



**HAL**  
open science

## Do random projections fasten an already fast NMF technique using Nesterov optimal gradient?

Farouk Yahaya, Matthieu Puigt, Gilles Delmaire, Gilles Roussel

### ► To cite this version:

Farouk Yahaya, Matthieu Puigt, Gilles Delmaire, Gilles Roussel. Do random projections fasten an already fast NMF technique using Nesterov optimal gradient?. 5ème Journée Régionale des Doctorants en Automatique et Traitement du Signal (JRDA), Jul 2018, Amiens, France. hal-01831111

**HAL Id: hal-01831111**

**<https://hal.science/hal-01831111v1>**

Submitted on 18 Aug 2018

**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.

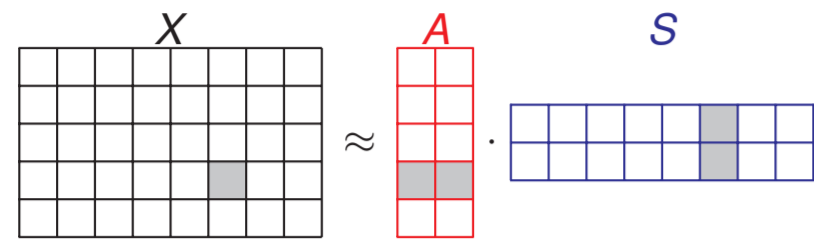
# Do random projections fasten an already fast NMF technique using Nesterov optimal gradient?

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

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

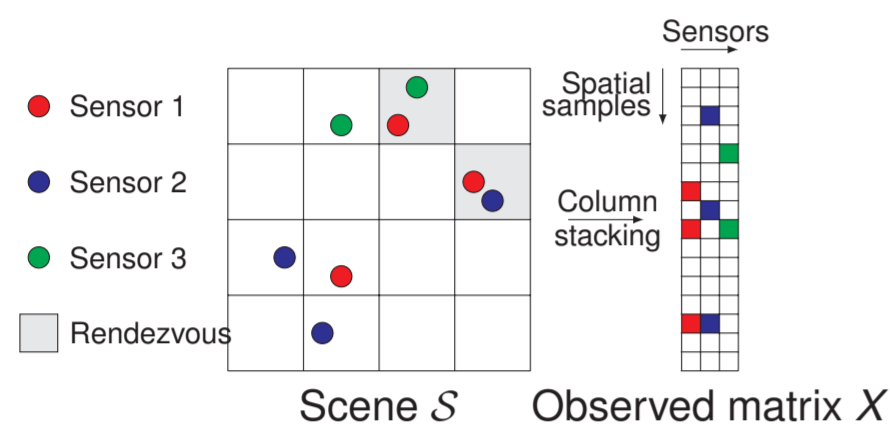
Firstname.LASTNAME@univ-littoral.fr

## Context and problem statement



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

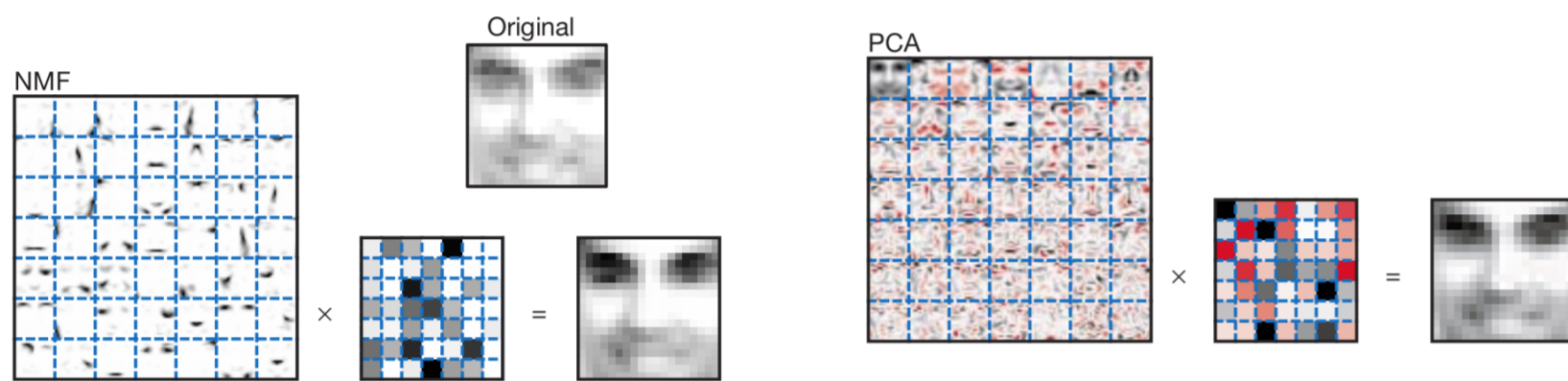
- Goal: estimate two nonnegative  $n \times p$  and  $p \times m$  matrices  $A$  and  $S$  such that an observed low-rank nonnegative  $n \times m$  matrix  $X$  can be written as  $X \approx A \cdot S$  :
- Some applications:
  - **Source separation**, dictionary learning, graph analysis, topic modelling, collaborative filtering, **low-rank nonnegative matrix completion**, **sensor calibration**...
- (informed) NMF for blind sensor calibration



Blind calibration revisited as a weighted NMF problem (affine model) [1]

$$W_0 \begin{bmatrix} x(1,1) & \dots & x(1,m-1) & y(1) \\ \vdots & & \vdots & \vdots \\ x(n,1) & \dots & x(n,m-1) & y(n) \end{bmatrix} \approx W_0 \begin{bmatrix} y(1) & 1 \\ \vdots & \vdots \\ y(n) & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{m-1} & 1 \\ \beta_1 & \beta_2 & \dots & \beta_{m-1} & 0 \end{bmatrix}$$

- Why is NMF so popular? **Better interpretability** than no-sign-constrained approaches



NMF and PCA applied to face dataset (source: [2])

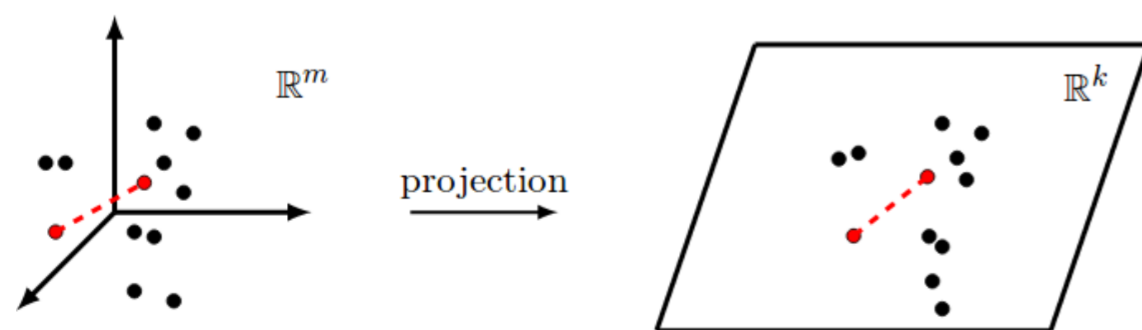
- How is NMF working?
  - Iterative procedure where  $A$  and  $S$  are alternately updated
  - Historical techniques known to be slow (multiplicative updates, projected gradients, nonnegative least squares, etc)
- NMF and Big Data: How to face the data deluge?
  - Distributed computing (e.g., [3])
  - Online factorization (e.g., [4])
  - Fast solver (e.g., [5], [6])
  - Randomized strategies (e.g., [7, 8, 9, 10])
- Most randomized NMF techniques are combined with a slow solver (except for separable NMF, which is out of the scope of this poster).
- ◊ **Is there any interest to combine a fast solver with randomization?**

## NMF with Nesterov iterations

- In 1983, Nesterov fastens the plain Gradient Descent by replacing the line search for optimal stepsize by its proposed iterative Optimal Gradient Descent
- ◊ NeNMF [5] much faster than MU-NMF [2] or PG-NMF [11]
- The NeNMF algorithm consists of:
  - an outer loop
    - initialize  $t = 1$ ,  $A^1$ , and  $S^1$
    - loop
      - $t \leftarrow t+1$ ;
      - $S^{t+1} \leftarrow \text{NeInnerLoopF}(X, A^t, S^t)$ ;
      - $A^{t+1} \leftarrow \text{NeInnerLoopG}(X, A^t, S^{t+1})$ ;
    - end loop
  - an inner loop (here detailed for an update of  $F$ )
    - define  $\alpha_0 = 1$ ,  $\alpha_{k+1} = \frac{1 + \sqrt{4\alpha_k^2 + 1}}{2}$ ,  $\forall k \in \mathbb{N}$
    - initialize  $Y_0 \triangleq S^t$ ,  $k = 0$
    - repeat
      - $S_k = (Y_k - \frac{1}{2} \nabla_{S^t} \mathcal{J}(A, Y_k))^+$  with  $\mathcal{L} \triangleq \|A\|_2^2$  Lipschitz constant
      - $Y_{k+1} = S_k + \frac{\alpha_k - 1}{\alpha_{k+1}} (S_k - S_{k-1})$ ,
    - until Stopping criterion reached at Iteration  $K$
    - return  $S^{t+1} = Y_K$
    - Stopping criterion:  $k = \text{MaxIter}$  or  $\left\| \nabla_{S^t}^+ \mathcal{J}(A, Y_k) \right\|_{\mathcal{F}} \leq \epsilon \cdot \left\| \left[ \nabla_{A^t}^+ \mathcal{J}(A^t, S^t), \nabla_{S^t}^+ \mathcal{J}(A^t, S^t) \right] \right\|_{\mathcal{F}}$
    - +  $k > \text{MinIter}$

## Principles of random projections

- In an *ideal world* (i.e., noiseless and errorless), only two rows and columns of  $X$  are needed to update  $S$  and  $A$  on the top figure, respectively
- But we are not living in an ideal world...
  - **Random projections** is a popular tool in machine learning to speed-up computations while preserving pairwise structure



Principles of the random projections (source: [12])

- Design of the compression scheme for NMF (e.g., using power iteration)

**Require:** a target rank  $\nu$  (with  $p \leq \nu \ll \min(n, m)$ ) and an integer  $q$  (e.g.,  $q = 4$  in [9])  
 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.

- NMF with random projections:

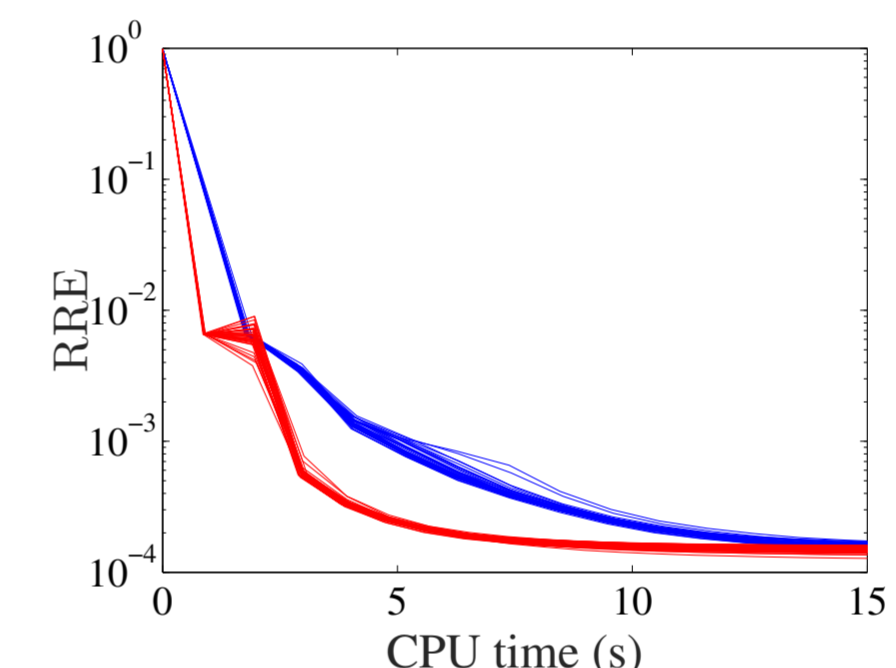
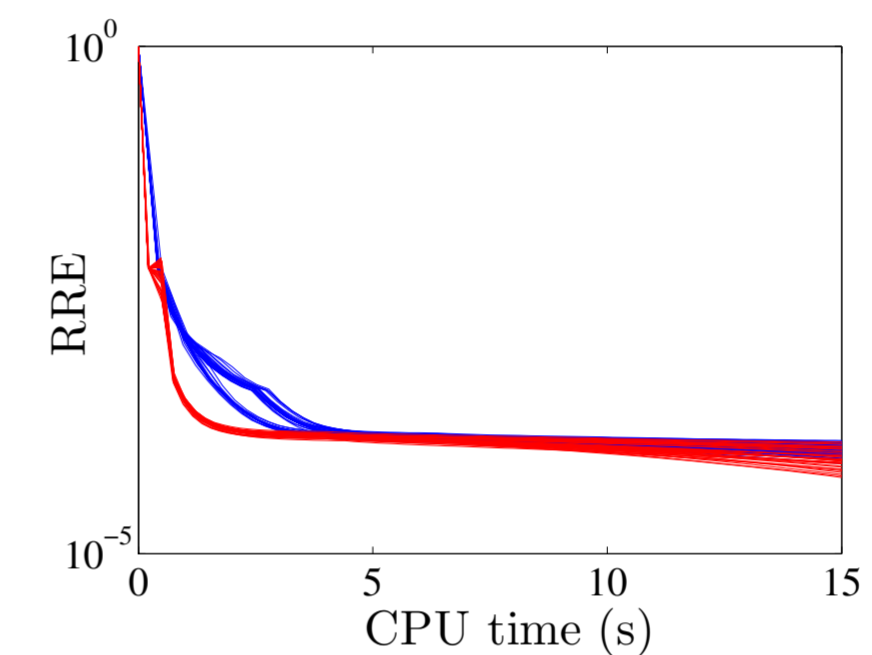
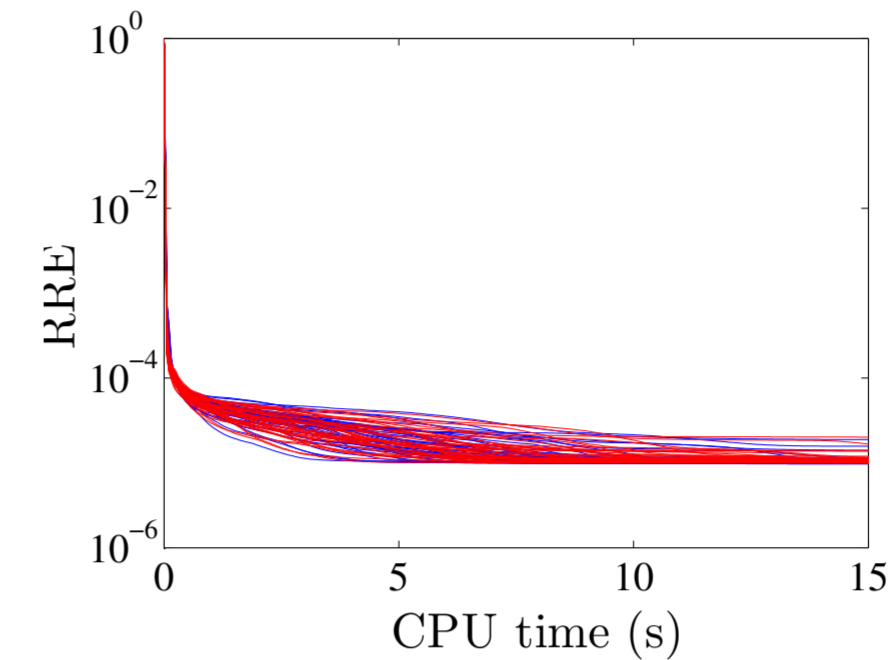
**Require:** initial matrices  $A$  and  $S$ , and compression matrices  $L$  and  $R$ .  
 Define  $X_L \triangleq L \cdot X$  and  $X_R \triangleq X \cdot R$   
**repeat**  
 Define  $S_R \triangleq S \cdot R$   
 Update  $A$  by resp. replacing  $X$  and  $S$  by  $X_R$  and  $S_R$   
 Define  $A_L \triangleq L \cdot A$   
 Update  $S$  by resp. replacing  $X$  and  $A$  by  $X_L$  and  $A_L$   
**until** a stopping criterion

## Experiments

- Simulations with rank-15 matrices  $X$  of size  $m = n = 500, 5000$ , and  $10000$ .
  - Theoretical matrices  $A$  and  $S$  are known
  - Input SNR  $\approx 30$  dB
  - Methods are run 15 s and their performance is regularly computed
  - Performance criterion:

$$\text{RRE} \triangleq \frac{\|X - A \cdot S\|_{\mathcal{F}}}{\|X\|_{\mathcal{F}}}$$

- Randomized NeNMF vs (already fast) vanilla NeNMF



Performance of the randomized NeNMF (in red) vs the vanilla one (in blue) with respect to time.  
 Top:  $n = 500$ . Middle:  $n = 5000$ . Bottom:  $n = 10000$ .

## Conclusion and perspectives

- An experimental study of the combination of random projections with an optimal solver for NMF
- The proposed approach is shown to be really fast (much more than vanilla NeNMF for very large matrices)
- In future work, we will extend the approach to weighted NMF (matrix completion, blind sensor calibration)

## References and acknowledgments

- [1] 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* (to appear).
- [2] 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.
- [3] 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*.
- [4] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. "Online learning for matrix factorization and sparse coding". In: *Journal of Machine Learning Research* 11. Jan (2010), pp. 19–60.
- [5] Naiyang Guan, Dacheng Tao, Zhigang Luo, and Bo Yuan. "NeNMF: An optimal gradient method for nonnegative matrix factorization". In: 60.6 (2012), pp. 2882–2898.
- [6] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel. "Fast nonnegative matrix factorization and completion using Nesterov iterations". In: *Proc. LVA/ICA'17. Vol. LNCS 10179. 2017*, pp. 26–35.
- [7] Fei Wang and Ping Li. "Efficient nonnegative matrix factorization with random projections". In: *Proc. SIAM ICDM'10. SIAM. 2010*, pp. 281–292.
- [8] G. Zhou, A. Cichocki, and S. Xie. "Fast Nonnegative Matrix/Tensor Factorization Based on Low-Rank Approximation". In: 60.6 (June 2012), pp. 2928–2940.
- [9] M. Tepper and G. Sapiro. "Compressed Nonnegative Matrix Factorization Is Fast and Accurate". In: 64.9 (May 2016), pp. 2269–2283.
- [10] N Benjamin Erichson, Ariana Mendible, Sophie Wihlborn, and J Nathan Kutz. "Randomized Nonnegative Matrix Factorization". In: *Pattern Recognition Letters* (2018).
- [11] C. Lin. "Projected Gradient Methods for Nonnegative Matrix Factorization". In: *Neural Computation* 19.10 (Oct. 2007), pp. 2756–2779.
- [12] N Benjamin Erichson, Sergey Voronin, Steven L Brunton, and J Nathan Kutz. "Randomized matrix decompositions using R". In: *arXiv preprint arXiv:1608.02148* (2016).



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