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 $K \in \mathbb{N}^*$, $m_1, \ldots, m_{K+1} \in \mathbb{N}$, 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 $k = 1 \ldots K$, let

$$M_k : \mathbb{R}^S \rightarrow \mathbb{R}^{m_k \times m_{k+1}},$$

be a linear map.

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

$$X = M_1(h_1) \cdots M_K(h_K) + e,$$  

for an unknown error term $e$ and parameters $h = (h_k)_{1 \leq k \leq K} \in \mathcal{M}^L \subset \mathbb{R}^{S \times K}$ for some $L$, 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 $h$ from $X$. A preliminary version of this work is presented in [2].

Set $N_K = \{1, \ldots, K\}$ and

$$\mathbb{R}^{S \times K}_2 = \{h \in \mathbb{R}^{S \times K}, \forall k \in N_K, \|h_k\| \neq 0\}.$$  

Define an equivalence relation in $\mathbb{R}^{S \times K}_2$: for any $h$, $g \in \mathbb{R}^{S \times K}$, $h \sim g$ if and only if there exists $(\lambda_k)_{k \in N_K} \in \mathbb{R}^K$ such that

$$\prod_{k=1}^K \lambda_k = 1$$  

and $\forall k \in N_K$, $h_k = \lambda_k g_k$.

Denote the equivalence class of $h \in \mathbb{R}^{S \times K}_2$ by $[h]$. We consider a metric denoted $d_p$ on $\mathbb{R}^{S \times K}/ \sim$. It is based on the $p^*$ 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 $h_k \in \mathbb{R}^{S \times K}$ such that $T$ is the outer product of the vectors $h_k$, for $k \in N_K$, that is, for any $i \in N_S^K$,

$$T_i = h_{i,1} \cdots h_{i,K}.$$  

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

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

$$P : \mathbb{R}^{S \times K} \rightarrow \Sigma_1 \subset \mathbb{R}^{S \times K}$$

$$h \mapsto (h_{1,1} \ h_{2,2} \cdots h_{K,i,K})_{i \in N_S^K}.$$  

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

$$(h_1, \ldots, h_K) \mapsto M_1(h_1)M_2(h_2) \cdots M_K(h_K),$$  

uniquely determines a linear map

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

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

$$M_1(h_1)M_2(h_2) \cdots M_K(h_K) = AP(h).$$

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

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

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

$$\|AP(h^*) - X\|^2 \leq \eta.$$  

Theorem 1. Sufficient condition for stable recovery

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

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

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

$$d_p([h^*],[\bar{h}]) \leq \frac{(KS)^{1/2}}{\sigma_{\min}} \min (\|P(\bar{h})\|_{\frac{1}{p}-1}, \|P(h^*)\|_{\frac{1}{p}-1}) (\delta + \eta).$$

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