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; $K \geq 3$: can be applied to many fast transforms (such as the FFT). In particular, we apply the theorems to deep convolutional network.**

Using a Lifting, we provide : a necessary and sufficient conditions for the identifiability of the factors (up to a scale indeterminacy); - 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.

The long article corresponding to this work is available in [1].

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

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

be a linear map.

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

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

for an unknown error term e and parameters $\mathbf{h} = (\mathbf{h}_k)_{1 \leq k \leq K} \in \mathcal{M}^{\overline{L}} \subset \mathbb{R}^{S \times K}$ for some $\overline{L}$, 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 [2].

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

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

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

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

The set of all the tensors of rank 1 is denoted by Σ_1.

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

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

Following [3], [4], [5], [6], [7], [8] where problems such that $K = 2$ are studied, we can *lift* the problem and show that the map

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

uniquely determines a linear map

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

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

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

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 $L' \in \mathbb{N}$

$$\mathrm{Ker}\,(\mathcal{A}) \cap \left(P(\mathcal{M}^L) - P(\mathcal{M}^{L'})\right) = \{0\}.$$

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 $\eta > 0$,

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

Definition 1. Deep-Null Space Property
Let $\gamma > 0$*, we say that* $\mathrm{Ker}\,(\mathcal{A})$ *satisfies the* deep-Null Space Property (deep-NSP) *with respect to the model collection* $\mathcal{M}$ *with constant* γ *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* $\|\mathcal{A}T\| \leq \varepsilon$ *and any* $T' \in \mathrm{Ker}\,(\mathcal{A})$*, we have*

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

Theorem 1. Sufficient condition for stable recovery
Assume $\mathrm{Ker}\,(\mathcal{A})$ *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* η *and* δ *sufficiently small, we have*

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

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

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

We also prove that the deep-NSP condition is **necessary** for the stable recovery of the factors. We detail how these results can be applied to obtain sharp conditions for the stable recovery of deep convolutional network as depicted on Figure 1.

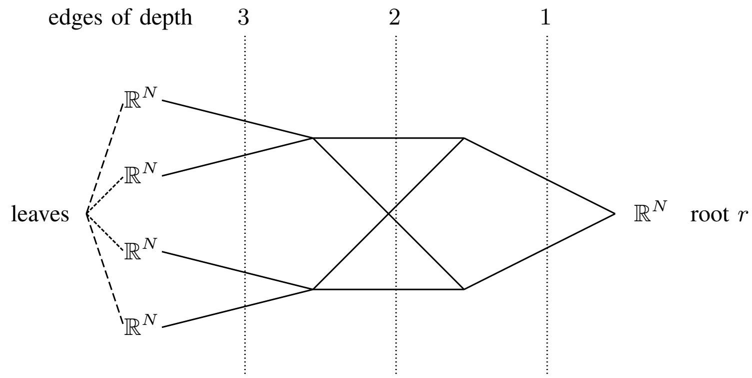

Fig. 1. Example of the considered convolutional network. To every edge is attached a convolution kernel. The network does not involve non-linearities or sampling.

References

[1] F. Malgouyres and J. Landsberg, "Stable recovery of the factors from a deep matrix product and application to convolutional network," *preprint hal-01494267*, 2017.

[2] ——, "On the identifiability and stable recovery of deep/multi-layer structured matrix factorization," in *IEEE, Info. Theory Workshop*, Sept. 2016.

[3] 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.

[4] S. Choudhary and U. Mitra, "Identifiability scaling laws in bilinear inverse problems," *arXiv preprint arXiv:1402.2637*, 2014.

[5] 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

[6] 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.

[7] S. Ling and T. Strohmer, "Blind deconvolution meets blind demixing: Algorithms and performance bounds," *arXiv preprint arXiv:1512.07730*, 2015.

[8] ——, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, p. 115002, 2015.