A Tensorial Approximations Method as a Universal Filter

A. Milnikov

Abstract:- A new method of elaboration of non-parametric universal filter elaborated. The method is based on low rank tensorial approximation of singular matrixes of SVD problem. Results of the approximation method were compared with classical methods of solutions of singular value decomposition problem. It was shown that the elaborated method can be successfully used as a non-parametric universal filter for both stationary and non-stationary time series filtration.

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

I. INTRODUCTION.

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: Singular Value Decomposition (SVD) being generalization of Eigen Value Decomposition (EVD) permits to compute singular (proper) values of non-square matrices. Also it is used in such important engineering fields as 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].

A. Milnikov is with Faculty of Computer Technologies and Engineering, Black Sea International University, David Agmashenebeli Alley 13km, Tbilisi 0131, Georgia

Tel. +995 32 2595005 +995 32 2595008

alexander.milnikov@gmail.com

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.

II. 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 $t \in V_1 \otimes \ldots \otimes V_k$ 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 $p \times n$ matrix $\mathbf{X}$ of rank $r \le n \le p$, and its singular value decomposition, $\mathbf{U}\Lambda\mathbf{V}'$, with the singular values arranged in decreasing sequence

$$\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge \dots \lambda_n \ge 0,$$

then there exists a $p \times n$ matrix B of rank s, $s \le r$, which minimizes the sum of the squared error between the elements of B and the corresponding elements of X when

$B = U\Lambda_s V^T$,

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

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

From the theorem follows that one can represent factorization of a K×L matrix X (with rank $r \le \min(K,L)$) 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. 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

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

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)

$$(X, X_j) = (\sum_{i=1}^{r} \lambda_i X_i, X_j) = (\lambda_i ((u_i \otimes v_i), (u_j \otimes v_j)) = \begin{cases} \lambda_j, & i = j \\ 0, & i \ne j \end{cases}.$$

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 $a_i^j \ (i, j = 1, \dots, K)$ and $b_i^j \ (i, j = 1, \dots, L)$. So, minimization of the S^2 leads to the system of equations

$$\sum_{j=1}^{L} x_{ij} b_j - a_i \sum_{j=1}^{L} (b_j)^2 = 0; \ (i = 1, \dots, K)$$

$$\sum_{m=1}^{K} x_{mn} a_m - b_n \sum_{m=1}^{K} (a_m)^2 = 0. \ (n = 1, \dots, L) \qquad (3)$$

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 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.

III. ALGORITHM

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

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

(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

$u_i = \frac{a_i}{|a_i|}$ 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 $X = \sum_{i=1}^{r} \lambda_i X_i = \sum_{i=1}^{r} H_i$.

IV. 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.

V. EXAMPLES OF TIME SERIES FILTRATION

a. Non-Stationery time series

One of the most important advantages of the method described in the article is its ability to be used as a nonparametric universal filter for time series modeling, filtration and forecasting. We have to outline that this field of science is one of the most quickly development, and a lot of advanced scientific researches are devoted to solutions of its main problems [10-12].

Analysis of the time series permits to conclude easily that a time series may contain the following components: 1. trend; 2. low-frequency oscillation over trend; 3. several components of comparatively higher frequency. If it is necessary to separate these components one has to design different type of band-pass filters. Designing such kinds of filter requires first of all determinations of corresponding cut off frequencies which in turn requires investigation of time series Fourier spectrum. But because initial time series may contain trend and that means it is not stationary so one needs to eliminate the trend first.

We would like to outline that the latter is typical parametrical approach because choosing of cut off frequencies depends on investigator's personal experience. Also important to outline that eliminating trend is also typical parametrical approach, because choosing of type and order of trend (in general) also depends on researcher's personal opinion and experience.

It is easy to see that a lot of different very specific functions are needed in the filtering processes and usage of those functions requires high level professional skill. It is also requires certain practical experience to choose important parameters. These should be considered as significant disadvantages of traditional filtering methods.

We represent usage of SSA as a universal filter with examples of two time series: 1. a stock's rates of return of 463 daily observations and 2. model data of a periodic signal corrupted with white nose. Clear, these data are quit different according to their nature, so, from conventional approaches point of view, their filtration requires correspondingly different approaches. Universality of the filter, based on the method under consideration, consist in that it can be directly applied to both of the mentioned problems.

463 daily observations of rates of return are represented (dashed line) in the fig.1. The time series contains a trend, thus usage of conventional methods requires several steps to filter it: 1. elimination of the trend, to obtain time series close to stationary (may be in wide sense); 2. design of a filter for residual time series, which also include certain steps: determination of Fourier Spectrum of the time series after elimination, choosing of cut-off frequencies etc.

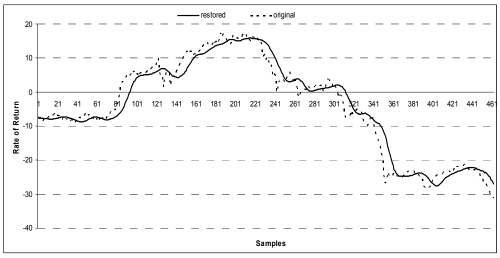

Fig.1. rates of return of 463 daily observations: observations - dashed curve; filtered observations – continuous curve

Filtered observations were restored by means of just matrix H_1 of decomposition (1), which was determined by means of elaborated method of computation of Singular values and vectors. Other factors of the decomposition represent high frequencies components of initial observations, which were filtered. Important to outline that non-stationary time series was filtered, without any preliminary transformation and specification of a type of trend. The latter shows non-parametric nature of the method.

b. Periodic signal with white nose additive component

Separation of Low frequencies components from the High ones is one of the most important problems in signal processing and applied time series analysis. Among others the most usable methods applied to solve the problem are: MUSIC and Pisarenko methods [2]. Both are based on estimation of pseudo spectrums of analyzed signal. The methods are parametric and efficient usage of them requires predetermination of main structure of mathematical model.

We illustrate results of usage of elaborated approach by means of solution of harmonic decomposition problem for a time series consisted of two periodic and additive noisy components. Time series is quite sort: 200 samples (Fig.1).

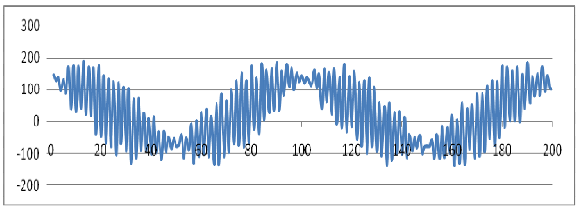

Fig.2. Periodic signal corrupted with white noise component .

Note that we took noise component quite strong: its magnitude comparable with the magnitude of basic signal. According to Singular Spectrum Analysis technique and on the base of these 200 samples, we designed trajectory matrix X with 20 rows and 10 columns.

Following the represented above Low Rank Tensorial Approximation approach, trajectory matrix X was decomposed according to (1).Total number of nonzero singular valus were 10 (it equals to rank of the trajectory matrix X). The singular values shown in Table 6 are as follows

Table 6

№	**Singular value**	**Relative singular values (Total=8711.84)**	**№**	**Singular value**	**Relative singular values (Total=8711.84)**
1	3172.71	36.42%	6	214.2	2.46%
2	3123.71	35.86%	7	207.21	2.38%
3	596.5	6.85%	8	206.37	2.37%
4	572.82	6.58%	9	201.12	2.31%
5	218.79	2.51%	10	198.41	2.28%

The first two values, which are significantly bigger than all others (their cumulative sum equals 72.2%), detected two the periodic components whereas all others represent random component, but among them the highest values of the third and fourth singular values connected with also high level of random components magnitude (cumulative sum of the first four singular values equals to 87%). Using hankelization procedure for: H_1 and H_2 matrices, two harmonic components were restored (fig.2 and fig.3) and for sum of matrices H_3+H_4 noise component was restored (fig.4)

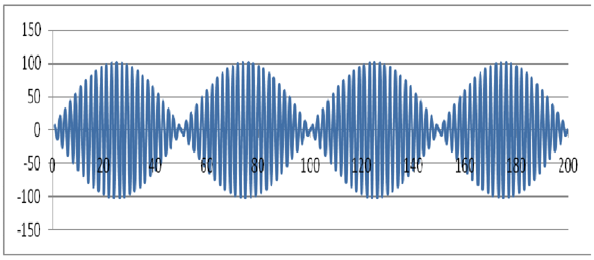

Fig.3. The first restored Periodic component.

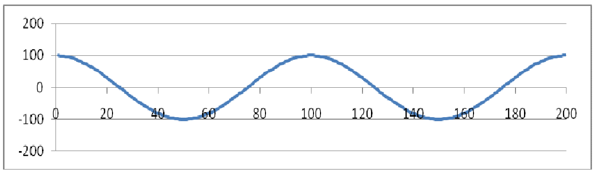

Fig.4.The second restored Periodic component

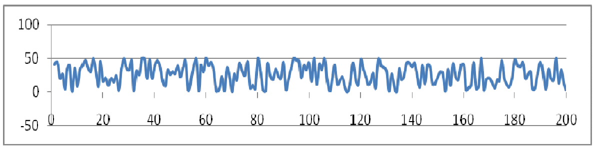

Fig. 5 . Restored White Noise component.

We have to underline two things: 1. high level of random components magnitude (its approximately 25 units) and 2. no preliminary Hypothesis were used to separate deterministic periodic ad random components.

The latter again confirms non-parametric nature of the method and its universality.

We compare SSA based approach with Music method. Application of the latter gives pseudo spectrum of the signal, which is represented in 6. Analyzing the Fig. it is easy to conclude that input signal consists of two periodic components because pseudo spectrum contains two peak points at frequencies 50 Hz and 100 Hz. MUSIC method does not give the shape of components of the signals, so one needs to construct these components analytically.

On the contrary elaborated method gives as its output all components of the initial signal. They are represented graphically below.

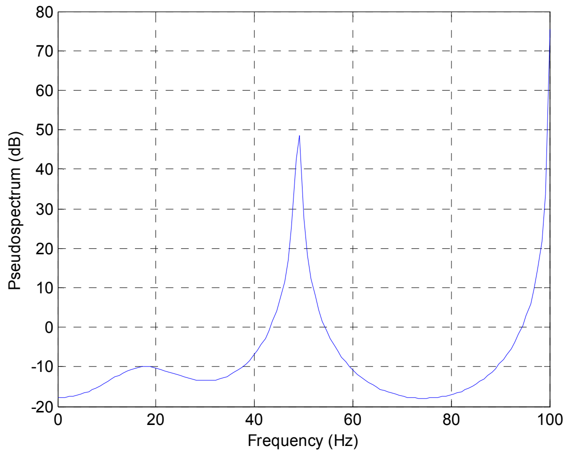

Fig. 6. Pseudo spectrum of the signal

VI. 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.

The elaborated method can be successfully used as a non-parametric universal filter for both stationary and non-stationary time series filtration.

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] A. Milnikov. A Low Rank Tensorial Approximations method of computation of Singular Values and Singular Vectors for SVD problem. Proceedings of the 6th International Conference on Applied Mathematics, Simulation, Modeling (ASM'12). Athens, Greece, March 7-9, 2012, pp.49-53.

[5] M.R.Allen, 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.

[10]
. Yagoubi. A geometric Approach to a Non-Stationary Process. Proceedings of the 6th International Conference on Mathematical Models and Engineering Sciences (MMES'II), Puerto De La Cruz, Spain, December 10-12, 2011, pp.179-183

[11]
. Ragulskis. Time Series Prediction based on Fuzzy and Neural Networks. Proceedings of the 6th International Conference on Mathematical Models and Engineering Sciences (MMES'II), Puerto De La Cruz, Spain, December 10-12, 2011, pp. 11-12

[12] A. Loredana, M. Neamtu. Deterministic and Stochastic Dynamic Model with Delays of the Financial Crisis Contagious. Proceedings of the 6th International Conference on Applied Mathematics, Simulation, Modeling (ASM'12). Athens, Greece, March 7-9, 2012, pp.92-97