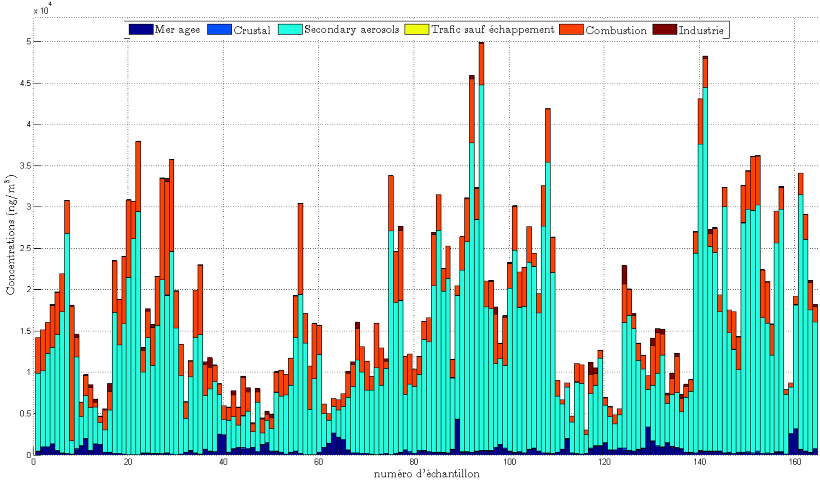
Separation of pollutant sources (Limem *et al.*, 2012–2014, Chreiky *et al.*, 2015–2017)



On estimated source profile (from S)

Application (1)

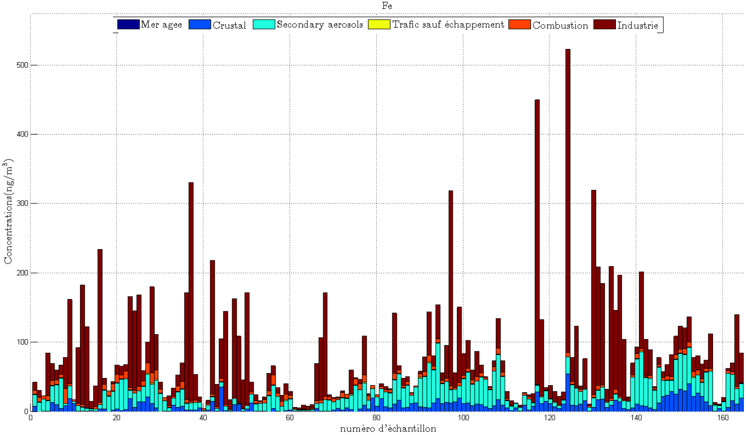
Separation of pollutant sources (Limem *et al.*, 2012–2014, Chreiky *et al.*, 2015–2017)



Global origin of sensed PM with respect to time (*A*)

Application (1)

Separation of pollutant sources (Limem *et al.*, 2012–2014, Chreiky *et al.*, 2015–2017)



Contribution of iron source in PM (from *A*)

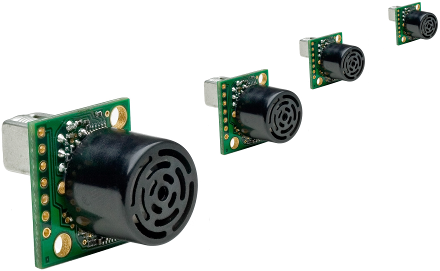
Application (2)

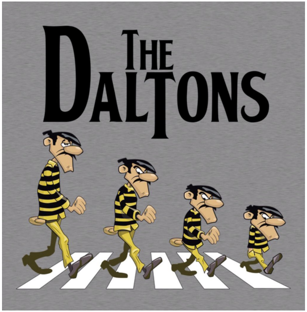
Mobile sensor calibration (Dorffer *et al.*, 2015–2018)

- Sensed phenomenon $\Longrightarrow$ voltage
- Voltage $\Longrightarrow$ phenomenon?

Application (2)

Mobile sensor calibration (Dorffer *et al.*, 2015–2018)

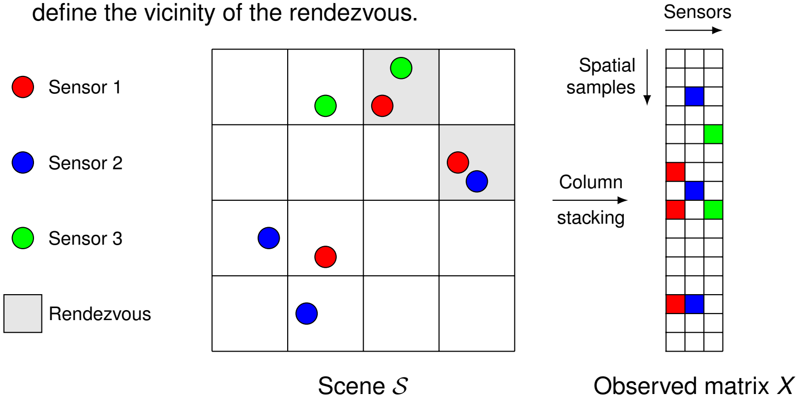




- Sensed phenomenon $\Longrightarrow$ voltage
- Voltage $\Longrightarrow$ phenomenon?
  - Sensor calibration needed
  - Not always physically possible
  - ➯ **Blind sensor calibration**

Application (2)
Definitions

- A rendezvous is a temporal and spatial vicinity between two sensors (Saukh *et al.*, 2013).
- A scene $\mathcal{S}$ is a discretized area observed during a time interval $[t, t+\Delta t)$. A spatial pixel has a size lower than Δd, where Δt and Δd define the vicinity of the rendezvous.



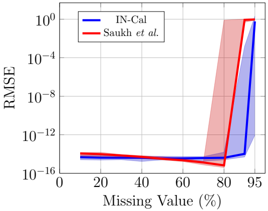
Blind calibration revisited as a weighted NMF problem (affine model)

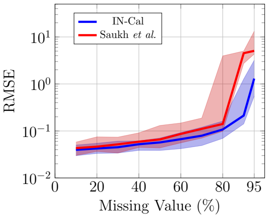
$$W \circ \underbrace{\begin{bmatrix} x(1,1) & \cdots & x(1,m-1) & y(1) \\ \vdots & & \vdots & \vdots \\ x(n,1) & \cdots & x(n,m-1) & y(n) \end{bmatrix}}_{X} \simeq W \circ \left(\underbrace{\begin{bmatrix} y(1) & 1 \\ \vdots & \vdots \\ y(n) & 1 \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{m-1} & 1 \\ \beta_1 & \beta_2 & \cdots & \beta_{m-1} & 0 \end{bmatrix}}_{S} \right)$$

We proposed methods:

- adding information on A (sparse assumptions) and S (sensor information)
- handling more complex calibration models (e.g., nonlinear)

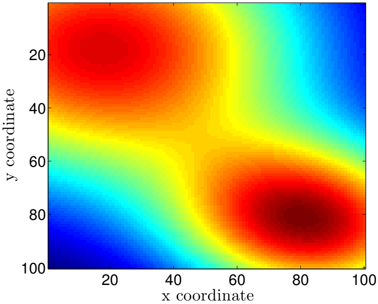
Estimation of S



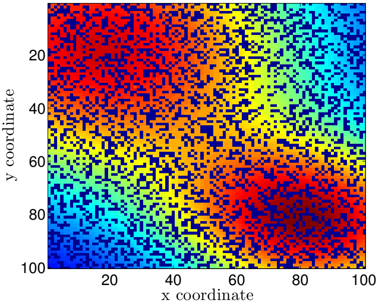


Estimation of *A*

Theoretical scene



First informed NMF (IN-Cal)

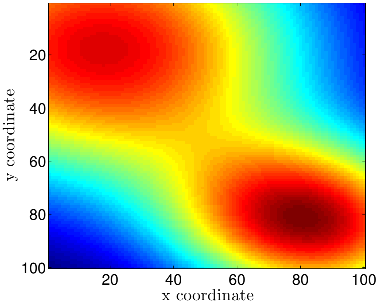


Application (2)

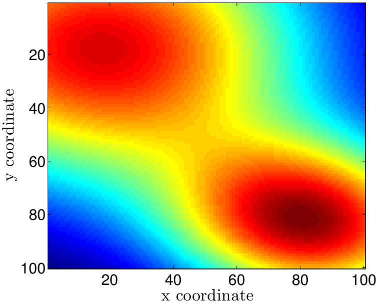
Reconstruction accuracy

Estimation of *A*

Theoretical scene



+ using sparse assumptions on *A* (SpIN-Cal)



The future of NMF... in LISIC

We now briefly discuss some of our theoretical work directions for the next 5 years:

1. being fast to process the data deluge (Ph.D. thesis of F. Yahaya)

The future of NMF... in LISIC

We now briefly discuss some of our theoretical work directions for the next 5 years:

1. being fast to process the data deluge (Ph.D. thesis of F. Yahaya)
2. Not a public information (submitted ANR)

The future of NMF... in LISIC

We now briefly discuss some of our theoretical work directions for the next 5 years:

1. being fast to process the data deluge (Ph.D. thesis of F. Yahaya)
2. Not a public information (submitted ANR)
3. Not a public information (starting work by G. Delmaire, G. Roussel, D. Courcot *et al.*)

The future of NMF... in LISIC

We now briefly discuss some of our theoretical work directions for the next 5 years:

1. being fast to process the data deluge (Ph.D. thesis of F. Yahaya)
2. Not a public information (submitted ANR)
3. Not a public information (starting work by G. Delmaire, G. Roussel, D. Courcot *et al.*)
4. Not a public information (starting work with S. Verel, G. Roussel *et al.*)

The future of NMF... in LISIC

We now briefly discuss some of our theoretical work directions for the next 5 years:

1. being fast to process the data deluge (Ph.D. thesis of F. Yahaya)
2. Not a public information (submitted ANR)
3. Not a public information (starting work by G. Delmaire, G. Roussel, D. Courcot *et al.*)
4. Not a public information (starting work with S. Verel, G. Roussel *et al.*)
5. Not a public information (starting work with G. Delmaire, G. Roussel *et al.*)

The future of NMF... in LISIC

We now briefly discuss some of our theoretical work directions for the next 5 years:

1. being fast to process the data deluge (Ph.D. thesis of F. Yahaya)
2. Not a public information (submitted ANR)
3. Not a public information (starting work by G. Delmaire, G. Roussel, D. Courcot *et al.*)
4. Not a public information (starting work with S. Verel, G. Roussel *et al.*)
5. Not a public information (starting work with G. Delmaire, G. Roussel *et al.*)

Being fast to process the data deluge

Historical NMF techniques and their extensions (including some of ours)

- Using slow techniques for the update rules (MU)

Being fast to process the data deluge

Historical NMF techniques and their extensions (including some of ours)

- Using slow techniques for the update rules (MU)

Fastening NMF

- Distributed computing (e.g., Liu *et al.*, 2010)
- Online factorization (e.g., Mairal *et al.*, 2010)
- Fast solver (e.g., Guan *et al.*, 2012, Dorffer *et al.*, 2017)
- Randomized strategies (e.g., Zhou *et al.*, 2012 or Tepper & Sapiro, 2016,

Being fast to process the data deluge

Historical NMF techniques and their extensions (including some of ours)

- Using slow techniques for the update rules (MU)

⇨

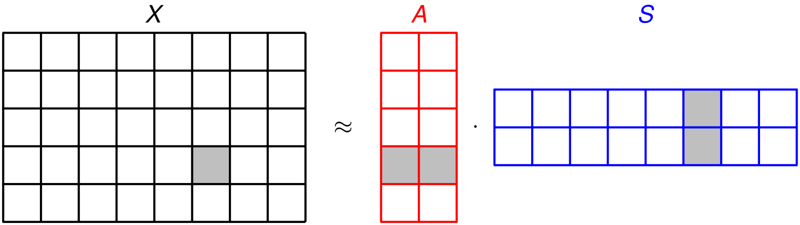
Fastening NMF

- Distributed computing (e.g., Liu *et al.*, 2010)
- Online factorization (e.g., Mairal *et al.*, 2010)
- Fast solver (e.g., Guan *et al.*, 2012, Dorffer *et al.*, 2017)
- Randomized strategies (e.g., Zhou *et al.*, 2012 or Tepper & Sapiro, 2016, Yahaya *et al.*, submitted)

Randomized strategies for NMF

Key idea

- Compress some matrices in the NMF problem in order to reduce its size



Randomized strategies for NMF

Key idea

- Compress some matrices in the NMF problem in order to reduce its size
- Compression based on **Random Projections** (preserving pairwise distances)

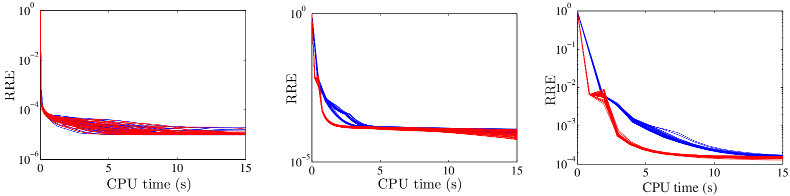
Randomized strategies for NMF

Key idea

- Compress some matrices in the NMF problem in order to reduce its size
- Compression based on **Random Projections** (preserving pairwise distances)

Reconstruction of rank-15 matrices of size 500×500 (left), 5000×5000 (middle) and 10000×10000 (right). Perf criterion:

$$\text{RRE} = \frac{\|X - AS\|_{\mathcal{F}}}{\|X\|_{\mathcal{F}}}$$



- In blue: the fast NeNMF method
- In red: our proposed faster-than-fast randomized NeNMF

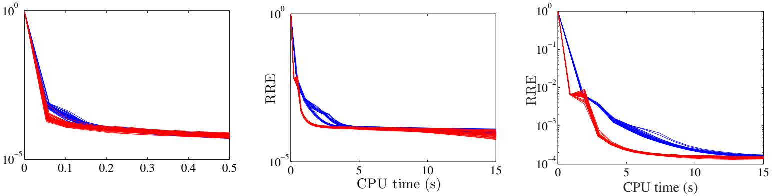
Randomized strategies for NMF

Key idea

- Compress some matrices in the NMF problem in order to reduce its size
- Compression based on **Random Projections** (preserving pairwise distances)

Reconstruction of rank-15 matrices of size 500×500 (left), 5000×5000 (middle) and 10000×10000 (right). Perf criterion:

$$\text{RRE} = \frac{\|X - AS\|_{\mathcal{F}}}{\|X\|_{\mathcal{F}}}$$



- In blue: the fast NeNMF method
- In red: our proposed faster-than-fast randomized NeNMF

Conclusion

- NMF $=$ powerful tool for signal, image processing and machine learning
- Good expertise in LISIC, with methodological developments
- Still many ideas and room for innovative work
- **Questions?**