



**HAL**  
open science

# Stable recovery of the factors from a deep matrix product

François Malgouyres, Joseph Landsberg

► **To cite this version:**

François Malgouyres, Joseph Landsberg. Stable recovery of the factors from a deep matrix product. SPARS, 2017, Lisbonne, Portugal. hal-01417943v1

**HAL Id: hal-01417943**

**<https://hal.science/hal-01417943v1>**

Submitted on 16 Dec 2016 (v1), last revised 20 Mar 2017 (v2)

**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.

# Stable recovery of the factors from a deep matrix product

François Malgouyres

Institut de Mathématiques de Toulouse ; UMR5219  
 Université de Toulouse ; CNRS  
 UPS IMT, F-31062 Toulouse Cedex 9, France  
<http://www.math.univ-toulouse.fr/~fmalgouy/>

Joseph Landsberg

Department of Mathematics  
 Mailstop 3368  
 Texas A& M University  
 College Station, TX 77843-3368  
 Email: jml@math.tamu.edu

**Abstract**—We study a deep matrix factorization problem. It takes as input the matrix  $X$  obtained by multiplying  $K$  matrices (called factors) and aims at recovering the factors. When  $K = 1$ , this is the usual compressed sensing framework;  $K = 2$ : Examples of applications are dictionary learning, blind deconvolution, self-calibration;  $K \geq 3$ : includes convolution network problems as well as many fast transforms.

Using a Lifting, we provide : a necessary and sufficient conditions for the identifiability of the factors (up to a scale rearrangement); - an analogue of the Null-Space-Property, called the Deep-Null-Space-Property which is necessary and sufficient to guarantee the stable recovery of the factors.

## I. INTRODUCTION

Let  $K \in \mathbb{N}^*$ ,  $m_1 \dots m_{K+1} \in \mathbb{N}$ , write  $m_1 = m$ ,  $m_{K+1} = n$ . We impose the factors to be structured matrices defined by a (typically small) number  $S$  of unknown parameters. More precisely, for  $k = 1 \dots K$ , let

$$\begin{aligned} M_k : \mathbb{R}^S &\longrightarrow \mathbb{R}^{m_k \times m_{k+1}}, \\ h &\longmapsto M_k(h) \end{aligned}$$

be a linear map.

We assume that we know the matrix  $X \in \mathbb{R}^{m \times n}$  which is provided by

$$X = M_1(\mathbf{h}_1) \cdots M_K(\mathbf{h}_K) + e, \quad (1)$$

for an unknown error term  $e$  and parameters  $\mathbf{h} = (\mathbf{h}_k)_{1 \leq k \leq K} \in \mathcal{M}^{\bar{L}} \subset \mathbb{R}^{S \times K}$  for some  $\bar{L}$ , where we assume that we know a collection of models  $\mathcal{M} = (\mathcal{M}^L)_{L \in \mathbb{N}}$  such that, for every  $L$ ,  $\mathcal{M}^L \subset \mathbb{R}^{S \times K}$ .

This work investigates models/constraints imposed on (1) for which we can (up to obvious scale rearrangement) identify or stably recover the parameters  $\mathbf{h}$  from  $X$ . A preliminary version of this work is presented in [1].

Set  $\mathbb{N}_K = \{1, \dots, K\}$  and

$$\mathbb{R}_*^{S \times K} = \{\mathbf{h} \in \mathbb{R}^{S \times K}, \forall k \in \mathbb{N}_K, \|\mathbf{h}_k\| \neq 0\}.$$

Define an equivalence relation in  $\mathbb{R}_*^{S \times K}$ : for any  $\mathbf{h}, \mathbf{g} \in \mathbb{R}_*^{S \times K}$ ,  $\mathbf{h} \sim \mathbf{g}$  if and only if there exists  $(\lambda_k)_{k \in \mathbb{N}_K} \in \mathbb{R}^K$  such that

$$\prod_{k=1}^K \lambda_k = 1 \quad \text{and} \quad \forall k \in \mathbb{N}_K, \mathbf{h}_k = \lambda_k \mathbf{g}_k.$$

Denote the equivalence class of  $\mathbf{h} \in \mathbb{R}_*^{S \times K}$  by  $[\mathbf{h}]$ . We consider a metric denoted  $d_p$  on  $\mathbb{R}_*^{S \times K} / \sim$ . It is based on the  $l^p$  norm.

We say that a tensor  $T \in \mathbb{R}_*^{S \times K}$  is of rank 1 if and only if there exists a collection of vectors  $\mathbf{h}_k \in \mathbb{R}_*^{S \times K}$  such that  $T$  is the outer product of the vectors  $\mathbf{h}_k$ , for  $k \in \mathbb{N}_K$ , that is, for any  $\mathbf{i} \in \mathbb{N}_S^K$ ,

$$T_{\mathbf{i}} = \mathbf{h}_{1,i_1} \cdots \mathbf{h}_{K,i_K}.$$

The set of all the tensors of rank 1 is denoted by  $\Sigma_1$ .

Moreover, we parametrize  $\Sigma_1 \subset \mathbb{R}^{S^K}$  by the Segre embedding

$$\begin{aligned} P : \mathbb{R}^{S \times K} &\longrightarrow \Sigma_1 \subset \mathbb{R}^{S^K} \\ \mathbf{h} &\longmapsto (\mathbf{h}_{1,i_1} \mathbf{h}_{2,i_2} \cdots \mathbf{h}_{K,i_K})_{\mathbf{i} \in \mathbb{N}_S^K} \end{aligned}$$

Following [2], [3], [4], [5], [6], [7] where problems such that  $K = 2$  are studied, we can lift the problem and show that the map

$$(\mathbf{h}_1, \dots, \mathbf{h}_K) \longmapsto M_1(\mathbf{h}_1)M_2(\mathbf{h}_2) \cdots M_K(\mathbf{h}_K),$$

uniquely determines a linear map

$$\mathcal{A} : \mathbb{R}^{S^K} \longrightarrow \mathbb{R}^{m \times n},$$

such that for all  $\mathbf{h} \in \mathbb{R}^{S \times K}$

$$M_1(\mathbf{h}_1)M_2(\mathbf{h}_2) \cdots M_K(\mathbf{h}_K) = \mathcal{A}P(\mathbf{h}).$$

When  $\|e\| = 0$ , we can prove that every element of  $\mathbf{h} \in \mathcal{M}$  is identifiable (i.e. the elements of  $[\mathbf{h}]$  are the only solutions of (1)) if and only if for any  $L$  and  $L' \in \mathbb{N}$

$$\text{Ker}(\mathcal{A}) \cap (P(\mathcal{M}^L) - P(\mathcal{M}^{L'})) = \{0\}.$$

When  $\|e\| \leq \delta$ , we further assume that we have a way to find  $L^*$  and  $\mathbf{h}^* \in \mathcal{M}^{L^*}$  such that, for some parameter  $\eta > 0$ ,

$$\|\mathcal{A}P(\mathbf{h}^*) - X\|^2 \leq \eta. \quad (2)$$

Using the following definition we can prove Theorem 1.

### Definition 1. Deep-Null Space Property

Let  $\gamma > 0$ , we say that  $\text{Ker}(\mathcal{A})$  satisfies the deep-Null Space Property (deep-NSP) with respect to the model collection  $\mathcal{M}$  with constant  $\gamma$  if there exists  $\varepsilon > 0$  such that for any  $L$  and  $L' \in \mathbb{N}$ , any  $T \in P(\mathcal{M}^L) - P(\mathcal{M}^{L'})$  satisfying  $\|AT\| \leq \varepsilon$  and any  $T' \in \text{Ker}(\mathcal{A})$ , we have

$$\|T\| \leq \gamma \|T - T'\|.$$

### Theorem 1. Sufficient condition for stable recovery

Assume  $\text{Ker}(\mathcal{A})$  satisfies the deep-NSP with respect to the collection of models  $\mathcal{M}$  and with the constant  $\gamma > 0$ . For any  $\mathbf{h}^*$  as in (2) with  $\eta$  and  $\delta$  sufficiently small, we have

$$\|P(\mathbf{h}^*) - P(\bar{\mathbf{h}})\| \leq \frac{\gamma}{\sigma_{\min}} (\delta + \eta),$$

where  $\sigma_{\min}$  is the smallest non-zero singular value of  $\mathcal{A}$ . Moreover, if  $\bar{\mathbf{h}} \in \mathbb{R}_*^{S \times K}$

$$d_p([\mathbf{h}^*], [\bar{\mathbf{h}}]) \leq \frac{7(KS)^{\frac{1}{p}} \gamma}{\sigma_{\min}} \min \left( \|P(\bar{\mathbf{h}})\|_{\infty}^{\frac{1}{K}-1}, \|P(\mathbf{h}^*)\|_{\infty}^{\frac{1}{K}-1} \right) (\delta + \eta).$$

We also prove that the deep-NSP condition is **necessary** for the stable recovery of the factors. We study a variante of the deep-NSP, dedicated to models  $\mathcal{M}$  enforcing sparsity, and discuss the applications of these results for the stable recovery of convolution network.

## REFERENCES

- [1] F. Malgouyres and J. Landsberg, "On the identifiability and stable recovery of deep/multi-layer structured matrix factorization," in *IEEE, Info. Theory Workshop*, Sept. 2016.
- [2] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1711–1732, 2014.
- [3] S. Choudhary and U. Mitra, "Identifiability scaling laws in bilinear inverse problems," *arXiv preprint arXiv:1402.2637*, 2014.
- [4] X. Li, S. Ling, T. Strohmer, and K. Wei, "Rapid, robust, and reliable blind deconvolution via nonconvex optimization," *CoRR*, vol. abs/1606.04933, 2016. [Online]. Available: <http://arxiv.org/abs/1606.04933>
- [5] S. Bahmani and J. Romberg, "Lifting for blind deconvolution in random mask imaging: Identifiability and convex relaxation," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2203–2238, 2015.
- [6] S. Ling and T. Strohmer, "Blind deconvolution meets blind demixing: Algorithms and performance bounds," *arXiv preprint arXiv:1512.07730*, 2015.
- [7] —, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, p. 115002, 2015.