ADAPTIVE MULTILINEAR SVD FOR STRUCTURED TENSORS
Remy Boyer, Roland Badeau

To cite this version:
Remy Boyer, Roland Badeau. ADAPTIVE MULTILINEAR SVD FOR STRUCTURED TENSORS. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’06), 2006, Toulouse, France. hal-00577274

HAL Id: hal-00577274
https://hal.archives-ouvertes.fr/hal-00577274
Submitted on 17 Mar 2011

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.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License
The Higher-Order SVD (HOSVD) is a generalization of the SVD to higher-order tensors (i.e., arrays with more than two indexes) and plays an important role in various domains. Unfortunately, the computational cost of this decomposition is very high since the basic HOSVD algorithm involves the computation of the SVD of three highly redundant block-Hankel matrices, called modes. In this paper, we present an ultra-fast way of computing the HOSVD of a third-order structured tensor. The key result of this work lies in the fact it is possible to reduce the basic HOSVD algorithm to the computation of the SVD of three non-redundant Hankel matrices whose columns are multiplied by a given weighting function. Next, we exploit an FFT-based implementation of the orthogonal iteration algorithm in an adaptive way. Even though for square \((I \times I \times I)\) tensors the complexity of the basic full-HOSVD is \(O(I^4)\) and \(O(I^3)\) for its \(r\)-truncated version, our approach reaches a linear complexity of \(O(rI\log_2(I))\).

1. INTRODUCTION

The subject of multilinear decomposition is now mature [1, 2]. There are essentially two families. The first one is known under the name of CANDECOMP/PARAFAC (CANonical DECOMposition or PARAAllel FACTors model) and was independently proposed in [6, 7]. This decomposition is very useful in several applications and is related to the tensor rank [9]. The second one is related to the multitudimensional rank [8] and is known under the name of Tucker decomposition [5]. This decomposition is a more general form which is often used. Orthogonality constraints are not required in the general Tucker decomposition but if needed one can refer to the Higher-Order Singular Value Decomposition (HOSVD) [8] or multilinear SVD.

The HOSVD is a generalization of the SVD to higher-order tensors (i.e., arrays with more than two indexes). This decomposition plays an important role in various domains, such as harmonic retrieval [3], image processing, telecommunication, biomedical applications (magnetic resonance imaging and electrocardiography), web search [13], computer facial recognition [11], handwriting analysis [12], statistical analysis by Independent Component Analysis (ICA) [8].

In [8], it has been shown that the HOSVD of a third-order tensor involves the computation of the SVD of three matrices called modes. As a consequence, the computational cost of this algorithm is very high. In the case of structured tensors, these modes are highly redundant matrices (i.e., many columns are repeated). In this paper, we propose a ultra-fast implementation of the \(r\)-truncated HOSVD for structured tensors which takes the redundant structure of each mode into account. The key result of this work lies in the fact that it is possible to reduce the basic HOSVD algorithm to the computation of the SVD of three non-redundant (all columns are different) Hankel matrices whose columns are multiplied by a given weighting function. Next, we exploit the orthogonal iteration algorithm in an adaptive way. Even though for square \((I \times I \times I)\) tensors the complexity of the basic full-HOSVD is \(O(I^4)\) and \(O(I^3)\) for its \(r\)-truncated version, our approach reaches a linear complexity of \(O(rI\log_2(I))\).

2. PRELIMINARIES IN MULTILINEAR ALGEBRA

2.1. Mode of a tensor

There are several ways to represent a tensor. All-orthogonality means that the matrices \(S_{i}^2\) are orthogonal for all possible values of \(s\). The Frobenius-norms of these matrices are called column rank (or 1-mode rank) \(R_i\). The second one is related to the multitudimensional rank [8] and is known under the name of Tucker decomposition [5]. This decomposition is a more general form which is often used. Orthogonality constraints are not required in the general Tucker decomposition but if needed one can refer to the Higher-Order Singular Value Decomposition (HOSVD) [8] or multilinear SVD.

The HOSVD is a generalization of the SVD to higher-order tensors (i.e., arrays with more than two indexes). This decomposition plays an important role in various domains, such as harmonic retrieval [3], image processing, telecommunication, biomedical applications (magnetic resonance imaging and electrocardiography), web search [13], computer facial recognition [11], handwriting analysis [12], statistical analysis by Independent Component Analysis (ICA) [8].

In [8], it has been shown that the HOSVD of a third-order tensor involves the computation of the SVD of three matrices called modes. As a consequence, the computational cost of this algorithm is very high. In the case of structured tensors, these modes are highly redundant matrices (i.e., many columns are repeated). In this paper, we propose a ultra-fast implementation of the \(r\)-truncated HOSVD for structured tensors which takes the redundant structure of each mode into account. The key result of this work lies in the fact that it is possible to reduce the basic HOSVD algorithm to the computation of the SVD of three non-redundant (all columns are different) Hankel matrices whose columns are multiplied by a given weighting function. Next, we exploit the orthogonal iteration algorithm in an adaptive way. Even though for square \((I \times I \times I)\) tensors the complexity of the basic full-HOSVD is \(O(I^4)\) and \(O(I^3)\) for its \(r\)-truncated version, our approach reaches a linear complexity of \(O(rI\log_2(I))\).

2.2. Multilinear SVD (HOSVD)

Theorem 1 (Third-Order SVD [5, 8]) Every \(I_1 \times I_2 \times I_3\) tensor \(A\) can be written as the product:

\[
A = S \times U^{(1)} \times U^{(2)} \times U^{(3)}
\]

in which \(\times\) represents the Tucker s-mode product [8], \(U^{(s)}\) is an unitary \(I_s \times I_s\) matrix and \(S\) is an all-orthogonal and ordered \(I_1 \times I_2 \times I_3\) tensor. All-orthogonality means that the matrices \(S_{i_s} = \alpha_i\), obtained by fixing the \(s\)-th index to \(i\), are mutually orthogonal w.r.t. the standard inner product. Ordering means that \(\|S_{i_1} - i_1\| \geq \|S_{i_2} - i_2\| \geq \ldots \geq \|S_{i_{I_s}} - I_s\| \geq 0\) for all possible values of \(s\). The Frobenius-norms...
of a tensor, corresponding to the nonzero \( s \)-values in (1)–(3). The values of this matrix unfolding. Note that the columns of \( U \) can be computed, which can be achieved faster than a complete SVD (e.g. Golub-Reinsch or R-SVD). Here, only the left term of the SVD needs to be computed, which can be achieved faster than a complete SVD (i.e. with a lower value of \( k_l \)). Besides, the matrix \( A_1 \) has \( n = I_1 \) rows and \( l = I_2 I_3 \) columns. Assuming that \( I_2 I_3 \) is often greater than \( I_1 \), the computational cost of the SVD of \( A_1 \) is \( k_l I_2^2 I_3 l \) flops.

The computational costs of the full HOSVD are summarized in Table 1. In particular, the maximal complexity over all values of \( I_1, I_2, I_3 \) satisfying \( N = I_1 + I_2 + I_3 = 1 \) is obtained for square tensors \((I_1 = I_2 = I_3 = I)\) and equals \((3k_f + 6)I^4\).

### 3.2. Computation of the truncated HOSVD

In many applications, we are interested in computing the HOSVD truncated at orders \((r_1, r_2, r_3)\) (\( r_i \) is often supposed to be much lower than \( I_i \)). Let \( r = \frac{1}{3}(r_1 + r_2 + r_3) \). This truncated HOSVD can be computed in the same way as the full HOSVD, except that the three SVDs involved in its computation are truncated at orders \( r_i \). Besides, the truncated SVD of an \( n \times I \) matrix can be computed faster than its full SVD, by means of the orthogonal iteration method [4, pp. 410–411] for instance. The average computational cost of this truncated SVD is \( k_r r n l \), where the constant \( k_r \) depends on the algorithm (\( k_r \) is generally greater than \( k_l \)). Therefore, applied to matrix \( A_1 \), its cost is \( O(r^3 I^3) \) flops for square tensors. Note that we assume \((1) I_1 \geqslant \log_2(I_1) \) and \( (2) \log_2(I_1) \) is the same order as \( r_1 \). The computational costs of the \( r \)-truncated HOSVD are summarized in Table 1.

### 4. FAST ALGORITHMS FOR HANKEL-STRUCTURED TENSORS

This section is dedicated to Hankel-structured tensors but other structures as for instance circulant matrices can be exploited in a similar way.

#### 4.1. Definition of an Hankel-structured tensor

A third-order \( l_1 \times I_2 \times I_3 \) Hankel-structured tensor is defined according to:

\[
[A]_{i,i_2,i_3} = x[i_1 + i_2 + i_3]
\]

where \( x[n] \) is the \( n \)-th sample of the analyzed signal and \( i_s \in [0 : I_s - 1] \) for \( s \in [1 : 3] \).

#### 4.2. Compressed and windowed modes

The structure of the considered tensor induces a strong redundancy in its modes or in other words, there are several columns in the modes which are repeated. Consequently, we introduce the compressed \( s \)-mode, denoted by \( H_s \) in the following manner:

\[
H_s = A_s I_s
\]

where \( J_s \) is a linear transformation which associates a redundant block-Hankel matrix \( A_s \) to a non-redundant Hankel matrix \( H_s \). For instance, we give the following example for the \( 4 \times 5 \times 3 \) 1-mode: 

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
2 & 3 & 4 & 5 & 6 \\
2 & 3 & 4 & 5 & 6 \\
2 & 3 & 4 & 5 & 6 \\
2 & 3 & 4 & 5 & 6 \\
2 & 3 & 4 & 5 & 6
\end{bmatrix}
\]
where for simplicity we have denoted \( x[n] \) by its index \( n \). In addition, we have underlined the common columns. Consequently, the matrix \( J_x \) is determined so as to produce the following "compressed" 1-mode:

\[
\begin{pmatrix}
0 & 1 & 2 & 3 & 4 & 5 \\
1 & 2 & 3 & 4 & 5 & 6 \\
2 & 3 & 4 & 5 & 6 & 7 \\
3 & 4 & 5 & 6 & 7 & 8
\end{pmatrix}
\] (10)

Remark the Hankel structure of matrix (10). If we note \( \mathbf{n}(\cdot) \) the number of columns, it comes:

\[
\mathbf{n}(H_x) = \sum_{s \neq j} l_s - 1 < \mathbf{n}(A_s) = \prod l_s
\] (11)

and thus \( H \) has less columns than \( A \). Decreasing the column dimension of the mode modifies the dominant singular space, denoted by \( R(\cdot) \). In other words, we have \( \mathbf{R}(H_x) \neq \mathbf{R}(A_s) \). So, for a given non-deficient matrix, \( D_j \), we have to satisfy:

\[
\mathbf{R}(H(D_j)) = \mathbf{R}(A_s).
\] (12)

Toward this end, consider the correlation matrix of the \( s \)-mode:

\[
[C^{(s)}]_{k,l} = \sum_{i=0}^{k-I_s-1} \mathbf{g}^T[i]k_i,j_i \mathbf{H}^s[j_i,j_i]
\] (13)

It can be noticed that some terms are redundant in this last equation. In order to remove this redundancy, we obtain after some derivations:

\[
[C^{(s)}]_{k,l} = \sum_{i=0}^{I_s+I_2-2} d_{(k,l)}[i+1] x[k+i] x^*[l+i]
\]

where:

\[
d_{(k,l)}[i] = \begin{cases}
  i & \text{if } 1 \leq i < \min(l_2,l_3) \\
  \min(l_2,l_3) & \text{if } \min(l_2,l_3) \leq i < \max(l_2,l_3) \\
  l_2 + l_3 - i & \text{if } \max(l_2,l_3) \leq i < l_2 + l_3 \\
  0 & \text{elsewise.}
\end{cases}
\]

The weighting function \( i \rightarrow d_{(k,l)}[i] \) is plotted in figure 1.

![Fig. 1. Weighting function \( d_{(k,l)}[i] \)](image)

It represents the number of times each column of the 1-mode is repeated. For instance, in the example of matrix (9), the weights are \( \{1,2,3,3,2,1\} \).

The correlation matrix of the 1-mode can then be rewritten as a function of the compressed 1-mode \( H_1 \):

\[
[C^{(1)}]_{k,l} = \sum_{i=0}^{K-I_1-1} d_{(k,l)}[i+1] [H_1]_{k,j} [H_1^*]_{l,j}.
\]

Then define the \( (N-I_1) \times (N-I_1) \) diagonal matrix \( D_1 \), whose diagonal coefficients are \([D_1]_{i,i} = \sqrt{d_{(l_1,l_1)}[i]}\). We obtain:

\[
C^{(1)} = H_1 D_1^T H_1^T.
\] (13)

Consequently, the left singular vectors of matrix \( A_1 \) can be obtained as the left singular vectors of matrix \( H_1 D_1 \) and thus \( \mathbf{R}(H(D_1)) = \mathbf{R}(A_1) \). In a similar way, we can build \( D_2 \) and \( D_3 \) such as \( \mathbf{R}(H(D_2)) = \mathbf{R}(A_2) \) and \( \mathbf{R}(H(D_3)) = \mathbf{R}(A_3) \). Obviously, we have:

\[
\mathbf{n}(H(D_s)) = \mathbf{n}(H) < \mathbf{n}(A_s).
\] (14)

Thus the SVD of \( H(D_s) \) can be computed much faster than that of \( A \). The fast algorithm for computing the full or truncated HOSVD of tensor \( \mathbf{H} \) is summarized in table 2.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Full cost</th>
<th>Truncated cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD of ( H_1 ) ( D_1 )</td>
<td>( k_1 r_1 I_1^2 )</td>
<td>( k_1 r_1 I_1^2 )</td>
</tr>
<tr>
<td>SVD of ( H_2 ) ( D_2 )</td>
<td>( k_1 r_1 I_1^2 )</td>
<td>( k_1 r_1 I_1^2 )</td>
</tr>
<tr>
<td>SVD of ( H_3 ) ( D_3 )</td>
<td>( k_1 r_1 I_1^2 )</td>
<td>( k_1 r_1 I_1^2 )</td>
</tr>
<tr>
<td>Tucker product</td>
<td>( 6rN )</td>
<td>( 6rN )</td>
</tr>
</tbody>
</table>

It can be noticed that the most computationally demanding step is now the Tucker product\(^1\). If tensor \( S \) does not have to be computed, then the complexity of the full HOSVD fast algorithm is always lower than \( 6k_0 r N^2 \), and that of the truncated HOSVD fast algorithm is lower than \( 6k_0 r N r^2 \). So, the compression and weighting of the modes allow a reduction of the complexity of one order of magnitude with respect to the basic HOSVD and the r-truncated HOSVD.

4.3. Implementation with Fast Fourier Transforms

To further reduce the complexity, we use the fact that the product between a Hankel matrix and a vector can be efficiently computed by means of Fast Fourier Transforms (FFT) [4, pp. 201–202].

By introducing those fast products into the orthogonal iteration method, the cost of the full SVD of a \( n \times l \) matrix with \( n < l \) and \( N = n + l - 1 \) is reduced to \( k_1 n r N \log_2(N) + k_2 n r^2 \), where the constant \( k_1 \) is related to the Hankel matrix / vector products, and the constant \( k_2 \) is related to the QR factorizations (see [4, pp. 410–411] for more details). In the same way, the cost of a rank-\( r \) truncated SVD is \( k_1 r N \log_2(N) + k_2 n r^2 \).

The ultra-fast algorithm for computing the full or truncated HOSVD of tensor \( \mathbf{H} \) is summarized in table 3. Again, it can be noticed that the most computationally demanding step is the Tucker product\(^2\). If tensor \( S \) does not have to be computed, then the complexity of the full HOSVD fast algorithm is always lower than \( 9k_1 r N^2 \log_2 l + 3k_2 r^2 l \), and that of the truncated HOSVD fast algorithm is lower than \( 9k_1 r N \log_2 l + 3k_2 r^2 l \). Therefore we achieve one more order of magnitude of complexity reduction.

---

\(^1\) In fact, the redundancy in \( \mathbf{H} \) implies that the Tucker product involves several times the same matrix/vector products. Therefore it can be computed faster than in the non-structured case. However this does not lead to a significant reduction of the complexity either.

\(^2\)The redundancy in \( \mathbf{H} \) also implies that the Tucker product involves several times the same Hankel matrix / vector products. Therefore it can be computed even faster than in the fast algorithm. Unfortunately, this does not lead to a significant reduction of the complexity either.
5. ADAPTIVE SIGNAL PROCESSING

In this section, tensor $\mathcal{H}(t)$ is supposed to have slow time variations, and our objective is to efficiently update its HOSVD.

A very classical approach for tracking the SVD of a time-varying matrix consists in interlacing the update of the data with one or a few steps of a standard SVD algorithm, such as the orthogonal iteration method [10]. Following this idea, an efficient way of updating $U((t))$ consists in replacing the exact SVD of $A((t))$ in table 1 by one step of the orthogonal iteration method:

- $Z((t)) = A_{((t))} A_{((t))}^T U((t)) (t - 1)$
- $U((t)) R((t)) = Z((t))$

The first operation is a matrix product; the second one is a QR-factorization. In the case of the full HOSVD, the QR factorization can be computed by means of the Householder QR method [4, pp. 224–225], or the Fast Givens QR method [4, pp. 228–229]. Its cost is $4r^3$ flops. In the case of the truncated HOSVD, the fastest approach is the Fast Givens QR method, whose cost is $2r^2$ flops. The overall cost of the sequential HOSVD algorithm obtained in this way is of the same order as that of the basic HOSVD algorithm, with a smaller multiplicative constant.

According to table 4, we can say that for square tensors the computational cost of the basic HOSVD is $O(r^3)$ and that of the r-truncated version is $O(r^3)$, whereas the ultra-fast truncated HOSVD derived in the same way has a final complexity of $O(r^2 \log r (I))$ (since we consider that $270 \log r (I) \gg 6r$).

6. NUMERICAL SIMULATIONS

In this part, we represent on figure 2 the complexity of the full Basic HOSVD, the r-truncated HOSVD, the fast (with compressed and windowed modes) r-truncated HOSVD and the sequential ultra-fast r-truncated HOSVD. The methodology is the following: we fix the two first dimensions of the tensor and we vary the last one.

As we can see a large complexity gain can be achieved. We note that the most important part of the complexity gain is obtained through the SVD of the compressed-windowed Hankel-structured modes. In addition, the adaptive/sequential approach is very attractive for tensors of very large dimensions.