Higher order SVD: theory and algorithms

Chia-Tung Kuo
tomkuo@ucdavis.edu

March 20, 2013

1 Abstract

In this report we describe a generalization of SVD in higher order (called HOSVD, short for higher order SVD), and discuss some of the well known properties of matrix SVD and compare them to their higher order counterparts. The theorem of the existence of HOSVD naturally leads to a closed form straightforward algorithm to compute it. We also present small numerical examples of order three tensors to illustrate the properties discussed.

2 Introduction/motivation

Singular value decomposition (SVD) has been studied and used extensively in diverse fields. Its properties, such as uniqueness, best low-rank approximation, etc, lend itself to many applications in scientific computing; e.g. image analysis, clustering, to name a few. Recently an increasing number of problems in different domains involve manipulations of higher dimensional data, and tensors are a natural generalization of matrices into higher dimensions.

An order N tensor is simply a N dimensional array in which each element is indexed by a N-tuple $(i_1, i_2, \ldots, i_N)$. Due to the usefulness of matrix SVD, it is natural to look for a similar higher order decomposition with similar properties. In this report, we present a proper generalization of such higher order SVD and discuss some of its properties in relation to the normal matrix SVD. It is worth mentioning that other SVD generalizations exist, and one could choose whichever to use based on which matrix SVD property one wants to carry over to higher dimension. The HOSVD we present here is probably one bearing the most resemblance in decomposition form to matrix SVD, and also possess strikingly analogous properties.

The report is organized as follows. Section 3 introduces the necessary definitions of some higher order operations analogous to the operations on matrices, and states the main theorem of HOSVD. The algorithm to compute such HOSVD is derived in section 4. In section 5, we show some properties of HOSVD analogous to the matrix SVD. After that, we demonstrate each property with a small numerical example in section 6. Section 7 then concludes this report.

3 Basic concepts and theory

Throughout this report, we use calligraphic letters to denote higher order tensors, normal capital letters for matrices, and lower case letters for scalars. Also for the ease of presentation and understanding, though definitions and theorems are stated for order N tensors, numerical examples and visualizations are given with order three tensors. Table 1 serves as a reference for the notations we use throughout this report.

$\mathcal{A}_{(n)}$	n-mode unfolding
$\|\|\mathcal{A}\|\|$	Frobenius norm ($\sqrt{\sum_{i_1,\dots,i_N} a^2_{i_1,\dots,i_N}}$)
$\sigma_i^{(n)}$	i-th singular value on mode n
$\times_n$	n-mode product
$\otimes$	Kronecker product

Table 1: Notations

We start with the matrix representations of a tensor, called unfolding, or matricization. This is useful when we consider the generalizations of left and right singular vectors. The basic idea is to fix one chosen mode and arrange all "slices" of other dimensions into columns of the matrix. The formal definition is given in Definition 1 in terms of the elements, but it is easier to understand with a small example. Let $\mathcal{A} \in \mathbb{R}^{3\times4\times2}$ have the following two frontal slices

$$\mathcal{A}_1 = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix} \text{ and } \mathcal{A}_2 = \begin{bmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{bmatrix}$$

then the unfoldings corresponding to each mode are

$$\mathcal{A}_{(1)} = \begin{bmatrix} 1 & 4 & 7 & 10 & 13 & 16 & 19 & 22 \\ 2 & 5 & 8 & 11 & 14 & 17 & 20 & 23 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 \end{bmatrix}$$

$$\mathcal{A}_{(2)} = \begin{bmatrix} 1 & 2 & 3 & 13 & 14 & 15 \\ 4 & 5 & 6 & 16 & 17 & 18 \\ 7 & 8 & 9 & 19 & 20 & 21 \\ 10 & 11 & 12 & 22 & 23 & 24 \end{bmatrix}$$

$$\mathcal{A}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & \dots & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & 17 & \dots & 21 & 22 & 23 & 24 \end{bmatrix}$$

Definition 1. *Let $\mathcal{A} \in \mathbb{C}^{I_1\times I_2\times\cdots\times I_N}$ be an order N tensor. The n-mode matrix unfolding (also called matricization) $\mathcal{A}_{(n)} \in \mathbb{C}^{I_n\times(I_{n+1}\times\cdots\times I_N\times I_1\times\cdots\times I_{n-1})}$ has tensor element $a_{i_1,i_2,\dots,i_N}$ at index (i_n, j) where*

$$j = 1 + \sum_{k=1,k\neq n}^{N} (i_k - 1)J_k \qquad \texttt{with} \qquad J_k = \prod_{m=1,m\neq n}^{N} I_m$$

Now we wish to define a higher order multiplication analogous to the multiplication on matrices. Let us look at the matrix product $G = UFV^H$. We note that the right multiplication by V^H is in fact very similar to the left multiplication by U, except it is multiplied to a different mode of F. While the columns of F are multiplied by U, the rows of F are multiplied by V in the same way. The generalization of this concept leads to the following definition for product in higher dimension.

Definition 2. *The n-mode product of tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ with matrix $U \in \mathbb{C}^{I_n \times J}$ is of size $I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$, and defined element-wise as*

$$(\mathcal{A} \times_n U)_{i_1 \ldots i_{n-1} j i_{n+1} \ldots i_N} = \sum_{i_n=1}^{I_N} a_{i_1 i_2 \ldots i_N} u_{j i_n}$$

A better characterization is that each mode-n column vector (by fixing indices for all other modes) is multiplied by matrix U; in matricized form we have

$$\mathcal{Y} = \mathcal{A} \times_n U \iff \mathcal{Y}_{(n)} = U \mathcal{A}_{(n)} \tag{1}$$

Figure 1 is a visualization of the equation $\mathcal{A} = \mathcal{B} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)}$. Multiplication of $\mathcal{B}$ with $U^{(1)}$ means that every column of $\mathcal{B}$ (fixing indices i_1 and i_2) is multiplied by $U^{(1)}$, and similarly for multiplications with $U^{(2)}$ and $U^{(3)}$. A crucial difference to the normal matrix multiplication is that this operation is clearly commutative since we directly specify which mode a matrix is multiplied onto. In the case of matrix multiplication above $G = UFV^H$ would be written as $G = F \times_1 U \times_2 V$.

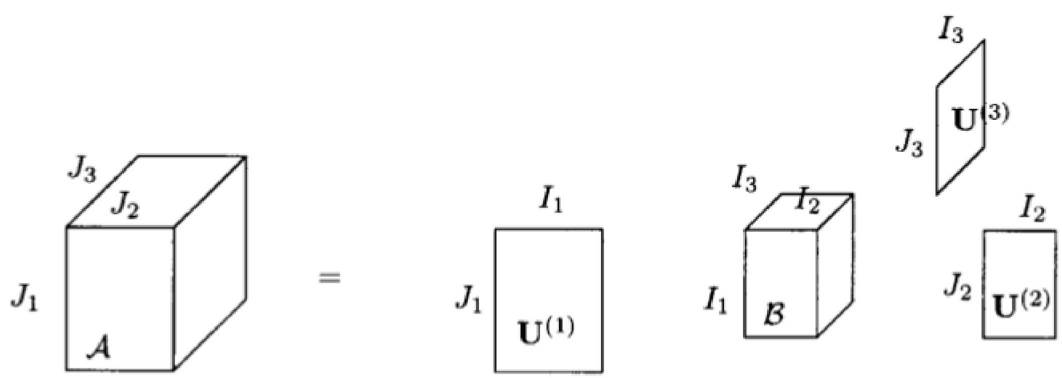

Figure 1: Visualization of the multiplications of order three tensor $\mathcal{B}$ with matrices $U^{(1)}, U^{(2)}$ and $U^{(3)}$ on each mode.

Here we introduce one more definition that enables us to write HOSVD in unfolded forms in a compact way.

Definition 3. *The Kronecker product of two matrices $A \in \mathbb{C}^{I \times J}$ and $B \in \mathbb{C}^{K \times L}$, denoted* *by $A \otimes B$, is a matrix of size $(IK) \times (JL)$ defined by*

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1J}B \\ a_{21}B & a_{22}B & \dots & a_{2J}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}B & a_{I2}B & \dots & a_{IJ}B \end{bmatrix}$$

Now we are ready to state the main theorem:

Theorem 1. *A $(I_1 \times I_2 \times \dots \times I_N)$-tensor $\mathcal{A}$ can be written as the product*

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

in which

1. *$U^{(n)}$ is a unitary $I_n \times I_n$ matrix*
2. *$\mathcal{S}$ is a $(I_1 \times I_2 \times \dots \times I_N)$-tensor of which the $(N-1)$-th order subtensor $\mathcal{S}_{i_n=\alpha}$ obtained by fixing the n-th index to α, have the following properties*
   (a) *all-orthogonality: for all n, subtensors $\mathcal{S}_{i_n=\alpha}$ and $\mathcal{S}_{i_n=\beta}$ are orthogonal for all α and β with $\alpha \neq \beta$.*
   (b) *ordering: for all n,*

$$||\mathcal{S}_{i_n=1}|| \geq ||\mathcal{S}_{i_n=2}|| \geq \dots \geq ||\mathcal{S}_{i_n=I_n}|| \geq 0 \tag{3}$$

   The Frobenius norm $||\mathcal{S}_{i_n=i}||$, denoted by $\sigma_i^{(n)}$, are the n-mode singular values of $\mathcal{A}$ and the columns of $U_i^{(n)}$ are the n-mode singular vectors.

In terms of an order three tensor $\mathcal{A}$, Theorem 1 states that these exist orthogonal (unitary) transformations ($U^{(n)}$) for each mode such that the core tensor $\mathcal{S} = \mathcal{A} \times_1 {U^{(1)}}^T \times_2 {U^{(2)}}^T \times_3 {U^{(3)}}^T$ is all-orthogonal and ordered. All-orthogonality means that the "horizontal slices" of $\mathcal{S}$ (fixing i_1 constant) are mutually orthogonal with respect to the inner product. The ordering condition is simply a convention that fixes a particular ordering of the columns of $U^{(n)}$, much like in the case of matrix SVD.

4 HOSVD Algorithm

In this section, we derive the algorithm to actually compute HOSVD. It is indeed a straightforward consequence of Theorem 1.

By equation (2) and the commutative property of n-mode product, we can easily obtain the following relation

$$\mathcal{S} = \mathcal{A} \times_1 {U^{(1)}}^H \cdots \times_N {U^{(N)}}^H \tag{4}$$

Repeated application of the relation (1), together with the definitions of Kronecker product leads to the n-mode unfolding of $\mathcal{A}$ as

$$\mathcal{A}_{(n)} = U^{(n)} \mathcal{S}_{(n)} (U^{(n+1)} \otimes \cdots \otimes U^{(N)} \otimes U^{(1)} \otimes \cdots \otimes (U^{(n-1)})^H \tag{5}$$

The all-orthogonality and ordering properties of $\mathcal{S}$ imply that $\mathcal{S}_{(n)}$ has mutually orthogonal rows with Frobenius norms equal to $\sigma_1^{(n)}, \sigma_2^{(n)}, \ldots, \sigma_{I_n}^{(n)}$.

Notice that since each $U^{(n)}$ is orthogonal, by the definition of Kronecker product, $U^{(n+1)} \otimes \cdots \otimes U^{(N)} \otimes U^{(1)} \otimes \cdots \otimes (U^{(n-1)}$ is also orthogonal. Therefore, we define

$$\Sigma^{(n)} = \mathtt{diag}(\sigma_1^{(n)}, \sigma_2^{(n)}, \ldots, \sigma_{I_n}^{(n)}) \tag{6}$$

and columnwise orthonormal V as

$$V^{(n)} = (U^{(n+1)} \otimes \cdots \otimes U^{(N)} \otimes U^{(1)} \otimes \cdots \otimes (U^{(n-1)}) \tilde{\mathcal{S}}_{(n)}^H \tag{7}$$

where $\tilde{\mathcal{S}}_{(n)}$ is the normalized version of $\mathcal{S}_{(n)}$ with each row scaled to unit norm (i.e. $\mathcal{S}_{(n)} = \Sigma^{(n)} \tilde{\mathcal{S}}_{(n)}$)

Putting (5), (6), and (7) together, we reach precisely the matrix SVD for the unfolding matrix

$$\mathcal{A}_{(n)} = U^{(n)} \Sigma^{(n)} {V^{(n)}}^H \tag{8}$$

Equations (4) and (8) lead to the following intuitive and straightforward algorithm outlined in Figure 2.

Algorithm 1 Procedure for computing HOSVD

1: **for** $n = 1, \ldots, N$ **do**
2: Compute $\mathcal{A}_{(n)} = U^{(n)} \Sigma^{(n)} {V^{(n)}}^H$
3: **end for**
4: $\mathcal{S} = \mathcal{A} \times_1 {U^{(1)}}^H \cdots \times_N {U^{(N)}}^H$
5: **return** $\mathcal{S}, U^{(1)}, \ldots, U^{(N)}$

Figure 2: Algorithm for computing HOSVD

5 Properties

In this section, we discuss some selected properties of the matrix SVD and their higher order counterparts.

Property 1. *Let the HOSVD of $\mathcal{A}$ be given as in Theorem 1, and let r_n be the largest index for which $||\mathcal{S}_{i_n=r_n}|| > 0$ in (3). Then the n-mode rank is*

$$R_n = \boldsymbol{rank}_n(\mathcal{A}) \equiv \boldsymbol{rank}(\mathcal{A}_{(n)}) = r_n$$

This result is analogous to the property in matrix SVD where the number of nonzero singular values is the rank of the matrix. It is, however, crucial to keep in mind that unlike the rank of a matrix, the n-rank (defined as $\text{rank}(\mathcal{A}_{(n)})$) is in general different for each n, and these numbers do not represent the number of rank-one tensors needed to construct $\mathcal{A}$. In fact the rank of a tensor (not n-rank) is precisely defined as the minimum number of rank-1 tensors that sum to a given tensor. In the case of matrices (order two tensors), rank and n-rank coincide. For general higher order tensors, there are no known simple methods to determine the rank of a given tensor.

Property 2. *Let the HOSVD of $\mathcal{A}$ be given as in Theorem 1. Then*

$$||\mathcal{A}||^2 = \sum_{i_1=1}^{R_1} {\sigma_{i_1}^{(1)}}^2 = \sum_{i_2=1}^{R_2} {\sigma_{i_2}^{(2)}}^2 = \cdots = \sum_{i_N=1}^{R_N} {\sigma_{i_N}^{(N)}}^2 = ||\mathcal{S}||^2$$

It can be easily verified that the Frobenius norm is orthogonally invariant in both matrix and higher order algebra (n-mode product). Therefore the well-known fact that the squared norm of a matrix equals the sum of its squared singular values can be generalized to higher order tensors.

Property 3. *Let the HOSVD of $\mathcal{A}$ be given as in Theorem 1 and let the n-mode rank of $\mathcal{A}$ be R_n. Define $\hat{\mathcal{A}}$ by discarding the smallest n-mode singular values $\sigma_{I'_n+1}^{(n)}, \sigma_{I'_n+2}^{(n)}, \ldots, \sigma_{R_n}^{(n)}$ for any given I'_n, i.e. set the corresponding "slices" of $\mathcal{S}$ to all zeros. Then we have*

$$||\mathcal{A} - \hat{\mathcal{A}}||^2 \leq \sum_{i_1=I'_1}^{R_1} {\sigma_{i_1}^{(1)}}^2 + \sum_{i_2=I'_2}^{R_2} {\sigma_{i_2}^{(2)}}^2 + \cdots + \sum_{i_N=I'_N}^{R_N} {\sigma_{i_N}^{(N)}}^2 \tag{9}$$

This property is the higher order link for the best low rank approximation in the matrix SVD. In terms of computation, this "truncated" tensor can be obtained by simply replacing $\mathcal{A}_{(n)} = U^{(n)}\Sigma^{(n)}{V^{(n)}}^H$ with a truncated SVD for each n in Algorithm 1. In this case, $\hat{\mathcal{A}}$ has n-rank equal to $\texttt{rank}_n(\hat{\mathcal{A}}) = I'_n$, respectively, for each mode. However, unlike matrix SVD, we only have an inequality here instead of an equality. More importantly, $\hat{\mathcal{A}}$ is in general not the best low rank approximation to $\mathcal{A}$ with the given n-rank constraints.

6 Numerical example

In this section, we give a small numerical example which illustrates each of the properties in section 5. Consider the $(3 \times 3 \times 3)$ tensor $\mathcal{A}$ defined by its 1-mode matrix unfolding

$$\mathcal{A}_{(1)} = \begin{bmatrix} 0.9073 & 0.7158 & -0.3698 & 1.7842 & 1.6970 & 0.0151 & 2.1236 & -0.0740 & 1.4429 \\ 0.8924 & -0.4898 & 2.4288 & 1.7753 & -1.5077 & 4.0337 & -0.6631 & 1.9103 & -1.7495 \\ 2.1488 & 0.3054 & 2.3753 & 4.2495 & 0.3207 & 4.7146 & 1.8260 & 2.1335 & -0.2716 \end{bmatrix}$$

The HOSVD for $\mathcal{A}$, written with its unfolded core and the singular vectors of each mode is

$$\mathcal{S}_{(1)} = \begin{bmatrix} 8.7088 & -0.0489 & -0.2797 & -0.1066 & 3.2737 & -0.3223 & -0.0033 & 0.1797 & -0.2223 \\ 0.0256 & 3.2546 & 0.2854 & 3.1965 & 0.2130 & 0.7829 & -0.2948 & -0.0378 & 0.3704 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}$$

$$U^{(1)} = \begin{bmatrix} 0.1121 & 0.7739 & 0.6233 \\ 0.5771 & -0.5613 & 0.5932 \\ 0.8090 & 0.2932 & -0.5095 \end{bmatrix}$$

$$U^{(2)} = \begin{bmatrix} 0.6208 & 0.4986 & 0.6050 \\ -0.0575 & 0.7986 & -0.5992 \\ 0.7818 & -0.3372 & -0.5244 \end{bmatrix}$$

$$U^{(3)} = \begin{bmatrix} 0.4624 & -0.0102 & 0.8866 \\ 0.8866 & 0.0135 & -0.4623 \\ -0.0072 & 0.9999 & 0.0152 \end{bmatrix}$$

The singular values for each mode (Frobenius norm of the corresponding slices in $\mathcal{S}$) are

$$\begin{aligned} \texttt{mode 1:} &\quad 9.3187, 4.6664, 0.0000 \\ \texttt{mode 2:} &\quad 9.2822, 4.6250, 1.0311 \\ \texttt{mode 3:} &\quad 9.3058, 4.6592, 0.5543 \end{aligned}$$

We can verify that $||\mathcal{A}||^2 = ||\mathcal{S}||^2 = 108.6136$, which is precisely (subject to roundoff error) the sum of the squares of the singular values on any single mode. We can also see that $\texttt{rank}_1(A) = 2$ as evidenced by the last singular value on mode 1 and the last horizontal slice of $\mathcal{S}$.

Now we demonstrate that the best low-rank approximation property in matrix SVD does not hold in HOSVD. Discarding $\sigma_3^{(2)}$ and $\sigma_3^{(3)}$ (the last singular values in both mode 2 and 3), we get a "truncated" core

$$\hat{\mathcal{S}}_{(1)} = \begin{bmatrix} 8.7088 & -0.0489 & 0.0000 & -0.1066 & 3.2737 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0256 & 3.2546 & 0.0000 & 3.1965 & 0.2130 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}$$

and the approximation $\hat{\mathcal{A}} = \hat{\mathcal{S}} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)}$ has an error of $||\mathcal{A} - \hat{\mathcal{A}}|| = 1.0880$. This clearly satisfies inequality (9). Consider another tensor $\mathcal{A}'$ defined by

$$\mathcal{A}' = \begin{bmatrix} 0.8188 & 0.8886 & -0.0784 & 1.7051 & 1.7320 & -0.0274 & 1.7849 & 0.2672 & 1.7454 \\ 1.0134 & -0.8544 & 2.1455 & 1.9333 & -1.5390 & 3.9886 & -0.2877 & 1.5266 & -2.0826 \\ 2.1815 & 0.0924 & 2.4019 & 4.3367 & 0.3272 & 4.6102 & 1.8487 & 2.1042 & -0.2894 \end{bmatrix}$$

which has n-rank = 2 on all three modes. But we have $||\mathcal{A} - \mathcal{A}'|| = 1.0848 < 1.0880 = ||\mathcal{A} - \hat{\mathcal{A}}||$.

7 Concluding remarks

This report presents a brief introduction to the generalization of SVD into higher order tensors. It has strikingly analogous form of decomposition and properties which are natural generalizations of matrix SVD. A small $(3 \times 3 \times 3)$ numerical example is given to illustrate each property in this report. It is, however, worth pointing out that other generalizations of SVD are possible, depending on which property one focuses on carrying into higher order. For example, the canonical decomposition, or parallel factors (CANDECOMP/PARAFAC) can be viewed as a special case of HOSVD in this report in which the core tensor $\mathcal{S}$ is required to be super-diagonal ($\mathcal{S}_{i_1,i_2,\ldots,i_N} \neq 0$ only if $i_1 = i_2 = \cdots = i_N$). We refer to [2, 3] for introductions to CANDECOMP/PARAFAC and other decompositions.

8 Acknowledgement

This report is largely based on the two well written and highly cited papers on tensors and HOSVD [4, 3]. The numerical experiment was performed using the MATLAB tensor toolbox from [1].

References

[1] Brett W. Bader and Tamara G. Kolda. Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. *ACM Transactions on Mathematical Software*, 32(4):635–653, December 2006.

[2] Rasmus Bro. Parafac. tutorial and applications. *Chemomemcs and Intelligent Laboratory Systems*, 38:149–171, 1997.

[3] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, September 2009.

[4] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl*, 21:1253–1278, 2000.