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. Signal Processing with Adaptive Sparse Structured Representations (SPARS), 2017, Lisbonne, Portugal. hal-01417943v2

HAL Id: hal-01417943
https://hal.archives-ouvertes.fr/hal-01417943v2
Submitted on 20 Mar 2017

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; \( K \geq 3 \) 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.

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

\[
M_{k} : \mathbb{R}^{S} \rightarrow \mathbb{R}^{m_{k} \times m_{k+1}},
\]

\[
h \mapsto M_{k}(h)
\]

be a linear map.

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

\[
X = M_{1}(h_{1}) \ldots M_{K}(h_{K}) + \epsilon,
\]

for an unknown error term \( \epsilon \) and parameters \( h = (h_{k})_{1 \leq k \leq K} \in \mathcal{M}^{L} \subset \mathbb{R}^{S \times K} \) for some \( \mathcal{T} \), 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 [1].

Set \( N_{K} = \{1, \ldots, K\} \) and

\[
\mathbb{R}^{S \times K} = \{ h \in \mathbb{R}^{S \times K}, \forall k \in N_{K}, \| h_{k} \| \neq 0 \}.
\]

Define an equivalence relation in \( \mathbb{R}^{S \times K} \): 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

\[
\prod_{k=1}^{K} \lambda_{k} = 1 \quad \text{and} \quad \forall k \in N_{K}, h_{k} = \lambda_{k} g_{k}.
\]

Denote the equivalence class of \( h \in \mathbb{R}_{S \times K} \) by \( [h] \). We consider a metric denoted \( d_{p} \) on \( \mathbb{R}^{S \times K} / \sim \), it is based on the \( p^{th} \) 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 \( h \in \mathbb{R}^{S \times K} \) such that \( T \) is the outer product of the vectors \( h_{i} \), for \( k \in N_{K} \), that is, for any \( i \in N_{K}^{S} \),

\[
T_{i} = h_{i,1} \ldots h_{K,i_{K}}.
\]

The set of all the tensors of rank \( 1 \) is denoted by \( \Sigma_{1} \).

Moreover, we parametrize \( \Sigma_{1} \subset \mathbb{R}^{S \times K} \) by the Segre embedding

\[
P : \mathbb{R}^{S \times K} \rightarrow \Sigma_{1} \subset \mathbb{R}^{S K}
\]

\[
h \mapsto (h_{1,1}, h_{2,1}, \ldots h_{K,1_{K}})_{i \in N_{S}^{K}}
\]

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

\[
(h_{1}, \ldots, h_{K}) \mapsto M_{1}(h_{1})M_{2}(h_{2}) \ldots M_{K}(h_{K}),
\]

uniquely determines a linear map

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

such that for all \( h \in \mathbb{R}^{S K} \)

\[
M_{1}(h_{1})M_{2}(h_{2}) \ldots M_{K}(h_{K}) = AP(h).
\]

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

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

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

\[
\| AP(h^{*}) - X \|^{2} \leq \eta.
\]

Definition 1. Deep-Null Space Property

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

\[
\| T' \| \leq \gamma \| T - T' \|.
\]

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

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

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

\[
d_{p}(\hat{h}^{*}; \hat{h}) \leq \frac{7(KS)^{1/2} \gamma}{\sigma_{\min}} \min \left( \| P(\hat{h}) \|^{1/2}, \| P(h^{*}) \|^{1/2} \right) (\delta + \eta).
\]

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


