A Low Rank Tensorial Approximations method of computation of Singular Values and Singular Vectors for SVD problem

A. Milnikov
Faculty of Computer Technologies and Engineering
Black Sea International University
David Agmashenebeli Alley 13km, 2
Tbilisi 0131, Georgia
alexander.milnikov@gmail.com
http://www.ibsu.edu.ge/

Abstract:- A new method of computation of singular values and left and right singular vectors of arbitrary non-square matrices has been proposed. The method permits to avoid solutions of high rank systems of linear equations of singular value decomposition problem, which makes it not sensitive to ill-conditioness of decomposed matrix. On base of Eckart-Young theorem, it was shown that each second order r-rank tensor can be represent as a sum of the first rank r-order "coordinate" tensors. A new system of equations for "coordinate" tensor's generators vectors was obtained. An iterative method of solution of the system was elaborated. Results of the method were compared with classical methods of solutions of singular value decomposition problem.

Key words: - singular values, singular vectors, tensor product, low rank tensor approximation, harmonic decomposition, periodic components, white noise components

1. Introduction.

Singular Value Decomposition (SVD) being generalization of Eigen Value Decomposition (EVD) permits to compute singular (proper) values of non-square matrices. Despite of its generality and efficiency in some Time Series Analysis problems (we mean Singular Spectrum Analysis or SSA [1, 2] it has not been used widely in many engineering fields, including mechanical engineering. Nowadays the method has only started to be used in several fields of mechanics and applied physics: processing of experimental data in vibrations problems, in numerical computation of the coefficients of amplitude equations and normal forms, in some problems of Hamiltonian Mechanics [3,4,5].

Usage of classical Singular Value Decomposition (SVD) leads to necessity of calculation of eigen values and eigen vectors of high dimensional matrixes [1, 2]. There are a lot of well known and widely used methods of their computation [6, 7]. First of all, we outline a big group of so called transformation methods: Schur, LR, QR, Jacobi, Givens, and Householder etc. Also found a wide use polynomial iteration methods (direct computation of det (A-λB) determinant's roots) and a group of methods (variational) based on stationary property of the eigen values (Rayleigh quotient) [2, 6].We shall not discuss them as all these methods are well known and one can find their detailed consideration in many both classical and modern text books and monographs.

Despite of their very different nature all these methods can be characterized with similar disadvantages: necessity of big computations volume, not reliable stability and sensitivity for ill-conditioning. The latter problem (ill-conditioning) is very important especially for SSA because the matrix X constructed on observed data can be turned to be ill-conditioning. To avoid these computational problems we elaborated a new approach and algorithms based on principally new approach.

2. Theoretical part.

The all results of the work is based on the conception of approximation by low rank tensors and Eckart-Young theorem [8,9].

Definition[7]. A best rank-r approximation to a tensor $t \in V_1 \otimes ... \otimes V_k$ is a tensor s_{min} with

$$\|s_{\min} - t\| = \inf_{rank(s) \le r} \|s - t\|,$$

where $\|\cdot\|$ - Frobenius norm[1]

The latter generates Eckart-Young problem [8]: find a best r-rank approximation for tensor of order k.

The problem is not solvable in general. But for matrixes it was proved as

Eckart-Young theorem [7,8]. Given a p x n matrix **X** of rank r ≤ n ≤ p, and its singular value decomposition, **U**Λ**V**′, with the singular values arranged in decreasing sequence

$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \lambda_n \geq 0$,

then there exists a p x n matrix B of rank s, s ≤ r, which minimizes the sum of the squared error between the elements of A and the corresponding elements of B[1] when

$B = U\Lambda_s V^T$,

where U and V matrices consist of left and right singular vectors the matrix X and Λ_s is diagonal matrix with the diagonal elements

$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \lambda_s > \lambda_{s+1} = \lambda_{s+2} = \ldots = \lambda_n = 0$.

The theorem states that the least squares approximation in s dimensions of a matrix X can be found by replacing the smallest n-s roots of Λ with zeroes and remultiplying **U**Λ**V**′.

From the theorem follows that one can represent factorization of a K × L matrix X (with rank r ≤ min(K,L)) by means of Singular Value Decomposition as

$$X = \sum_{i=1}^{r} \lambda_i X_i = \sum_{i=1}^{r} H_i \ , \qquad (1)$$

where X_i - 1-rank matrices, which can be represented as a Kronecker product $X_i = u_i \otimes v_i$ of left u_i and right v_i singular vectors, corresponded to the singular value λ_i and $H_i = \lambda_i X_i$ – also 1-rank matrices. Note that H_i are decomposable, so $H_i = a_i \otimes b_i$, where a_i and b_i are linearly independent vectors[2]. They may be expressed via left u_i and right v_i singular vectors.

$a_i = \sqrt{\lambda_i} u_i$ and $b_i = \sqrt{\lambda_i} v_i$.

Each of the two systems of vectors u_i (i=1,2,…,K) and v_i (i=1,2,…,L) are orthonormal systems, therefore full contraction of X_i matrices satisfies[3]

$$(X_i, X_j^*) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} . \qquad (2)$$

One can consider (1) as a decomposition of the second order tensor (r-rank) X by a system of “coordinate” tensors X_i (1-rank). It is interesting to underline that singular values λ_i can interpreted as magnitudes of the projections of tensor X onto tensors X_i (i=1,2,…,r). The justification of such interpretation follows from orthonormality of vectors u_i (i=1,2,…,K) and v_i (i=1,2,…,L) : $(X, X_j) = (\sum_{i=1}^{r} \lambda_i X_i, X_j) = (\lambda_i ((u_i \otimes v_i), (u_j \otimes v_j)) = \begin{cases} \lambda_j, & i = j \\ 0, & i \neq j \end{cases}$.

If singular values and both types of singular vectors are known, one may use decomposition (1). Now we are interested in inverse problem: define singular values and both types of singular vectors, using matrix X and decomposition (1). It can be done by means of consequent computation of matrices H_1, by means of minimization of the sum of the squared errors between the elements of X and the corresponding elements of H_1. The squared sum of errors can be represented as follows

$$S^2 = \sum_{i=1}^{K} \sum_{j=1}^{L} (x_{ij} - h_{ij})^2 = \sum_{i=1}^{K} \sum_{j=1}^{L} (x_{ij} - a_i b_j)^2 .$$

Clear, that it is a function of $(K+L)^2$ unknown variables $a_i^j \ (i, j = 1, ..., K)$ and $b_i^j \ (i, j = 1, ..., L)$. So, minimization of the S^2 leads to the system of equations

$$\sum_{j=1}^{L} x_{ij} b_j - a_i \sum_{j=1}^{L} (b_j)^2 = 0; \ (i = 1, ..., K)$$
$$\sum_{m=1}^{K} x_{mn} a_m - b_n \sum_{m=1}^{K} (a_m)^2 = 0 . \ (n = 1, ..., L) \qquad (3)$$

Solution of the system gives vectors a and b, which define the best approximation of matrix X by 1-rank matrix H_1. In fact, the matrix H_1 is the first term in decomposition (1). Then, applying the same procedure to matrix $X_2 = X - H_1$, we are getting the second term H_2 and so on.

Now, there is a problem - how to solve the system (3), because we have already reduced the problem of computation of (1) to the problem of solution of the system (3). Few analysis permits to conclude, that the system can’t be solved analytically, so we elaborated numerical approach, which is the core of an algorithm of SVD by means of 1-rank tensors approximation. Below we

---

[1] That is in Frobenius norm sense.

[2] a_i is K –dimensional vector and b_i –is L-dimensional vector.

[3] It follows from the orthonormality of the systems of vectors u and v and the multilinearity properties of tensor product ($(\cdot,\cdot)$ stands for dot product): 1. for any real α , $x \otimes \alpha = \alpha \otimes x = \alpha x$ and 2. $((x \otimes y), (v \otimes u)) = ((y, v)x, u) = (y, v)(x, u)$. The last one is the full contraction of tensors $(x \otimes y)$ and $(v \otimes u)$.

ISBN: 978-1-61804-076-3

represent full algorithm of the system (3) solution and SVD by means of approximation by 1-rank tensors, which is completely based on the above theoretical consideration.

3. 1-rank tensors approximation algorithm of SVD problem solution.

Now we can represent the method, which, in fact, is a method of solution of the system (3). It starts with the choosing of any arbitrary matrix (vector) $a^{(1)}$ with the dimensions K×1.

The elaborated method consists of cycles and iterations. Total number of cycles equals to r where r is the rank of the matrix X or number of singular values of the matrix X. Each cycle consists of iterations and at the end of cycle i we have H_i where H_i is a component of decomposition $X = \sum_{i=1}^{r} H_i$ and i is the number of current cycle. Iterations are computed by means of the following steps.

Step 1: Choose arbitrary vector $a^{(0)}$.

Step 2: Construct a matrix using tensor product $w^{(0)} = a^{(0)} \otimes b^{(0)}$ where $b^{(0)}$ is a vector with unknown components; upper index in brackets shows number of iterations. These components can be computed by means of minimizing of Frobenius norm [2, 6] of differences between matrices X and $w^{(0)}$

$$\min_{1 \le a_i \le m} \left(\sum_{i=1}^{K} \sum_{j=1}^{L} (x_{ij} - a_i^{(0)} b_j^{(0)})^2 \right). \qquad (4)$$

Clear that minimizing of this norm is a special case of least square method [8]. As a result we shall have to get normal equations with respect to unknown components of vector b

$$\frac{\partial \left(\sum_{i=1}^{K} \sum_{j=1}^{L} \left(x_{ij} - a_i^{(0)} b_j^{(0)} \right)^2 \right)}{\partial b_j^{(0)}} = -2 \sum_{i=1}^{L} \left(a_i x_{ij} - \left(a_i^{(0)} \right)^2 \left(b_j^{(0)} \right) \right) = 0 \; .$$

(j=1,2,…,K)

The latter is a normal equation for minimization problem of (4). It is easy to define now unknown values of b_j:

$$b_j^{(0)} = \frac{\sum_{i=1}^{L} x_{ij} a_i^{(0)}}{\sum_{i=1}^{L} \left(a_i^{(0)} \right)^2}, \text{ where } j = 1, \ldots, K. \qquad (5)$$

Step 3: Next step of the algorithm consists of calculation of $a_i^{(1)}$ on the base of solution of the following problem

$$\min_{1 \le a_i \le K} \left(\sum_{i=1}^{K} \sum_{j=1}^{L} \left(x_{ij} - a_i^{(1)} b_j^{(0)} \right)^2 \right). \qquad (6)$$

Similar to (5), it is easy to represent the solution of (6) as

$$a_i^{(1)} = \frac{\sum_{j=1}^{K} x_{ij} b_j^{(0)}}{\sum_{j=1}^{K} \left(b_j^{(0)} \right)^2} \text{ , where } i = 1, \ldots, L. \qquad (7)$$

Using (7) one can construct a new matrix $w^{(1)} = a^{(1)} \otimes b^{(0)}$. If Frobenius norm of difference of matrices $w^{(0)}$ and $w^{(1)}$

$$\left| w^{(0)} - w^{(1)} \right|^2 = \sum_{i=1}^{L} \sum_{j=1}^{K} \left(a_i^{(0)} b_j^{(0)} - a_i^{(1)} b_j^{(0)} \right)^2$$

is greater than predefined accuracy ε, then we start new iteration going to step 2. In general while iteration i , we have matrix

$$w^{(j)} = \begin{cases} a^{(k)} \otimes b^{(k-1)}, j = 2k-1 \\ a^{(k)} \otimes b^{(k)}, j = 2k \end{cases} .$$

At the end of each iteration we check inequality $\left| w^{(j-1)} - w^{(j)} \right|^2 \le \varepsilon$. If it holds we have to stop iterations and this is the end of current cycle and denote matrix $w^{(i)}$ as $H_{(1)}$. Note that $H_{(1)}$ is the first component in SVD of matrix X.

To start next cycle we calculate X- $H_{(1)}$= $X_{(2)}$. The matrix defines new system of type (3), then we apply all above mentioned iteration to the system and so on till we get matrix $X_{(r)}$.

So we will get $X=H_1+H_2+...+H_r+X_n$ where X_n is very small which can be neglect able. So as a result

$$X = \sum_{i=1}^{r} H_i = \sum_{i=1}^{r} a_i \otimes b_i = \sum \lambda_i u_i \otimes v_i, \qquad (8)$$

where $a_i = \sqrt{\lambda_i} u_i$ and $b_i = \sqrt{\lambda_i} v_i$. The latter follows that left and right singular vectors can be represented as

$$u_i = \frac{a_i}{|a_i|} \text{ and } v_i = \frac{b_i}{|b_i|}$$

and taking into account (8) singular values can be represented as

$$\lambda_i = |a_i| |b_i| .$$

Thus, the represented algorithm solves the inverse problem defined above: define singular values and both types of singular vectors, using matrix X and decomposition

$$X = \sum_{i=1}^{r} \lambda_i X_i = \sum_{i=1}^{r} H_i .$$

4. Numerical Example

Below we represent result of application of suggested algorithm to computation of singular values and both (left and right) singular values of 7 x 9 singular matrix X (Table 1). Corresponding procedures were written in MatLab programming language.

7 x 9 singular matrix X Table 1

91	56	28	41	70	47	53	39	87
84	69	61	95	21	50	49	80	47
22	90	67	91	57	90	5	95	74
89	39	99	68	4	78	7	11	27
39	96	27	96	78	99	95	9	37
30	80	22	33	21	22	81	98	99
10	100	95	22	2	53	5	94	43

All results of computations are represented in below Tables.

First four left singular vectors. Table 2

Vector 1	Vector 2	Vector 3	Vector 4
-0.3557	-0.3059	-0.2090	0.4682
-0.4095	-0.0077	0.1119	0.3138
-0.4477	0.2461	0.0682	-0.3997
-0.3127	-0.0969	0.7104	0.3154
-0.4126	-0.6178	-0.0791	-0.5697
-0.3551	0.2047	-0.6447	0.2434
-0.3336	0.6425	0.1121	-0.1981

Last three left singular vectors. Table 3

Vector 5	Vector 6	Vector 7
0.5038	-0.3850	0.3390
-0.3163	0.6646	0.4238
0.6403	0.3451	-0.2134
-0.1101	-0.2063	-0.4837
-0.3067	-0.1478	0.0356
-0.2748	-0.0153	-0.5305
-0.2332	-0.4759	0.3785

First five right singular vectors. Table 4

Vector1	Vector3	Vector3	Vector4	Vector5
-0.2916	-0.3114	0.2578	0.7500	0.0168
-0.4380	0.1586	-0.1380	-0.3240	-0.2463
-0.3177	0.3366	0.5353	0.0826	-0.1586
-0.3775	-0.2585	0.2301	-0.1780	-0.0370
-0.2180	-0.3641	-0.1909	-0.2300	0.5661
-0.3647	-0.1641	0.3146	-0.3702	0.1334
-0.2449	-0.3958	-0.4533	0.0526	-0.6383
-0.3522	0.6099	-0.2917	0.0943	0.0386
-0.3393	0.0983	-0.3860	0.3041	0.4066

Last four right singular vectors. Table 5

Vector 6	Vector 7	Vector 8	Vector 9
-0.0182	0.3276	-0.2030	-0.1999
-0.2955	0.3698	-0.5787	0.1984
-0.3216	-0.0651	0.4746	0.3666
0.7447	-0.1607	-0.0908	0.3406
-0.1309	0.4495	0.4070	0.1589
-0.1901	-0.3244	-0.0322	-0.6663
-0.0983	-0.1404	0.3668	-0.0608
0.3972	0.2240	0.2462	-0.3741
-0.1834	-0.5905	-0.1690	0.2397

It easy verify that the same results could be obtained by means of corresponding MatLab software.

5. Conclusion.

A new method of computation of singular values and left and right singular vectors of arbitrary non-square matrices has been proposed. The method permits to avoid solutions of high rank systems of linear equations of singular value decomposition problem. On the base of Eckart-Young theorem, it was shown that each second order r-rank tensor can be represent as a sum of the first rank r-order "coordinate" tensors.

A new system of equations for "coordinate" tensor's generators vectors was obtained. An iterative method of solution of the system was elaborated. Results of the method were compared with classical methods of solutions of singular value decomposition problem.

References:

[1] Gattenmacher F.P. Matrix Theory. M., Nauka, 1967, p. 575.

[2] Elsner, J.B. and Tsonis, A.A.: *Singular Spectral Analysis. A New Tool in Time Series Analysis*, Plenum Press, 1996.

[3] M. Chu A differential equation approach to the singular value decomposition of bidiagonal matrices. Lin. Alg. Appl. 80:71- 80. 1986.

[4] K.P. Chen, D. D. Application of the singular value decomposition to the numerical computation of the coefficients of amplitude equations and normal forms. Applied Numerical Mathematics Volume 6, Issue 6, 1990, Pages 425- 430.

[5] Allen, M.R., and A.W. Robertson: "Distinguishing modulated oscillations from coloured noise in multivariate datasets", *Clim. Dyn.*, 12, 775–784, 1996.

[6] Bretscher, Otto (2005), *Linear Algebra with Applications* (3rd ed.), Prentice Hall.

[7] Golub, Gene H.; Van Loan, Charles F. (1996), *Matrix Computations* (3rd ed.), Johns Hopkins, ISBN 978-0-8018-5414-9.

[8] Mc'Connell A.J. Introduction in Tensor Analysis. M., Physmathgiz, 1963, p. 411.

[9] Postnikov M.M. Smooth Varieties. Term III. M., Nauka, 1988, p. 478.