Accelerating Weighted (Nonnegative) Matrix Factorization with Random Projections

Matthieu PUIGT

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

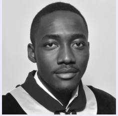
matthieu.puigt[at]univ-littoral.fr
http://www-lisic.univ-littoral.fr/~puigt/

Future of Random Matrices ♯4 – LightOn – December 19, 2019

Acknowledgements

Many thanks to:

- F. Yahaya, G. Delmaire, and G. Roussel



Région Hauts-de-France

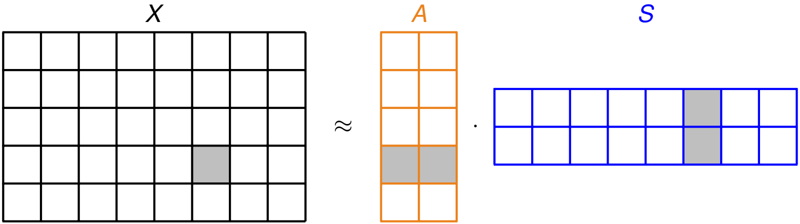
- The Ph.D. thesis of F. Yahaya is partly funded by the Région Hauts-de-France
- Experiments conducted using the CALCULCO computing platform, supported by SCoSI/ULCO

And also to:

- C. Dorffer, within the OSCAR & DoMasQ'Air projects

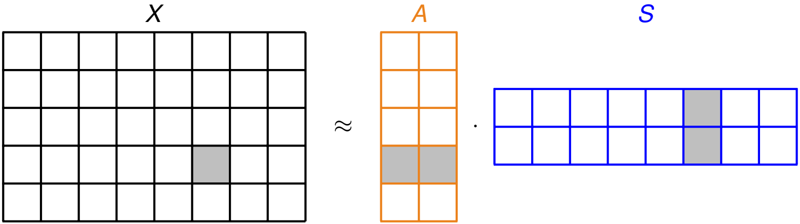
Foreword: matrix factorization

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 $X \approx A \cdot S$:



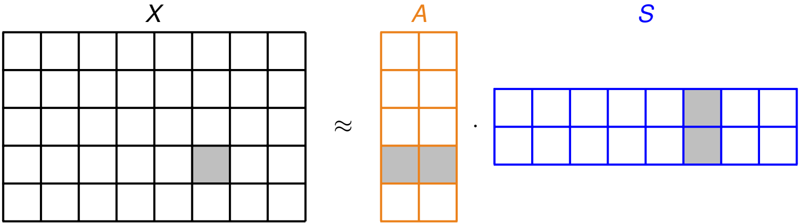
source localization/separation

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 (or if sensor array geometry is known)
- ➪ source separation (or beamforming)

Foreword: matrix factorization

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



Low-rank matrix approximation

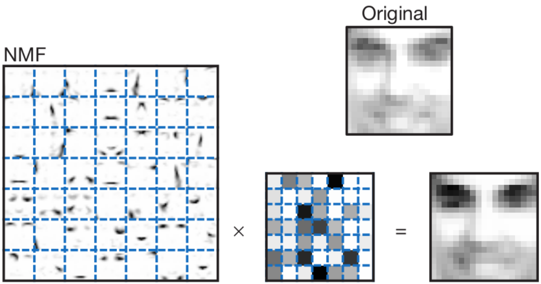
X noisy matrix (possibly with missing entries), A weight matrix, S latent variable matrix

- Topic modeling (e.g., online news)
- Collaborative filtering (e.g., Netflix Prize)
- Graph analysis (after a transformation into a matrix – e.g., bike sharing system)

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 yields **better interpretability**

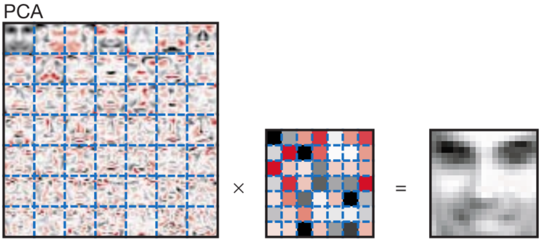


NMF applied to face dataset (source: Lee & Seung, 1999)

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 yields **better interpretability**



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

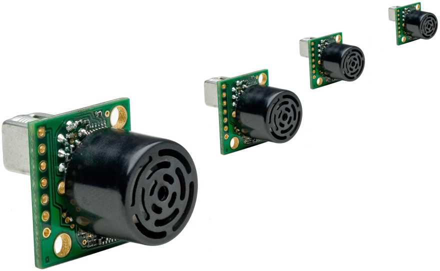
Application (motivation) – 1

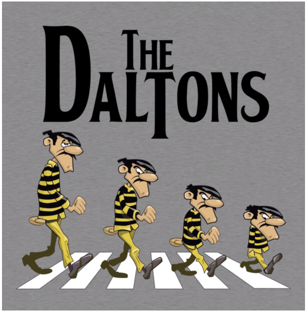
Mobile sensor calibration (Dorffer *et al.*, 2015–2018)

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

Application (motivation) – 1

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



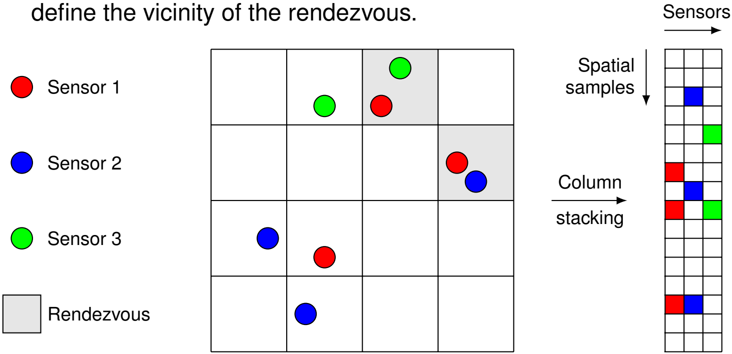


- Sensed phenomenon $\Longrightarrow$ voltage
- Voltage $\Longrightarrow$ phenomenon?
  - Sensor calibration needed
  - Not always physically possible
  - ➪ **Blind sensor calibration**

Application (motivation) – 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.



Scene $\mathcal{S}$ Observed matrix X

Application (motivation) – 3

Factorization

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

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

where

$$w(i,j) = \begin{cases} 0 & \text{if } x(i,j) \text{ not available} \\ \rho_j & \text{otherwise} \end{cases}$$

➪ Solution based on specific problem parameterization to handle information and MU

- We also proposed methods:
  - adding information[1] on A (column-wise sparse assumptions) and S (sensor information)
  - handling more complex calibration models[2] (e.g., nonlinear)

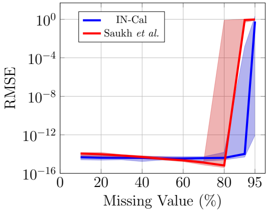
[1] C. Dorffer, MP, G. Delmaire, G. Roussel, *Informed Nonnegative Matrix Factorization Methods for Mobile Sensor Network Calibration*, IEEE TSIPN, 2018

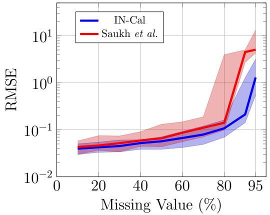
[2] C. Dorffer, MP, G. Delmaire, G. Roussel, *Nonlinear mobile sensor calibration using informed semi-nonnegative matrix factorization with a Vandermonde factor*, Proc. IEEE SAM'16

Application (motivation) – 4

Reconstruction accuracy

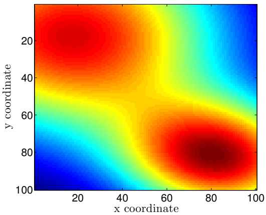
Estimation of *S*



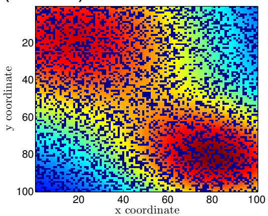


Estimation of *A*

Theoretical scene



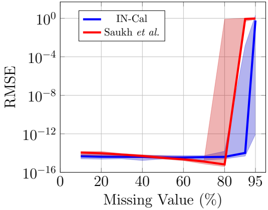
First informed NMF (IN-Cal)

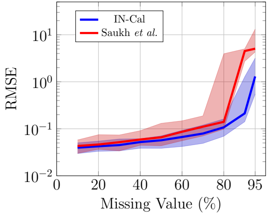


Application (motivation) – 4

Reconstruction accuracy

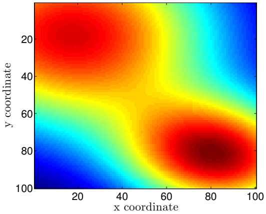
Estimation of *S*



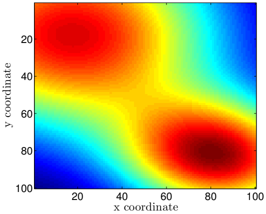


Estimation of *A*

Theoretical scene



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



Current work: NMF & Big Data (Yahaya *et al.*, 2018-)

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

- Using slow techniques for the update rules (MU, PG)
- Not well-suited for tall & skinny matrices (as met in mobile sensor calibration)

Current work: NMF & Big Data (Yahaya *et al.*, 2018-)

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

- Using slow techniques for the update rules (MU, PG)
- Not well-suited for tall & skinny matrices (as met in mobile sensor calibration)

Fastening NMF

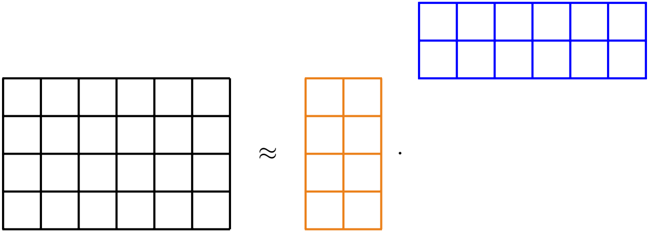
- Distributed computing (e.g., Liu *et al.*, 2010)
- Online factorization (e.g., Mairal *et al.*, 2010)
- Fast solver (e.g., Guan *et al.*, 2012)
- **Randomized strategies** (e.g., Zhou *et al.*, 2012, Tepper & Sapiro, 2016, Erichson *et al.*, 2018, or Yahaya⋆ *et al.*, 2018)

⋆ F. Yahaya, MP, G. Delmaire, G. Roussel, *Faster-than-fast NMF using random projections and Nesterov iterations*, Proc. iTwist'18.

Compressing NMF

One initially aim to solve:

$$X \approx A \cdot S$$

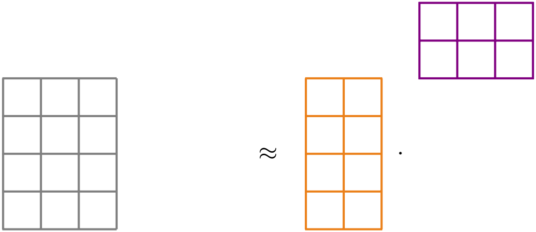


Compressing NMF

Dimensionality reduction of X by right multiplication with R:

$$\underbrace{X \cdot R}_{X_R} \approx A \cdot \underbrace{S \cdot R}_{S_R}$$

➪ **Update of A still possible**

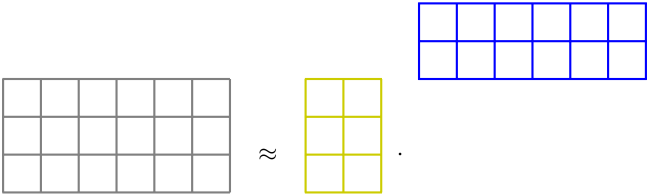


Compressing NMF

Dimensionality reduction of X by left multiplication with L:

$$\underbrace{L \cdot X}_{X_L} \approx \underbrace{L \cdot A}_{A_L} \cdot S$$

➪ **Update of S still possible**



How to design R and L ?

- **Random projections** is a popular tool in big data optimization
- Mathematical foundations based on the **Johnson-Lindenstrauss Lemma**:
  Given $0 < \varepsilon < 1$, a set X of n points in $\mathbb{R}^m$, and a number $k > 8\ln(n)/\varepsilon^2$, there is a linear map $f : \mathbb{R}^m \to \mathbb{R}^k$ such that :

$$\forall u, v \in X, \quad (1-\varepsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2$$

➪ Tentative compression matrices applied to NMF

- Gaussian random matrices as projection matrices (Zhou *et al.*, 2012)
- *Structured* random projections (Tepper & Sapiro, 2016)
  - ➪ Randomized Power Iterations (RPIs, Tepper & Sapiro, 2016)

$$B_L \triangleq (XX^T)^q \cdot X \cdot \Omega_L \qquad \text{and} \qquad L = \mathrm{QR}(B_L)$$

  - ➪ Randomized Subspace Iterations (RSIs, Yahaya *et al.*, 2018): similar to RPIs in theory but less sensitive to round-off errors

Problem statement

- In many problems, observed data matrix X with **missing entries or confidence measures** associated to each entry

➩ **Some applications:** collaborative filtering, source apportionment, mobile sensor calibration

➩ Weighted NMF (WNMF):

$$\min_{A,S\geq 0} \|W \circ X - W \circ (A \cdot S)\|_{\mathcal{F}}$$

- WNMF and Big Data:
  1. Most techniques are based on slow solvers (MU, PG)
  2. A few with similar fastening strategies as used in unweighted NMF, e.g., (Dorffer et al., 2017)[3]
  3. **No existing approach with random projections**

➩ How to use random projections in WNMF?

[3] C. Dorffer, MP, G. Delmaire, G. Roussel, *Fast nonnegative matrix factorization and completion using Nesterov iterations*, Proc. LVA-ICA'17

Strategies to solve Weighted NMF

Weighted Extensions of NMF aim to solve:

$$\min_{A,S\geq 0} \|W \circ X - W \circ (A \cdot S)\|_{\mathcal{F}}$$

1. Direct computations (Ho, 2008):
   - Incorporating the matrix W in the update rules
2. EM-based strategy (Zhang *et al.*, 2006)
   - E-step: Estimate the unknown entries of X

     $$X^{\text{comp}} = W \circ X + (\mathbb{1}_{n,m} - W) \circ (A \cdot S), \tag{1}$$

     where $\mathbb{1}_{n,m}$ is the $n \times m$ matrix of ones.
   - M-step: Apply any standard NMF technique to X^{comp}
3. Stochastic gradient (Hsieh & Dhillon, 2011, Yu *et al.*, 2012)
   - Update a single variable at a time
   - Only suitable for binary weights

Proposed Method

Require: initial matrices A and S
repeat
{**E-step**}
Compute X^{comp} as in (1)
Apply RPIs or RSIs to X^{comp} to compute L and R
Define $X_L^{\text{comp}} \triangleq L \cdot X^{\text{comp}}$ and $X_R^{\text{comp}} \triangleq X^{\text{comp}} \cdot R$
{!! Compression makes E-step 3 times slower in our experiments}
{**M-step**}
for compt=1 **to** $\text{Max}_{\text{OutIter}}$ **do**
Define $S_R \triangleq S \cdot R$
Solve

$$\min_{A \geq 0} \|X_R^{\text{comp}} - A \cdot S_R\|_{\mathcal{F}}$$

Define $A_L \triangleq L \cdot A$
Solve

$$\min_{S \geq 0} \|X_L^{\text{comp}} - A_L \cdot S\|_{\mathcal{F}}$$

{!! Each pass in the loop is 10-100 times faster than SotA EM-W-NMF methods in our experiments}
end for
until a stopping criterion

Experimental Validation

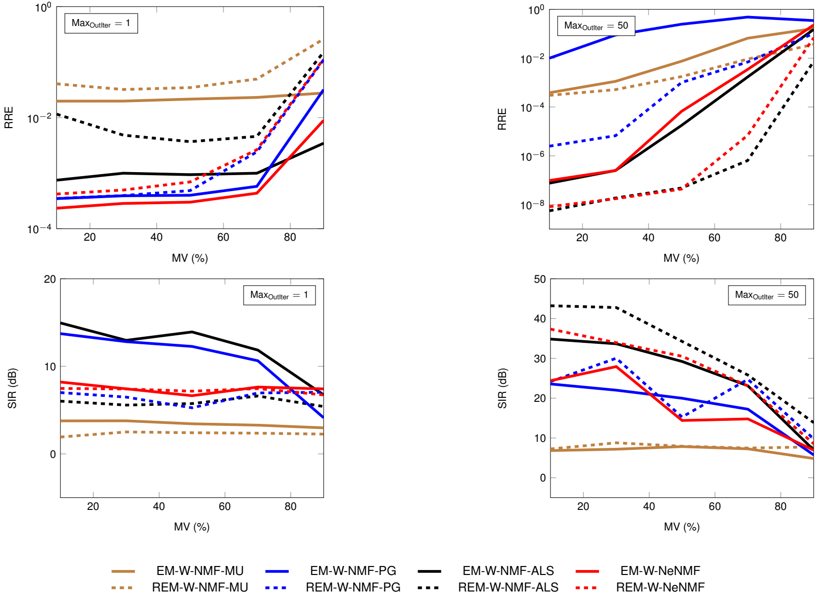
We run each test for 60 seconds and repeat 15 times the following:

- we randomly generate nonnegative factor matrices A^{theo} and S^{theo}—with $n = m = 10000$ and $p = 5$—$X^{\text{theo}} = A^{\text{theo}} \cdot S^{\text{theo}}$ without additive noise
- we randomly suppress data to generate X, with a sampling rate varying from 10 to 90% (with a step-size of 20%)
- Compression strategy: RPIs with $q = 4$ and oversampling $\nu = 10$ (i.e., L and R of size $(p+\nu) \times n$ and $m \times (p+\nu)$, resp.)
- Investigating the performance of the proposed REM-W-NMF strategy wrt. EM-W-NMF with several solvers, i.e., MU, PG, ALS, and Nesterov iterations
- Performance criteria: accuracy of reconstruction of X (Relative Reconstruction Error—RRE) and accuracy of estimation of S (Signal-to-Interference Ratio—SIR)

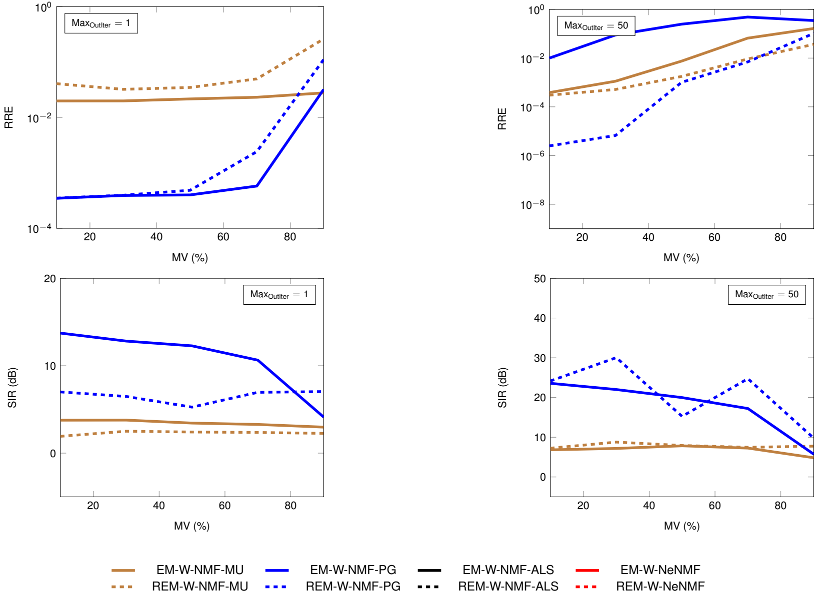
$$\text{RRE} \triangleq \left|\left|X^{\text{theo}} - A \cdot S\right|\right|_{\mathcal{F}}^{2} / \left|\left|X^{\text{theo}}\right|\right|_{\mathcal{F}}^{2}$$

$$\text{SIR} = \sum_{j=1}^{p} 10 \log_{10} \left(\left|\left|\hat{s}_j^{\text{coll}}\right|\right|^{2} / \left|\left|\hat{s}_j^{\text{orth}}\right|\right|^{2} \right)$$

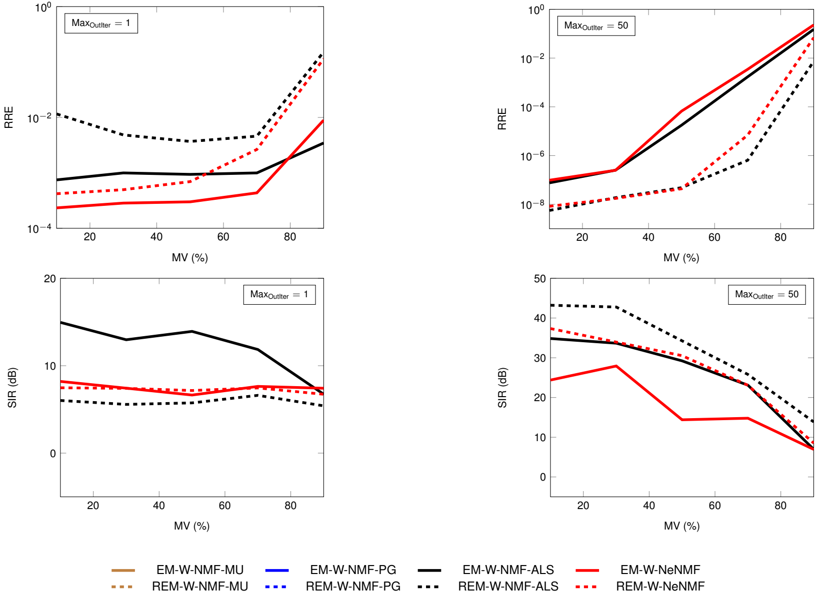
Results (EUSIPCO'19)



Results (EUSIPCO'19)



Results (EUSIPCO'19)



Further Experiments

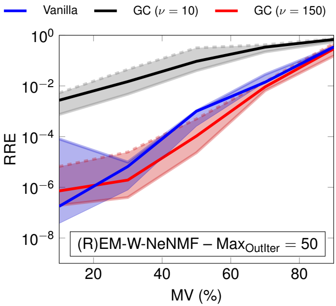
Preliminary Results

- Stability issues with ALS (zeroing negative entries)
- Other solvers: Active Set (Kim & Park, 2008), Nesterov, and their extrapolated extensions (Ang & Gillis, 2019)
  - ➩ Consistent results

Further Experiments

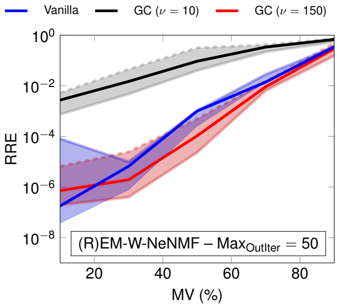
Preliminary Results

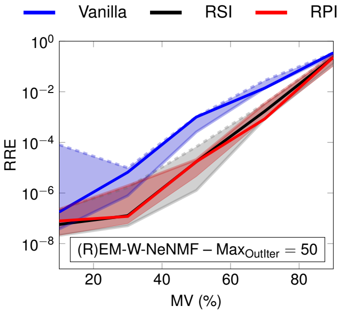
- Stability issues with ALS (zeroing negative entries)
- Other solvers: Active Set (Kim & Park, 2008), Nesterov, and their extrapolated extensions (Ang & Gillis, 2019)
  - ➩ Consistent results
- NMF performance with Gaussian compression:
  - When $\nu = 10$, bad!
  - When ν increases, better (e.g., $\nu = 150$ ➩ $\approx$ as good as Vanilla)



Further Experiments
Preliminary Results

- Stability issues with ALS (zeroing negative entries)
- Other solvers: Active Set (Kim & Park, 2008), Nesterov, and their extrapolated extensions (Ang & Gillis, 2019)
  - ➪ Consistent results
- NMF performance with Gaussian compression:
  - When $\nu = 10$, bad!
  - When ν increases, better (e.g., $\nu = 150$ ➪ $\approx$ as good as Vanilla)
- RPIs vs RSIs:
  - Similar median performance, slightly narrower envelope with RPIs
  - But simulations with noise not tested yet!





Conclusion

- A novel framework to combine random projections and weighted matrix factorization
- Based on an EM scheme
- Outperforms non-randomized SotA EM techniques under mild conditions
- Bottleneck: computing compression matrices at each E-step
  - Improving compression strategies
  - Using OPU
- In future work, we aim to extend the proposed strategy to informed and structured NMF techniques applied to mobile sensor calibration

Thank you!

Discover our work and try our codes

- Unweighted NMF with random projections: `https://gogs.univ-littoral.fr/puigt/Faster-than-fast_NMF`
- Mobile sensor calibration using Informed WNMF: `https://gogs.univ-littoral.fr/puigt/Informed_NMF_Mobile_Sensor_Calibration/`