Randomized methods for higher-order subspace separation

Michele N. da Costa, Renato R. Lopes, João Marcos T. Romano
DSPCom Laboratory, University of Campinas (UNICAMP), Campinas, Brazil

***Abstract*—This paper presents an algorithm for signal subspace separation in the context of multidimensional data. The proposal is an extension of the randomized Singular Value Decomposition (SVD) for higher-order tensors. From a set derived from random sampling, we construct an orthogonal basis associated with the range of each mode-space of the input data tensor. Multilinear projection of the input data onto each mode-space then transforms the data to a low-dimensional representation. Finally, we compute the Higher-Order Singular Value Decomposition (HOSVD) of the reduced tensor. Furthermore, we propose an algorithm for computing the randomized HOSVD based on the row-extraction technique. The results reveal a relevant improvement from the standpoint of computational complexity.**

***Index Terms*—higher-order singular value decomposition, randomized algorithm, signal subspace method, tensor decomposition, dimension reduction, row-extraction technique.**

I. Introduction

Matrix factorizations decompose a matrix into two or more matrix factors. In general, the decompositions can be used to store (or to process) all data in an inexpensive way regarding memory and computational cost, as well as to discover certain latent features underlying the data and also when part of it is missing or inaccurate.

Generally speaking, decompositions are distinguished by additional constraints on the factors and a particular factorization could be more appropriate depending upon the problem to be solved. The best computational method for computing factorizations depends on several factors, such as matrix properties, availability of memory, acceptable accuracy, and robustness. In many applications, one looks for an approximation to the input data in fewer degrees of freedom to reduce the resulting computational complexity.

Randomized sampling techniques can be used to construct a low-rank approximation to a given matrix [1]–[4]. In general, these methods construct a low-dimensional subspace that approximates the range of the matrix and then computes a factorization of the reduced matrix via variations of classical deterministic methods.

Halko *et al.* [4] propose a dimensionality-reduction technique based on a randomized low-rank factorization. The idea is to sample randomly the matrix aiming at identifying a subspace that approximates the range of an input data in which most of the information is captured. In the sequence, the matrix is projected to the approximate range before computing the Singular Value Decomposition (SVD) of the reduced matrix.

E-mail: nazareth@decom.fee.unicamp.br; rlopes@decom.fee.unicamp.br; romano@decom.fee.unicamp.br.

An attractive advantage of their algorithm is the robustness and accuracy attained with a low computational complexity cost. For an $m \times n$ input matrix with k dominant components of the SVD, their method requires $\mathcal{O}\big(mn\log(k) + (m+n)k^2\big)$ floating-point operations (flops) in contrast with $\mathcal{O}(mnk)$ for classical algorithms. The asymptotic algorithmic complexity reduction is achieved by employing a structured random matrix to provide an orthogonal basis for the range of the input matrix and by computing the SVD via row-extraction technique [4].

The use of a matrix approach for applications in which data structure representation naturally fits more than two domains could be inappropriate, especially when the computational load is affected by multiple iterations. Furthermore, the inherent conditions of the classical bidimensional analysis are more restrictive than those of the multidimensional analysis [5], [6].

A multilinear generalization of the SVD called the Higher-Order Singular Value Decomposition (HOSVD) was introduced by Lathauwer [7]. Although this decomposition is a particular case of the Tucker model [8], the HOSVD is analogously defined as the matrix SVD, regarding singular vectors and values, and has some properties similar to those of the standard SVD [7], [9].

For the matrix approach, the SVD is a factorization of the form $\mathbf{X} = \mathbf{USV}^{\mathrm{H}}$, where $\mathbf{S}$ is a pseudo-diagonal matrix. In the higher-order case, the *core matrix* $\mathbf{S}$ is replaced by the *core tensor* $\mathcal{S}$, which is in general a full tensor with a weaker condition than being pseudo-diagonal (i.e. $s_{i_1,i_2,\ldots,i_N} \neq 0$ only if $i_1 = i_2 = \ldots = i_N$). As a consequence, the HOSVD does not allow the direct interpretation of tensor rank concerning the minimum number of rank-one tensors sufficient to decompose a tensor. Analogously to the matrix approach, the HOSVD can be applied to decompose input data into orthogonal subspaces, as firstly proposed in [10].

Although the usual definition of tensor rank [11] is a generalization of the definition of matrix rank, determining the tensor rank is not easy and can only be achieved in some special cases [12]–[14]. The problem of determining the best approximation of the rank of a given higher-order tensor has been extensively investigated [9], [15]–[19], being useful as a tool for dimensionality reduction, data compression, and signal subspace estimation. A good rank approximation can be obtained by restricting the rank of *matrix unfoldings*, called the *multilinear rank* of a tensor. Nonetheless, this approach leads to a suboptimal solution in general, contrarily to the matrix case [9], [15].

The present article is motivated by applications involving high-dimensional data and proposes a method for computing a lower dimensionality representation of the input data tensor

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

as a pre-processing step to estimate the signal subspace. It also leads to an improvement of the computation complexity. In fact, our method can be viewed as an extension of the randomized SVD to higher-order tensors. Based on the row-extraction technique employed in [4] for computing the SVD at lower computational cost, we also propose an extension of this method for computing the randomized HOSVD.

In this sense, the focus is not to determine the best approximation of a given multilinear rank to $\mathcal{X}$, neither to investigate a new technique for truncating the HOSVD, since there are several alternative strategies for truncating the HOSVD as *higher-order orthogonal iteration* (HOOI) [15], *sequentially truncated HOSVD* (ST-HOSVD) [19], and optimization-based on the trust-region and on the conjugate gradients methods [16], [18].

The paper is organized as follows. First, we provide a brief overview of some basic definitions and also an introduction to HOSVD, which will be useful throughout this paper. In Section III, we present the randomized method for higher-order subspace separation. A performance analysis of the proposed methods is carried out in Section IV. Finally, Section V draws some conclusions and perspectives for future works.

II. Definitions

In this paper, N-th order tensors (for $N \geq 3$), matrices (second-order tensors), vectors (first-order tensors), and scalars (zero-order tensors) are denoted by calligraphic letters $(\mathcal{A}, \mathcal{B}, \dots)$, boldface upper-case $(\mathbf{A}, \mathbf{B}, \dots)$, boldface lower-case $(\mathbf{a}, \mathbf{b}, \dots)$, and lower-case $(a, b, \dots)$, respectively. $\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{\mathrm{H}}$, and $\mathbf{A}^{*}$ stand for transpose, Hermitian transpose, and complex conjugate of $\mathbf{A}$, respectively. The selection vector $\hat{\mathbf{e}}_l \triangleq [0 \cdots 0\, 1\, 0 \cdots 0]^{\mathrm{T}} \in \mathbb{R}^L$, $l \in \{1, \dots, L\}$, is defined as an unit vector with 1 in the l-th row and zeros elsewhere. The Kronecker and Khatri-Rao products are denoted by $\otimes$ and $\diamond$ respectively.

Definition 1. *The n-mode product of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ and a matrix $\mathbf{U} \in \mathbb{C}^{J_n \times I_n}$ is an $(I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N)$-tensor given by* $[\mathcal{A} \times_n \mathbf{U}]_{i_1,\dots,i_{n-1},j_n,i_{n+1},\dots,i_N} \triangleq \sum_{i_n=1}^{I_n} a_{i_1,\dots,i_n,\dots,i_N} u_{j_n,i_n}$, *for all index values.*

Definition 2. *The n-mode product of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ and a vector $\mathbf{u} \in \mathbb{C}^{I_n}$ is an $(I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N)$-tensor given by* $[\mathcal{A} \times_n \mathbf{u}]_{i_1,\dots,i_{n-1},i_{n+1},\dots,i_N} \triangleq \sum_{i_n=1}^{I_n} a_{i_1,\dots,i_n,\dots,i_N} u_{i_n}$, *for all index values.*

Definition 3. (Tensor rank). *Any N-th order tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ can be written as a sum of R rank-one tensors, i.e.*

$$\mathcal{X} = \sum_{r=1}^{R} \lambda_r \, \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \cdots \circ \mathbf{a}_r^{(N)}, \tag{1}$$

where $R \triangleq \operatorname{rank}(\mathcal{X})$ is the minimum number that satisfies (1).

The decomposition of a tensor $\mathcal{X}$ in the form of (1) is referred to as the *Canonical Decomposition* (CANDECOMP) [20] or *Parallel Factors* (PARAFAC) [21] model. This definition is a generalization of a minimum sum of rank one matrices to higher-order tensors.

The Higher-Order Singular Value Decomposition (HOSVD) of an N-th order tensor decomposes multilinear data into N orthogonal bases, each associated with a mode (or dimension) of the tensor. Calculating the HOSVD consists of the computation of N SVDs of different matrices constructed from the elements of the N-th order tensor (or *matrix unfoldings*).

The HOSVD of an N-th order tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ will be written as the product

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)}. \tag{2}$$

III. A new strategy for HOSVD

In this section, we derive two different methods for performing HOSVD by employing the randomized SVD [4].

By definition, the matrix unfoldings of an N-th order tensor, $\mathbf{X}^{(n)}$ for $n \in \{1, .., N\}$, are, in general, rectangular matrices with size $(I_n \times I_1 \dots I_{n-1} I_{n+1} \dots I_N)$ for $n \in \{1, \dots, N\}$, since I_n tends to be smaller than $I_1 \dots I_{n-1} I_{n+1} \dots I_N$ for any n.

According to the randomized SVD procedure [4], a randomized matrix $\mathbf{W}^{(n)} \in \mathbb{C}^{I_n \times K_n}$ can be employed to obtain a series of vectors that span the range of the unfolded matrix $\mathbf{X}^{(n)}$. Then, we can use the QR decomposition of these vectors to obtain an orthonormal basis $\mathbf{Q}^{(n)}$ for the range of $\mathbf{X}^{(n)}$, i.e.,

$$\begin{aligned} \mathbf{Y}^{(n)} &= \mathbf{X}^{(n)^{\mathrm{T}}} \mathbf{W}^{(n)} \\ &= \mathbf{Q}^{(n)} \mathbf{R}^{(n)} \in \mathbb{C}^{I_1 \dots I_{n-1} I_{n+1} \dots I_N \times K_n + p}, \end{aligned} \tag{3}$$

where p denotes an *oversampling parameter*, typically set to 5-10 [4]. This parameter is used to ensure that the columns of $\mathbf{Y}^{(n)}$ actually cover the entire range of each matrix unfolding.

In the second stage of the randomized SVD, we compute the SVD on the reduced matrix $\mathbf{B}^{(n)} \in \mathbb{C}^{I_n \times K_n + p}$ given by

$$\mathbf{B}^{(n)} = \mathbf{X}^{(n)} \mathbf{Q}^{(n)}. \tag{4}$$

Note that $\mathbf{B}^{(n)}$ is much smaller than $\mathbf{X}^{(n)}$, so this decomposition can be obtained with low complexity.

Given the SVD of $\mathbf{B}^{(n)}$, i.e., $\mathbf{B}^{(n)} = \mathbf{U}^{(n)} \mathbf{\Sigma}^{(n)} \mathbf{V}^{(n)^{\mathrm{H}}}$, and the approximation $\mathbf{X}^{(n)} \mathbf{Q}^{(n)} \mathbf{Q}^{(n)^{\mathrm{H}}} \approx \mathbf{X}^{(n)}$, the SVD of each matrix unfolding $\mathbf{X}^{(n)}$ can be directly computed as follows

$$\mathbf{X}^{(n)} = \mathbf{U}^{(n)} \mathbf{\Sigma}^{(n)} \tilde{\mathbf{V}}^{(n)^{\mathrm{H}}}, \tag{5}$$

with $\tilde{\mathbf{V}}^{(n)} \triangleq \mathbf{Q}^{(n)} \mathbf{V}^{(n)} \in \mathbb{C}^{I_1 \dots I_{n-1} I_{n+1} \dots I_N \times K_n + p}$.

A. Generalization of the randomized SVD to higher-order tensors

The main feature of the randomized SVD is that it transforms a matrix $\mathbf{X}^{(n)}$ into a smaller matrix $\mathbf{B}^{(n)}$, and then computes the simpler SVD of $\mathbf{B}^{(n)}$. The transformation is carried out using an orthonormal basis of the range space of $\mathbf{X}^{(n)}$, so that obtaining the SVD of $\mathbf{X}^{(n)}$ from the SVD of $\mathbf{B}^{(n)}$ is straightforward. In this section, we extend this idea to a higher-order tensor $\mathcal{X}$.

For the sake of simplicity, we restrict the analysis to the simplest case of higher-order tensors, i.e. for $N = 3$. Let the

random vectors $\{\mathbf{w}_1^{(n)},\ldots,\mathbf{w}_{K_n}^{(n)}\}$ associated with each n-th mode form a linearly independent set, such that

$$\begin{aligned}\mathbf{y}_k^{(1)} &= \mathcal{X}\times_2 \mathbf{w}_k^{(2)} \times_3 \mathbf{w}_k^{(3)} \in \mathbb{C}^{I_1},\\ \mathbf{y}_k^{(2)} &= \mathcal{X}\times_1 \mathbf{w}_k^{(1)} \times_3 \mathbf{w}_k^{(3)} \in \mathbb{C}^{I_2},\\ \mathbf{y}_k^{(3)} &= \mathcal{X}\times_1 \mathbf{w}_k^{(1)} \times_2 \mathbf{w}_k^{(2)} \in \mathbb{C}^{I_3},\end{aligned} \tag{6}$$

for $k\in\{1,\ldots,K_n\}$ and $n\in\{1,2,3\}$. Each expression in (6) is as a linear mapping from $\mathbb{C}^{I_1\times I_2\times I_3}$ to $\mathbb{C}^{I_1}$, $\mathbb{C}^{I_2}$, and $\mathbb{C}^{I_3}$, respectively.

From (1), one can show that

$$\begin{aligned}\mathbf{y}_k^{(1)} &= \sum_{r=1}^{R}\gamma_{r,k}^{(1)}\,\mathbf{a}_r^{(1)}, \quad \gamma_{r,k}^{(1)} \triangleq \lambda_r\left(\mathbf{a}_r^{(2)^{\mathrm{T}}}\mathbf{w}_k^{(2)}\right)\left(\mathbf{a}_r^{(3)^{\mathrm{T}}}\mathbf{w}_k^{(3)}\right),\\ \mathbf{y}_k^{(2)} &= \sum_{r=1}^{R}\gamma_{r,k}^{(2)}\,\mathbf{a}_r^{(2)}, \quad \gamma_{r,k}^{(2)} \triangleq \lambda_r\left(\mathbf{a}_r^{(1)^{\mathrm{T}}}\mathbf{w}_k^{(1)}\right)\left(\mathbf{a}_r^{(3)^{\mathrm{T}}}\mathbf{w}_k^{(3)}\right),\\ \mathbf{y}_k^{(3)} &= \sum_{r=1}^{R}\gamma_{r,k}^{(3)}\,\mathbf{a}_r^{(3)}, \quad \gamma_{r,k}^{(3)} \triangleq \lambda_r\left(\mathbf{a}_r^{(1)^{\mathrm{T}}}\mathbf{w}_k^{(1)}\right)\left(\mathbf{a}_r^{(2)^{\mathrm{T}}}\mathbf{w}_k^{(2)}\right).\end{aligned} \tag{7}$$

For a tensor of N dimensions, (6) and (7) can be written as

$$\begin{aligned}\mathbf{y}_k^{(n)} &= \mathcal{X}\times_1 \mathbf{w}_k^{(1)}\cdots\times_{n-1}\mathbf{w}_k^{(n-1)}\times_{n+1}\mathbf{w}_k^{(n+1)}\cdots\times_N \mathbf{w}_k^{(N)}\\ &= \sum_{r=1}^{R}\gamma_{r,k}^{(n)}\,\mathbf{a}_r^{(n)} \in \mathbb{C}^{I_n},\end{aligned} \tag{8}$$

with

$$\gamma_{r,k}^{(n)} \triangleq \lambda_r \prod_{\substack{m=1\\ m\neq n}}^{N}\left(\mathbf{a}_r^{(m)^{\mathrm{T}}}\mathbf{w}_k^{(m)}\right), \tag{9}$$

$k\in\{1,\ldots,K_n\}$, and $n\in\{1,\ldots,N\}$, for an N-th order tensor $\mathcal{X}$.

Taking into account the vectors given in (8), for each mode n, we can construct a sample matrices with K_n column vectors $\{\mathbf{y}_1^{(n)},\ldots,\mathbf{y}_{K_n}^{(n)}\}$ as

$$\mathbf{Y}^{(n)} \triangleq \begin{bmatrix}\mathbf{y}_1^{(n)} & \cdots & \mathbf{y}_{K_n}^{(n)}\end{bmatrix} \in \mathbb{C}^{I_n\times K_n}, \tag{10}$$

where K_n is the rank of $\mathbf{X}^{(n)}$. It is possible to show that

$$\mathbf{Y}^{(n)} = \mathbf{X}^{(n)}\mathbf{\Omega}^{(n)}, \tag{11}$$

with

$$\begin{aligned}\mathbf{\Omega}^{(n)} &\triangleq \bigdiamond_{\substack{m=1\\ m\neq n}}^{N}\mathbf{W}^{(m,n)} \in \mathbb{C}^{I_1\ldots I_{n-1}I_{n+1}\ldots I_N\times K_n},\\ \mathbf{W}^{(m,n)} &\triangleq \begin{bmatrix}\mathbf{w}_1^{(m)} & \cdots & \mathbf{w}_{K_n}^{(m)}\end{bmatrix} \in \mathbb{C}^{I_m\times K_n}.\end{aligned} \tag{12}$$

An orthonormal basis $\mathbf{Q}^{(n)}\in\mathbb{C}^{I_n\times K_n}$ for the range of each n-th mode vector space of $\mathcal{X}$ can be computed from the QR factorization of each $\mathbf{Y}^{(n)}$, which yields

$$\mathbf{Y}^{(n)} = \mathbf{Q}^{(n)}\mathbf{R}^{(n)}, \tag{13}$$

for $n=\{1,\ldots,N\}$.

From N orthonormal bases $\{\mathbf{Q}^{(1)},\ldots,\mathbf{Q}^{(N)}\}$, and since $K_n\leq I_n$, a reduced tensor $\mathcal{B}\in\mathbb{C}^{K_1\times K_2\times\cdots\times K_N}$ can be computed by the following product

$$\mathcal{B} = \mathcal{X}\times_1 \mathbf{Q}^{(1)^{\mathrm{H}}} \times_2 \mathbf{Q}^{(2)^{\mathrm{H}}}\cdots\times_N \mathbf{Q}^{(N)^{\mathrm{H}}}. \tag{14}$$

Since $\mathbf{X}^{(n)}\approx\mathbf{Q}^{(n)}\mathbf{Q}^{(n)^{\mathrm{H}}}\mathbf{X}^{(n)}$, then

$$\mathcal{X} \approx \mathcal{B}\times_1 \mathbf{Q}^{(1)} \times_2 \mathbf{Q}^{(2)}\cdots\times_N \mathbf{Q}^{(N)}. \tag{15}$$

We can calculate the HOSVD of this reduced tensor

$$\mathcal{B} = \mathcal{S}\times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)}\cdots\times_N \mathbf{U}^{(N)}. \tag{16}$$

As shown in [7], this decomposition can be computed from the SVD of the unfolded matrices $\mathbf{B}^{(n)}\in\mathbb{C}^{K_n\times K_1\ldots K_{n-1}K_{n+1}\ldots K_N}$

$$\mathbf{B}^{(n)} = \mathbf{U}^{(n)}\,\mathbf{\Sigma}^{(n)}\,\mathbf{V}^{(n)^{\mathrm{H}}}. \tag{17}$$

The orthogonal matrices $\mathbf{U}^{(n)}\in\mathbb{C}^{K_n\times K_n}$, $n\in\{1,\ldots,N\}$ of the HOSVD in (16) can be directly deduced from (17). The core tensor $\mathcal{S}\in\mathbb{C}^{K_1\times K_2\times\cdots\times K_N}$ can be obtained from its matrix unfoldings, given by

$$\mathbf{S}^{(n)} = \mathbf{\Sigma}^{(n)}\mathbf{V}^{(n)^{\mathrm{H}}}\left(\bigotimes_{\substack{m=1\\ m\neq n}}^{N}\mathbf{U}^{(m)}\right)^{*}. \tag{18}$$

Applying (16) to (15), we can recover the original tensor $\mathcal{X}$

$$\begin{aligned}\mathcal{X} &\approx \mathcal{S}\times_1\left(\mathbf{Q}^{(1)}\mathbf{U}^{(1)}\right)\times_2\left(\mathbf{Q}^{(2)}\mathbf{U}^{(2)}\right)\cdots\times_N\left(\mathbf{Q}^{(N)}\mathbf{U}^{(N)}\right)\\ &\approx \mathcal{S}\times_1\tilde{\mathbf{U}}^{(1)}\times_2\tilde{\mathbf{U}}^{(2)}\cdots\times_N\tilde{\mathbf{U}}^{(N)}.\end{aligned} \tag{19}$$

Observe that (19) can be related with the HOSVD decomposition in (2), in which the orthogonal matrices $\{\tilde{\mathbf{U}}^{(1)},\ldots,\tilde{\mathbf{U}}^{(N)}\}$ are directly obtained from the orthogonal transformation of the columns of $\mathbf{U}^{(n)}$, i.e. $\tilde{\mathbf{U}}^{(n)}\triangleq\mathbf{Q}^{(n)}\mathbf{U}^{(n)}$.

Analogously to the procedure described for the randomized SVD, it is important to ensure that the range of each mode-space of $\mathcal{X}$ is entirely covered. Consequently, we can also consider an oversampling parameter $p>0$ for constructing the sample matrices $\mathbf{Y}^{(n)}\in\mathbb{C}^{I_n\times K_n+p}$ from the random vectors $\{\mathbf{w}_1^{(m)},\ldots,\mathbf{w}_{K_n+p}^{(m)}\}$, $m\in\{1,\ldots,n-1,n+1,\ldots,N\}$.

B. Structured random matrices

In randomized methods, a structured random matrix $\mathbf{W}\in\mathbb{C}^{n\times k}$, such as the *subsampled randomized Fourier transform* (SRFT) [4] or the *subsampled randomized Hadamard transform* (SHRT) [22] among others presented by Liberty [23], can be employed in order to produce an orthogonal basis for the range of an input data $\mathbf{X}\in\mathbb{C}^{m\times n}$, allowing a complexity reduction of the multiplication $\mathbf{Y}=\mathbf{X}\mathbf{W}$.

According to Woolfe *et al.* [1], an SRFT can be constructed as

$$\mathbf{W} = \sqrt{\frac{n}{k}}\,\mathbf{D}\,\mathbf{F}\,\mathbf{E}, \tag{20}$$

where $\mathbf{D}$ is an $n\times n$ diagonal matrix whose complex entries are randomly drawn from the unit circle, $\mathbf{F}$ is the $n\times n$ unitary discrete Fourier transform with $f_{n_1,n_2}=e^{-2\pi j(n_1-1)(n_2-1)/n}$, and $\mathbf{E}\triangleq[\,\hat{\mathbf{e}}_{i_1}\cdots\hat{\mathbf{e}}_{i_k}]$, with $i_1\neq\cdots\neq i_k$ and $i_l\in\{1,\ldots,n\}$, is an operator that randomly selects k of n columns of the product $\mathbf{DF}$.

By employing the SRFT given by (20), we obtain a reduction in the multiplication cost to $\mathcal{O}(mn\log(k))$ as opposed to

Algorithm 1: Randomized HOSVD - SVD via row-extraction.

1. Draw a SRFT matrix $\mathbf{\Omega}^{(n)}$ according to (21), $\forall n \in \{1, ..., N\}$.
2. Compute $\mathbf{Y}^{(n)} = \mathbf{X}^{(n)}\mathbf{\Omega}^{(n)}$, $\forall n \in \{1, ..., N\}$.
3. Compute an orthonormal basis $\mathbf{Y}^{(n)} = \mathbf{Q}^{(n)}\mathbf{R}^{(n)}$, $\forall n \in \{1, ..., N\}$.
4. Initialize $\tilde{\mathbf{B}}^{(1)} = \mathbf{X}^{(1)}$.
5. Compute an ID $\mathbf{Q}^{(n)} = \mathbf{\Lambda}^{(n)}\mathbf{Q}^{(n)}(\gamma, :)$, with $\gamma = [i_1, ..., i_{K_n+p}]$.
6. Extract $\tilde{\mathbf{B}}^{(n)}(\gamma, :)$ and compute a QR factorization of $\left(\tilde{\mathbf{B}}^{(n)}(\gamma, :)\right)^{\mathrm{H}}$: $\tilde{\mathbf{B}}^{(n)}(\gamma, :) = \tilde{\mathbf{R}}^{(n)^{\mathrm{H}}}\tilde{\mathbf{Q}}^{(n)^{\mathrm{H}}}$.
7. Form the product $\mathbf{Z}^{(n)} = \mathbf{\Lambda}^{(n)}\tilde{\mathbf{R}}^{(n)^{\mathrm{H}}}$.
8. Compute an SVD $\mathbf{Z}^{(n)} = \tilde{\mathbf{U}}^{(n)}\tilde{\mathbf{\Sigma}}^{(n)}\tilde{\mathbf{V}}^{(n)^{\mathrm{H}}}$.
9. Compute a reduced tensor $\mathcal{B} = \tilde{\mathcal{B}} \times_n \mathbf{Q}^{(n)^{\mathrm{H}}}$.
10. Update $\tilde{\mathcal{B}} \leftarrow \mathcal{B}$.
11. Repeat step 5 to 10 for $n = 1, ..., N$.
11. Compute $\mathcal{X}^{\text{signal}} = \mathcal{X} \times_1 \tilde{\mathbf{U}}^{(1)}\tilde{\mathbf{U}}^{(1)^{\mathrm{H}}} \cdots \times_N \tilde{\mathbf{U}}^{(N)}\tilde{\mathbf{U}}^{(N)^{\mathrm{H}}}$.

$\mathcal{O}(mnk)$ flops for a standard Gaussian matrix, as proposed by Martinsson *et al.* [3]. Note that the Fourier transform could be replaced by the Walsh-Hadamard transform, as suggested in [1].

There are several possibilities for constructing structured random matrices based on the fast Fourier transform and similar algorithms [4], [22], [23] for the previously mentioned purpose. For the sake of simplicity, let us consider the SRFT given in (20).

Applying (20) to (12), on can show that

$$\mathbf{\Omega}^{(n)} = \sqrt{\frac{I_1 \dots I_N}{K_1 \dots K_N}}\, \bar{\mathbf{D}}^{(n)}\bar{\mathbf{F}}^{(n)}\bar{\mathbf{E}}^{(n)}. \tag{21}$$

C. Higher-order subspace method

According to the higher-order subspace method, it is possible to decompose given input data $\mathcal{X}$ into the sum of a signal and noise subspaces as $\mathcal{X} = \mathcal{X}^{\text{signal}} + \mathcal{X}^{\text{noise}}$.

The signal subspace $\mathcal{X}^{\text{signal}}$ can be constructed using the L_n singular vectors of $\mathbf{U}^{(n)}$ associated with the L_n-th largest singular values of each n-th mode, i.e.

$$\mathcal{X}^{\text{signal}} = \mathcal{X} \times_1 \mathbf{P}^{(1)} \cdots \times_N \mathbf{P}^{(N)}, \tag{22}$$

where $\mathbf{P}^{(n)} \triangleq \tilde{\mathbf{U}}^{(n)}\tilde{\mathbf{U}}^{(n)^{\mathrm{H}}}$, for $n \in \{1, \dots, N\}$, are orthogonal projectors along each n-th mode and $\tilde{\mathbf{U}}^{(n)} \triangleq \left[\mathbf{u}_1^{(n)} \cdots \mathbf{u}_{L_n}^{(n)}\right]$ are matrices containing the first L_n singular vectors.

When the singular values associated with each n-th matrix unfolding decay slowly, the smallest singular values can interfere with the approximation, leading to an approximation of poor quality. A known way to reduce this interference is by taking power q of the matrix to be analyzed, i.e., $\mathbf{X}^{(n)}$ [4].

IV. Simulation Results

In this section, we compare the standard HOSVD, the HOSVD by computing the randomized SVD (via row extraction) and the proposed randomized HOSVD.

For classical algorithms, the SVD of each unfolded matrix $\mathbf{X}^{(n)}$ requires approximately $\mathcal{O}(I_1 \cdots I_N K_n)$ in contrast with $\mathcal{O}\left(I_1 \cdots I_N \log(K_n) + \left(I_n + \prod_{\substack{m=1 \\ m\neq n}}^{N} I_N\right) K_n^2\right)$ flops for the HOSVD computed by the randomized SVDs, referred to as

TABLE I
Computational complexity for computing the singular vectors associated with each n mode.

Method	Computational complexity (flops)
Standard HOSVD	$\mathcal{O}\left(K_n \prod_{m=1}^{N} I_m\right)$
HOSVD-RandSVD	$\mathcal{O}\left(C_n^{(1)} + K_n^2 \prod_{\substack{m=1 \\ m\neq n}}^{N} I_m\right)$
Randomized HOSVD†	$\mathcal{O}\left(C_n^{(1)} + C_n^{(2)} + K_n^2 \prod_{m=1}^{n-1} K_m \prod_{m=n+1}^{N} I_m\right)$
Randomized HOSVD‡	$\mathcal{O}\left(C_n^{(1)} + C_n^{(2)} + K_n \prod_{m=1}^{N} K_m\right)$

† By computing direct HOSVD.
‡ By employing an extension of the row-extraction technique proposed in [4].
$C_n^{(1)} \triangleq \log(K_n) \prod_{m=1}^{N} I_m + K_n^2 I_n$, $C_n^{(2)} \triangleq \prod_{m=1}^{n} K_m \prod_{m=n}^{N} I_m$.

HOSVD-RandSVD. The cost of randomized SVD is reduced thanks to the use of a structured random matrix instead of a general dense matrix $\mathbf{W}^{(n)}$ and the SVD computed via row-extraction [4].

Similarly, the use of a structured random matrix, as proposed in (21), allows us to employ a fast method for computing the product $\mathbf{X}^{(n)}\mathbf{\Omega}^{(n)}$ and for obtaining an orthonormal basis $\mathbf{B}^{(n)}$. We propose two ways to compute the randomized HOSVD: by computing direct HOSVD of the reduced tensor $\mathcal{B}$, and by employing an extension of the row-extraction technique proposed in [4]. The last proposed algorithm is described in Algorithm 1. Table I shows the computation complexity for computing the singular vectors associated with each n mode.

In Figure 1 we give some examples of the impact of design parameters by fixing $K_1 = K_2 = K_3$ and $I_1 = I_2 = I_3$ on the complexity. Thus, the HOSVD by computing the randomized SVD and standard HOSVD provide the lowest and highest complexity, respectively. For these design parameters, we do not observe a vast difference between the randomized HOSVD computed via direct HOSVD and row-extraction.

To evaluate the proposed methods, we consider a $100 \times 100 \times 100$ random tensor $\mathcal{X}$ with $R = 10$, generated by using (3) and corrupted by different levels of noise. The performances associated with the proposed methods are evaluated making use of 50 Monte Carlo simulations. The normalized mean squared error (NMSE), in dB, is regarded in signal

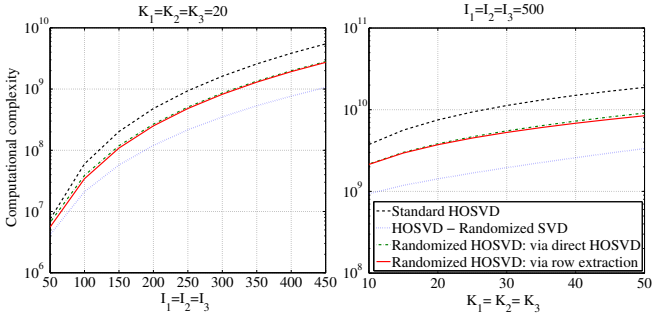

Fig. 1. Computational complexity for an $(I_1 \times I_2 \times I_3)$-tensor.

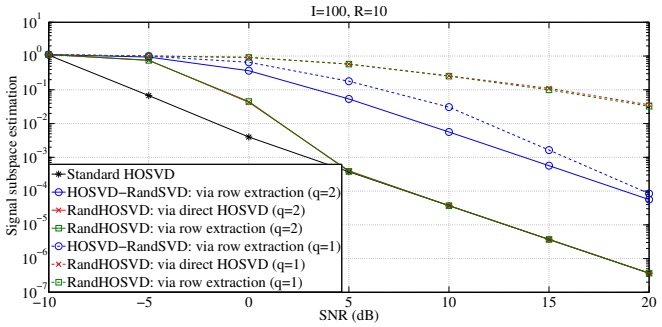

Fig. 2. Approximation of signal subspace estimation for two values of q.

subspace estimation.

From Figure 2, it can be noticed that both randomized HOSVD algorithms, referred to as RandHOSVD, tend to present closer performances. The HOSVD by computing randomized SVDs, referred to as HOSVD-RandSVD, provides better performance than the randomized HOSVD algorithms. However, the randomized HOSVD algorithms can improve the approximation by increasing q near to the standard HOSVD. As expected, the standard HOSVD estimates better the signal subspace, but at the cost of higher computational complexity.

V. Conclusion

In this paper, we have proposed a method for improving dimensionality reduction before computing the HOSVD, motivated by applications involving high-dimensional data in which only a few sources have significant contributions. It is a generalization of the randomized SVD for higher-order tensors. We present two different ways to employ the randomized HOSVD and compare with the standard HOSVD and HOSVD by computing the randomized SVDs concerning the NMSE of signal subspace estimation and computation complexity taking the most onerous operations.

The greatest advantage of all proposed methods is to provide lower dimensionality representation and computational complexity. The randomized HOSVD algorithms provide an interesting tradeoff between the NMSE performance and computational complexity. The randomized HOSVD method can even be improved by conveniently adjusting the structured random matrices. It can also be jointly employed with a better truncation strategy for computing the HOSVD for the purpose of reducing the number of required operations and of improving the estimation error. Furthermore, our method can be applied in a pre-processing step for dimensionality reduction, which can lead to an improvement of the existing adaptive methods for low-rank approximations.

Acknowledgment

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil, under Grant 2014/23936-4. The authors thank Prof. Romis Attux (UNICAMP, Brazil) for the discussions and valuable suggestions.

References

[1] F. Woolfe, E. Liberty, V. Rokhlin, and M. Tygert, "A fast randomized algorithm for the approximation of matrices," *Applied and Computational Harmonic Analysis*, vol. 25, no. 3, pp. 335–366, 2008.

[2] V. Rokhlin, A. Szlam, and M. Tygert, "A randomized algorithm for principal component analysis," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1100–1124, 2009.

[3] P. G. Martinsson, V. Rokhlin, and M. Tygert, "A randomized algorithm for the decomposition of matrices," *Applied and Computational Harmonic Analysis*, vol. 30, pp. 47–68, 2011.

[4] N. P. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Review*, vol. 53, no. 2, pp. 217–288, 2011.

[5] A. Cichocki, D. Mandic, A.-H. Phan, C. Caiafa, G. Zhou, Q. Zhao, and L. D. Lathauwer, "Tensor decompositions for signal processing applications from two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2014.

[6] A. Cichocki, "Era of Big Data Processing: A new approach via tensor networks and tensor decompositions," *A Computing Research Repository (CoRR)*, vol. abs/1403.2048, 2014. [Online]. Available: http://arxiv.org/abs/1403.2048

[7] L. D. Lathauwer, "Signal processing based on multilinear algebra." Ph.D. dissertation, Universit Catholique de Leuven, Netherlands, 1997.

[8] L. Tucker, "Some mathematical notes of three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[9] L. de Lathauwer, B. de Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[10] N. L. Bihan and G. Ginolhac, "Three-mode data set analysis using higher order subspace method: application to sonar and seismo-acoustic signal processing," *Signal Processing*, vol. 84, no. 5, pp. 919–942, 2004.

[11] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and its Applications*, vol. 18, pp. 111–117, 1977.

[12] P. Comon and B. Mourrain, "Decomposition of quantics in sums of powers of linear forms," *Signal Processing*, vol. 53, pp. 93–108, 1996, special Issue on Higher Order Statistics.

[13] L. de Lathauwer, B. de Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 2, pp. 295–327, 2004.

[14] P. Comon, J. M. F. Ten Berge, L. de Lathauwer, and J. Castaing, "Generic and typical ranks of multi-way arrays," *Linear Algebra and its Applications*, vol. 430, no. 11, pp. 2997–3007, 2009.

[15] L. de Lathauwer, B. de Moor, and J. Vandewalle, "On the best rank-1 and rank-$(R_1, R_2, ..., R_n)$ approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

[16] M. Ishteva, P.-A. Absil, S. V. Huffel, and L. D. Lathauwer, *Recent Advances in Optimization and its Applications in Engineering*, 1st ed. Springer-Verlag Berlin Heidelberg, November 2010, ch. On the best low multilinear rank approximation of higher-order tensors, pp. 145–164.

[17] B. Savas and L.-H. Lim, "Quasi-Newton methods on Grassmannians and multilinear approximations of tensors," *SIAM Journal on Scientific Computing*, vol. 32, no. 6, pp. 3352–3393, 2010.

[18] M. Ishteva, P.-A. Absil, S. V. Huffel, and L. D. Lathauwer, "Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme," *SIAM Journal on Matrix Analysis and Applications*, vol. 32, no. 1, pp. 115–135, 2011.

[19] N. Vannieuwenhoven, R. Vandebril, and K. Meerbergen, "A New Truncation Strategy for the Higher-Order Singular Value Decomposition," *SIAM Journal on Scientific Computing*, vol. 34, no. 2, pp. A1027–A1052, 2012.

[20] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, no. 35, pp. 283–319, 1970.

[21] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[22] F. Woolfe, E. Liberty, V. Rokhlin, and M. Tygert, "A fast randomized algorithm for the approximation of matrices," Yale University, Department of Computer Science, Preliminary report, April 2007.

[23] E. Liberty, "Accelerated dense random projections." Ph.D. dissertation, Yale University, New Haven, CT, 2009.