



HAL
open science

NMF for Big Data with Missing Entries: A Random Projection Based Approach

Farouk Yahaya, Matthieu Puigt, Gilles Delmaire, Gilles Roussel

► **To cite this version:**

Farouk Yahaya, Matthieu Puigt, Gilles Delmaire, Gilles Roussel. NMF for Big Data with Missing Entries: A Random Projection Based Approach. 6ème Journée Régionale des Doctorants en Automatique et Traitement du Signal (JRDA), Jul 2019, Lille, France. hal-02166582

HAL Id: hal-02166582

<https://hal.science/hal-02166582>

Submitted on 1 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NMF for Big Data with Missing Entries: A Random Projection Based Approach

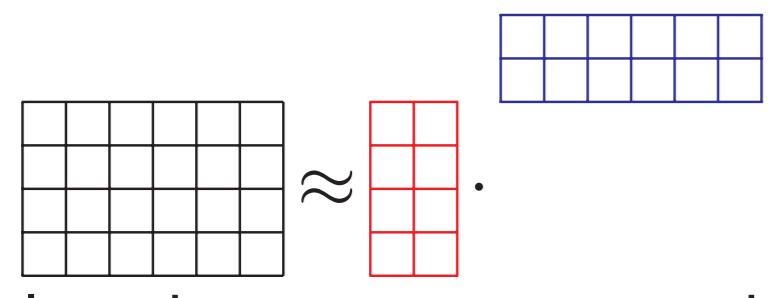
Farouk YAHAYA, Matthieu PUIGT, Gilles DELMAIRE, and Gilles ROUSSEL

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

Firstname.LASTNAME@univ-littoral.fr

Context and problem statement

$$X \approx W \cdot H$$



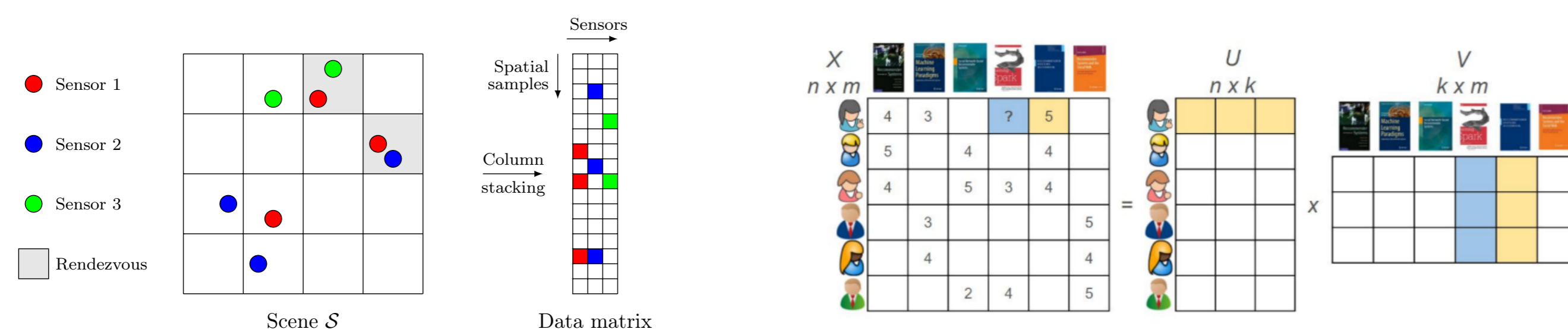
► **Nonnegative Matrix Factorization (NMF)** is a popular tool in Signal/Image Processing and Machine Learning

- Goal: estimate two nonnegative $n \times p$ and $p \times m$ matrices W and H such that an observed low-rank nonnegative $n \times m$ matrix X can be written as $X \approx W \cdot H$
- Some applications:
 - **Source separation**, dictionary learning, graph analysis, topic modelling, **hyperspectral unmixing**...
- Why is NMF so popular? **Better interpretability** than no-sign-constrained approaches



NMF and PCA applied to face dataset (source: [1])

- How is NMF working?
 - Iterative procedure where W and H are alternately updated
 - Historical techniques known to be slow (multiplicative updates, projected gradients, nonnegative least squares, etc)
- NMF and Big Data: How to face the data deluge?
 - Distributed computing (e.g., [2])
 - Online factorization (e.g., [3])
 - Fast solver (e.g., [4])
 - Randomized strategies (e.g., [5, 6, 7])
- In many problems, observed data matrix X with **missing entries or confidence measures** associated to each entry
 - Some applications: collaborative filtering, **source apportionment**, **low-rank nonnegative matrix completion**, **mobile sensor calibration** (aim of the Ph.D. thesis)



mobile sensor calibration (source: [8]).

recommender systems (source: [9])

◊ **Weighted NMF (WNMF):**

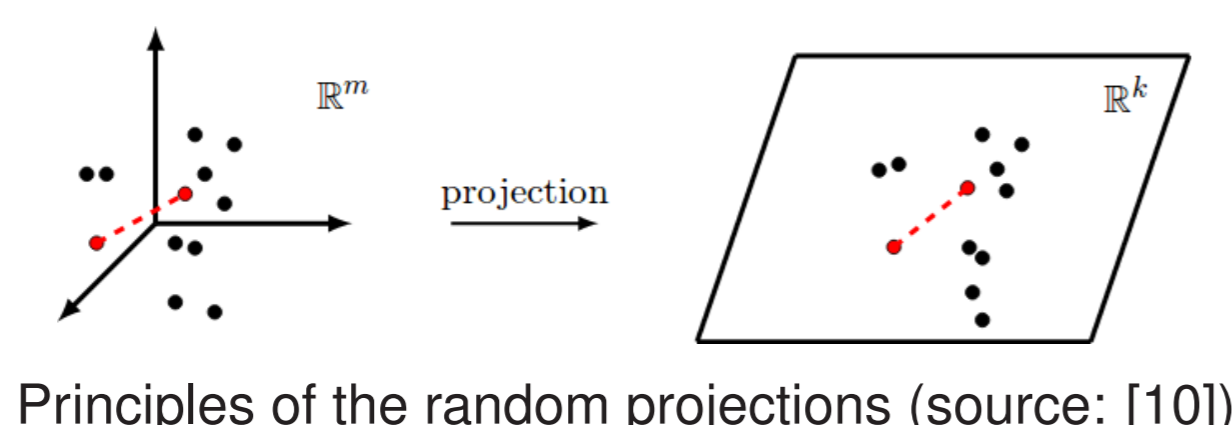
$$\min_{W, H \geq 0} \|Q \circ X - Q \circ (W \cdot H)\|_{\mathcal{F}}$$

- WNMF and Big Data:
 1. Most techniques are based on slow solvers
 2. A few with the fastening strategies for unweighted NMF
 3. **No existing approach with random projections**

◊ **How to use random projections in WNMF?**

Principle of random projections applied to NMF

- **Random projections** is a popular tool in machine learning to speed-up computations while preserving pairwise structure
- Mathematical foundations based on the **Johnson-Lindenstrauss Lemma**



Principles of the random projections (source: [10])

- Gaussian random matrices as projection matrices [5]
- **Structured** random projections [6, 7]
 - ◊ **Randomized Power Iterations** (RPIs) or **Randomized Subspace Iterations** (RSIs)

Require: a target rank ν (with $p \leq \nu \ll \min(n, m)$) and an integer q

Draw Gaussian random matrices $\Omega_L \in \mathbb{R}^{m \times \nu}$ and $\Omega_R \in \mathbb{R}^{\nu \times n}$

Define $B_L \triangleq (X X^T)^q \cdot X \cdot \Omega_L$ and $B_R \triangleq \Omega_R \cdot X \cdot (X^T X)^q$

Derive $L \in \mathbb{R}^{\nu \times n}$ and $R \in \mathbb{R}^{m \times \nu}$ by QR decompositions of B_L and B_R , respectively.

- RSIs are similar but add intermediate QR decompositions in the computations ◊ less sensitive to round-off errors than RPIs

◊ **Why we have designed "compressions matrices" R and L ?**

Dimensionality reduction of X by right (resp. left) multiplication with R (resp. L)

$$\underbrace{X \cdot R}_{X_R} \approx \underbrace{W \cdot H}_{H_R} \quad \underbrace{L \cdot X}_{X_L} \approx \underbrace{L \cdot W}_{W_L} \cdot H$$



NMF with weights

Weighted Extensions of NMF aim to solve:

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

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

$$X^{\text{comp}} = Q \circ X + (\mathbb{1}_{n,m} - Q) \circ (W \cdot H), \quad (1)$$

where $\mathbb{1}_{n,m}$ is the $n \times m$ matrix of ones.

- M-step: Apply any standard NMF technique to X^{comp}

◊ **Direct incorporation of random projections is not trivial...**

Proposed Method

Require: initial matrices W and H

repeat

{**E-step**} {Compression makes E-step 3 times slower in our experiments}

Compute X^{comp} as in (1)

Apply RSIs or RPIs to X^{comp} to compute L and R

Define $X_L^{\text{comp}} \triangleq L \cdot X^{\text{comp}}$ and $X_R^{\text{comp}} \triangleq X^{\text{comp}} \cdot R$

{**M-step**} {10-100 times faster than SotA EM-W-NMF methods in our experiments}

for $\text{comp}=1$ **to** MaxOuter **do**

Define $H_R \triangleq H \cdot R$

Solve $\min_{W \geq 0} \|X_R^{\text{comp}} - W H_R\|_{\mathcal{F}}$

Define $W_L \triangleq L \cdot W$

Solve $\min_{H \geq 0} \|X_L^{\text{comp}} - W_L H\|_{\mathcal{F}}$

end for

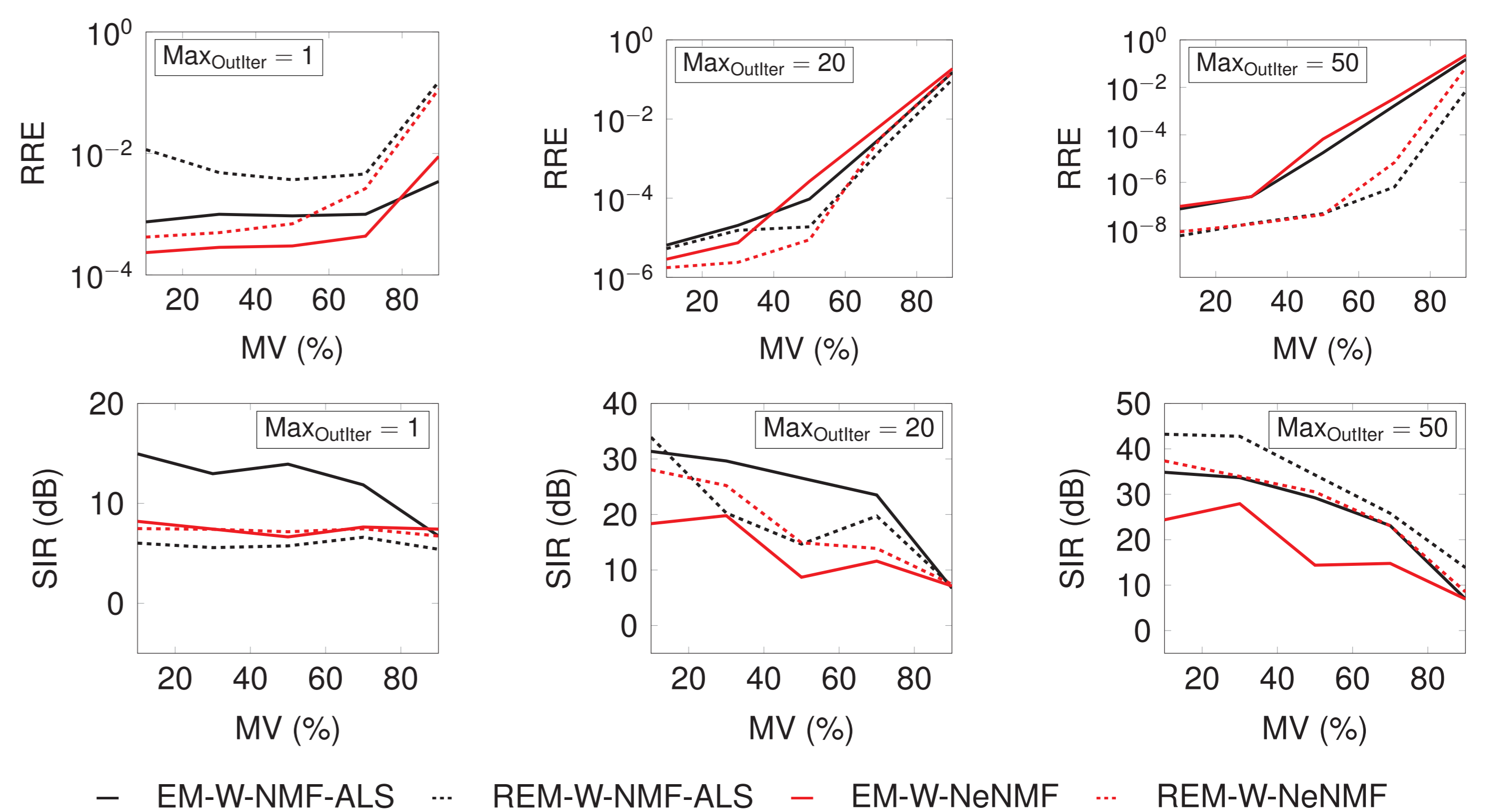
until a stopping criterion

Experiments and Results

To test our method we run each test for 60 seconds and repeat 15 times the following:

- we randomly generate nonnegative factor matrices W^{theo} and H^{theo} , with $n = m = 10000$ and $p = 5$.
- we randomly suppress data to generate X , with a sampling rate varying from 10 to 90% (with a step-size of 20%)
- We compare the proposed REM-W-NMF strategy to the EM-W-NMF with several solvers, i.e., MU, PG, ALS & Nesterov iterations
- measure the accuracy of reconstruction of X Relative Reconstruction Error (RRE) and the accuracy of estimation of H (Signal-to-Interference Ratio—SIR)

$$\text{RRE} \triangleq \frac{\|X^{\text{theo}} - W \cdot H\|_{\mathcal{F}}^2}{\|X^{\text{theo}}\|_{\mathcal{F}}^2} \quad (2) \quad \text{SIR} = \sum_{j=1}^p 10 \log_{10} \left(\frac{\|\hat{h}_j^{\text{coll}}\|^2}{\|\hat{h}_j^{\text{orth}}\|^2} \right) \quad (3)$$



Conclusion and perspectives

- Novel framework to combine random projections and weighted matrix factorization.
- Based on an EM scheme
- The proposed strategy outperforms non-randomized state-of-the-art EM techniques
- In future work, we aim to apply the proposed strategy to informed and structured NMF techniques applied to mobile sensor calibration

References and acknowledgments

- [1] D.D. Lee and H.S. Seung. "Learning the parts of objects by non negative matrix factorization". In: *Nature* 401.6755 (1999), pp. 788–791.
- [2] C. Liu, H.-C. Yang, J. Fan, L.-W. He, and Y.-M. Wang. "Distributed Nonnegative Matrix Factorization for Web-Scale Dyadic Data Analysis on MapReduce". In: *Proc. WWW Conf.'10*. Apr. 2010.
- [3] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. "Online learning for matrix factorization and sparse coding". In: *Journal of Machine Learning Research* 11.Jan (2010), pp. 19–60.
- [4] N. Guan, D. Tao, Z. Luo, and B. Yuan. "NeNMF: An optimal gradient method for nonnegative matrix factorization". In: *IEEE Trans. on Signal Proc.* 60.6 (2012), pp. 2882–2898.
- [5] F. Wang and P. Li. "Efficient nonnegative matrix factorization with random projections". In: *Proc. SIAM ICDM'10*. SIAM. 2010, pp. 281–292.
- [6] M. Tepper and G. Sapiro. "Compressed Nonnegative Matrix Factorization Is Fast and Accurate". In: *IEEE Trans. on Signal Proc.* 64.9 (May 2016), pp. 2269–2283.
- [7] N. B. Erichson, A. Mendible, S. Wihlbom, and J. N. Kutz. "Randomized Nonnegative Matrix Factorization". In: *Pattern Recognition Letters* (2018).
- [8] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel. "Informed Nonnegative Matrix Factorization Methods for Mobile Sensor Network Calibration". In: *IEEE Trans. Sig. Inf. Proc. Networks* 4.4 (Dec. 2018), pp. 667–682.
- [9] O. Gindele. *Recommender systems in Tensorflow*. https://www.slideshare.net/seldon_io/tensorflow-london-by-oliver-gindele-recommender-systems-in-tensorflow. Feb. 2018.
- [10] N. B. Erichson, S. Voronin, S. L. Brunton, and J. N. Kutz. "Randomized matrix decompositions using R". In: *Journal of Statistical Software, Articles* 89.11 (2019).



F. Yahaya gratefully acknowledges the Région Hauts-de-France to partly fund his PhD fellowship.