

## **Method of Adaptive Forecasting Based On Multidimensional Linear Extrapolation**

Mantula E.V.<sup>1</sup>, Mashtalir S.V.<sup>1</sup>

<sup>1</sup> (Informatics department, Kharkiv National University of Radio Electronics, Ukraine)

---

**ABSTRACT:** *The paper is sanctified to the problem of time series analysis. The linear extrapolation methods of multidimensional time series with Euclidian metric and proximity function is offered. As practical application the problem of spatio-temporal segmentation of video is considered.*

**Keywords** – linear extrapolation, multidimensional, time series, metric, video

---

### **I. INTRODUCTION**

A necessity of forecasting time series heterogeneous in nature often arises for different applications, viz. technical, financial, economic, medico-biological, environmental, agricultural ones, etc. Nowadays, there exists a variety of methods to solve such problems: from traditional statistical and adaptive [1-5] approaches to more sophisticated techniques with artificial neural networks in use [6-10].

More complicated situation appears when a predicted sequence is multidimensional. However, this problem can also be solved either by decomposition of the original multidimensional series to the set of one-dimensional ones or by means of multi-dimensional (MIMO) predictive models which parameters can be successfully identified by the same statistics, adaptive or neural network algorithms.

It should be emphasized that successful application of these procedures assumes availability of sufficiently representative sample of observations, which gives possibility to create a fairly accurate predictive model. Unfortunately, often there are situations when the sample is either small or predictable process is not stationary, so its prehistory cannot be used to find parameters of the model. It is very hard to think about effective predictive model in this situation and therefore, prediction methods should be introduced, that do not use the concept of the model. Such approach becomes especially important when it is required to process enormous data volumes on-line or even in real time, e.g. its application to shot (basic unit of video content analysis) boundary detection for video annotation and retrieval due to inherently temporal aspect of video streams. The key to explaining this fact may lie in that content-based video retrieval systems have become an active field of research along with both the rapid increase in the amount of video acquisition and wide range of video applications, what is additionally emphasized by arbitrariness of video structures and semantic concepts. In any case, shot boundary detection can be a ground for effective parsing, as time series processing is an essential component of preliminary video analysis and forecasting.

The paper is devoted to the development and improvement of one of the possible approaches to multidimensional time series forecasting. The structure of the paper is as follows. The second section sketches multidimensional linear extrapolation problem. Section 3 enlarges upon an adaptive multidimensional linear extrapolation algorithm with Euclidean metric. Section 4 presents a new algorithmic approach to the adaptive multidimensional linear extrapolation based on the proximity function. The final section proposes summarized results.

### **II. The method of multidimensional linear extrapolation**

The basis of traditional mathematical forecasting methods (statistical, adaptive, neural networking, etc.) consists in different kinds of mathematical models obtained as a result of structural and parametric identification. Problems of time extrapolation are usually solved on the basis of mathematical models, and such models explicitly or implicitly include discrete time as an argument. Predictive model synthesis is not possible if there is not enough data to construct the mathematical model. In this case, spatial prediction (extrapolation) can be used instead of time extrapolation, which comes to estimation of vector field under particular observations. Multidimensional linear extrapolation [11] should be marked among the most promising methods of spatial extrapolation, it has proved its effectiveness for solving some actual design problems and control of complex multi-dimensional nonlinear objects. Consider multidimensional linear extrapolation method applied to the problem of one-step prediction of  $n$ -dimensional nonlinear nonstationary time series  $y(k), k = 1, 2, \dots, N$  where  $k$  denotes the discrete time. Suppose the analyzed series (its  $i$ -th component) can be described by nonlinear relation (NARX-model) in the form

$$\begin{aligned} \hat{y}_i(k) &= f_i(y_i(k-1), \dots, y_i(k-n_{A,i}), x_1(k-1), \dots, x_1(k-n_B), \\ &\quad x_2(k-1), \dots, x_2(k-n_B), \dots, x_p(k-1), \dots, x_q(k-n_B)) = \\ &= f_i(z_{i1}(k), \dots, z_{i,n_{A,i}}(k), \dots, z_{i,n_{A,i}+n_B q}(k)), \end{aligned} \quad (1)$$

where  $f_i(\circ)$  is a priori unknown nonlinear relation which should be restored from available observations;  $\hat{y}_i(k)$  stands for an evaluation (prediction) of the controlled sequence  $y_i(k)$ , performed using data available by  $(k-1)$  point in time;  $i=1, \dots, n$ ;  $n_{A,i}$  denotes the considered history depth of the analyzed sequence (generally  $n_A, n_B$  are memory parameters of observations);  $x_p(k-l)$  is  $p$ -th exogenous component of multidimensional signal that affects  $y_i(k)$ ;  $l=1, \dots, n_B$ ;  $p=1, \dots, q$ . Expression (1) can also be rewritten in vector-matrix form

$$\hat{y}(k) = F(y(k-1), \dots, y(k-n_A), x(k-1), \dots, x(k-n_B)) = F(z(k)) \quad (2)$$

where  $\hat{y}(k) = (\hat{y}_1(k), \hat{y}_2(k), \dots, \hat{y}_n(k))^T$ ,  $y(k) = (y_1(k), \dots, y_n(k))^T$ ,  $x(k-1) = (x_1(k-1), \dots, x_q(k-1))^T$ ,  $z(k) = (y^T(k-1), \dots, y^T(k-n_A), x^T(k-1), \dots, x^T(k-n_B))^T$  is  $(n \cdot n_A + qn_B) \times 1$  prehistory vector.

Non-linear transformations  $f_i(\circ)$  and  $F(\circ)$  can be retrieved in the process of any artificial neural network learning when a learning sample exists. However, if the sample is small enough, the neural network approach is unusable, while multivariate linear extrapolation will provide quite accurate results. The problem of multidimensional linear extrapolation applied to prediction of multidimensional time series can be described using notations (1), (2) as follows [11]. Let matrix of precedents (prehistory)

$$P = \begin{pmatrix} z^T(1), & y^T(1) \\ z^T(2), & y^T(2) \\ \vdots & \\ z^T(N), & y^T(N) \end{pmatrix}$$

be defined with dimension  $N \times (nn_A + qn_B + n)$ . Extrapolation, in fact, is reduced to evaluation of

$$\hat{y}(N+1) = \psi(z(N+1), P), \quad (3)$$

in time point  $N$ , where  $\psi(\circ)$  is extrapolation algorithm which should meet a number of requirements. The main requirement is that after its application, all the consequences matrix of precedents should be accurately recovered, i.e.

$$\hat{y}(k) = y(k) = \psi(z(k), P), \quad k=1, 2, \dots, N. \quad (4)$$

It should be noted that prediction methods based on various mathematical models almost never provide the condition (4). From the other requirements it can be noted that the algorithm  $\psi(\circ)$  should be constructed in a vector form, i.e.

$$\psi = (\psi_1, \dots, \psi_n), \quad y_i = \psi_i(Z, P), \quad i=1, \dots, n; \quad (5)$$

the complexity should increase by  $n$  and  $N$  not faster than linearly; the algorithm must be able to operate for all  $N$  (even if  $N=1$ ). It is clear that for small  $N$ , and also  $N=1$ , sufficiently accurate mathematical model cannot in principle be constructed.

### III. An adaptive multidimensional linear extrapolation algorithm with Euclidean metric

According to L.A. Rastrigin [11], consider multidimensional linear extrapolation algorithm as the following sequence of steps:

i). generation of predicted process history in the form of matrices

$Z = (z(1), z(2), \dots, z(N))$  is matrix of  $(nn_A + qn_B) \times N$  dimension,

$Y = (y(1), y(2), \dots, y(N))$  is  $n \times N$  matrix;

ii). finding the weight vector  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)^T$  that provides minimum for norm function

$$\|z(N+1) - \sum_{k=1}^N \lambda(k)z(k)\|^2 = \|Z(N+1) - Z\lambda\|^2; \quad (6)$$

iii). formation of optimal prediction as a linear combination

$$\hat{y}(N+1) = \sum_{k=1}^N \lambda_k y(k) = Y\lambda. \quad (7)$$

Minimizing of expression (7) can be performed by means of the standard method of least squares and

as a result, it is obvious that

$$\lambda = (Z^T Z)^{-1} Z^T z(N+1) \tag{8}$$

which exists only if the matrix  $Z^T Z$  is nonsingular. It is essential to see that under sufficiently small  $N$  ( $N < nm_A + qn_B$ ) this is not true, therefore it is proposed to use pseudoinverse matrix [12] and finally we have the expression

$$\lambda = Z^+ z(N+1) \tag{9}$$

from which it follows that, in fact, the problem is reduced to finding orthogonal projection of the vector  $z(N+1)$  on the linear hull generated by the prehistory vectors  $Z$ .

From computational point of view there are no difficulties in the implementation of this procedure, however, the solution becomes more complicated if the processing data are sequential in real time. In this case, all previous relations can be rewritten in the following form

$$Z_k = (z(1), z(2), \dots, z(k)),$$

$$Y_k = (y(1), y(2), \dots, y(k)),$$

$$\|z(k+1) - \sum_{l=1}^k \lambda_l z(l)\|^2 = \|z(k+1) - Z_k \Lambda_k\|^2, \tag{10}$$

$$\Lambda_k = (\lambda_1, \lambda_2, \dots, \lambda_k)^T,$$

$$\hat{y}(k+1) = \sum_{l=1}^k \lambda_l y(l) = Y_k \Lambda_k, \tag{11}$$

$$\Lambda_k = Z_k^+ z(k+1). \tag{12}$$

In [11] it is proposed to use Greville formulae for the matrix  $Z_{k+1}^+$  calculation from available  $Z_k^+$  and incoming values  $z(k+1)$ ,  $y(k+1)$ , although it is much more preferable to replace it by regularized version of (8) in the form

$$\Lambda_k = (Z_k^T Z_k + \gamma I_k)^{-1} Z_k z(k+1) \tag{13}$$

where  $\gamma$  is regularization parameter,  $I_k$  denotes  $(k \times k)$  identity matrix.

For processing of non-stationary time series, which characteristics change unpredictably over time, it is appropriate to solve the problem in the ‘sliding window’ instead of all available sample processing. ‘Sliding window’ consists of  $\chi$  most recent observations, and in this case relations (10) – (12) can be rewritten as follows

$$Z_{k,\chi} = (z(k-\chi+1), z(k-\chi+2), \dots, z(k)),$$

$$Y_{k,\chi} = (y(k-\chi+1), y(k-\chi+2), \dots, y(k)),$$

$$\|z(k+1) - \sum_{l=k-\chi+1}^k \lambda_l z(l)\|^2 = \|z(k+1) - Z_{k,\chi} \Lambda_{k,\chi}\|^2, \tag{14}$$

$$\Lambda_{k,\chi} = (\lambda_{k-\chi+1}, \dots, \lambda_k)^T,$$

$$\hat{y}(k+1) = \sum_{l=k-\chi+1}^k \lambda_l y(l) = Y_{k,\chi} \Lambda_{k,\chi}, \tag{15}$$

$$\Lambda_{k,\chi} = Z_{k,\chi}^+ z(k+1). \tag{16}$$

For implementation of this procedure in real time, the recurrent algorithm of pseudoinversion on ‘sliding window’ had been proposed in [11], but it is inconvenient and computationally complex. In this regard it is expedient to use a modification of (13) in ‘window’ view, and its packet form can be written as

$$\Lambda_{k,\chi} = (Z_{k,\chi}^T Z_{k,\chi} + \gamma I_\chi)^{-1} Z_{k,\chi} z(k+1), \tag{17}$$

and respectively recurrent form [13] can be represented as

$$\begin{cases} \Lambda_{k+1,\chi} = \Lambda_{k,\chi} + \frac{P(k)(z(k+1) - \Lambda_{k,\chi}^T z(k))}{1 + z^T(k)P(k)z(k)} z(k), \\ \tilde{P}(k-1) = P(k-1) + \frac{P(k-1)z(k-\chi)z^T(k-\chi)P(k-1)}{1 - z^T(k-\chi)P(k-1)z(k-\chi)}, \\ P(k) = \tilde{P}(k-1) - \frac{\tilde{P}(k-1)z(k)z^T(k)\tilde{P}(k-1)}{1 + z^T(k)\tilde{P}(k-1)z(k)}, \\ \Lambda_{0,\chi} = 0, P(k) = \gamma^{-1}I_{\chi}. \end{cases} \quad (18)$$

Actually the forecast is calculated according to (15).

It is evident that expressions (15), (18) simplify prediction considerably, but there arises a question of reasonable choice of window  $\chi$  size which is usually specified by some heuristics, that in the end reduces the overall efficiency of the approach.

#### IV. Adaptive method of multidimensional linear extrapolation based on the proximity function

Multi-dimensional extrapolation method proposed below is based on the proximity (distance) between the last history vector  $z(N+1)$  and all previous data  $z(1), \dots, z(N)$ , and also it is based on making predictions  $\hat{y}(N+1)$  using the same function.

Implementation of the method consists in the following steps:

i). Calculation of the distance between the vector  $z(N+1)$  and all the previous functions  $z(k)$  on the basis of proximity function  $d(N+1, k)$  (in the simplest case, this is the Euclidean metric)

$$d(N+1, k) = \|x(N+1) - x(k)\| \quad \forall k;$$

ii). Arrangement of these distances in increasing order (ranking)

$$d^1(N+1, k_1) < d^2(N+1, k_2) < \dots < d^N(N+1, k_N);$$

iii). Selection of the first  $\chi$  vectors, for which the following condition is true

$$d^{\chi}(N+1, k_{\chi}) \leq \varepsilon$$

where  $\varepsilon$  is given or computed threshold;

iv). Finding a set of weights  $\lambda_l$

$$\lambda_l = \frac{(d^l)^{-1}}{\sum_{l=1}^{\chi} (d^l)^{-1}}, \quad 1 \leq l \leq \chi;$$

v). Forecast computing

$$\hat{y}(N+1) = \sum_{l=1}^{\chi} \lambda_l z(l).$$

All iterations are repeated under acquisition of a new experimental observation  $y(N+1)$ .

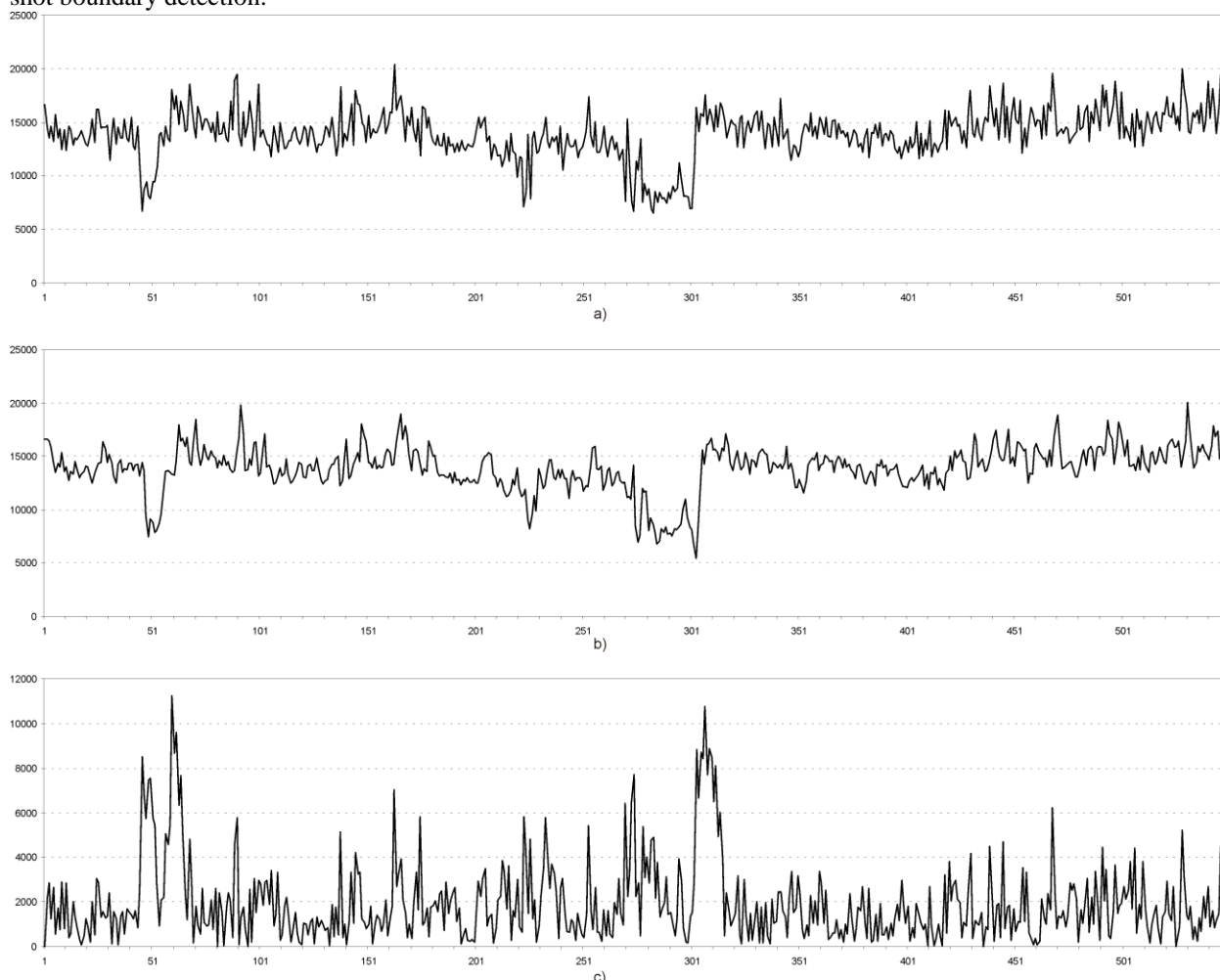
Thus,  $\chi$  observations are also involved at each step of forecast formation, but this value may change, and it is clear that the less  $\chi$  is, the more non-stationary the signal is. It is also easy to see that if  $y(k) = const$ , then  $\chi = N$ ,  $\lambda_l = \frac{1}{N}$ .

One of the issues that can be solved with the above mentioned forecasting approach is video analysis. This is mostly due to multidimensional time series nature of video data. Besides, one of the peculiarities common for video data should be noted. It consists in close similarity of consecutive video frames (from a single shot), which permits forecasting future frames based on the previous ones. In particular, this video property is often used in a number of compression algorithms. For shot boundary detection, where a search for fragments with a common sense is an issue, video segmentation should be considered.

Thus, if spatio-temporal video segmentation is considered [14], an input is presented as a series of frames, being partitioning of initial data. In turn, each frame can be characterized by its spatial segments, i.e. it is possible to present each frame as a vector which values correspond to geometric characteristics of segments, such as an area, perimeter, tilting about axes, etc.

In order to extract segments with a common sense from initial data, it is possible to use the aforementioned approach in a following way: if the forecast (15) for  $N+1$  frame differs significantly from its actual value, it should be interpreted as a change of a shot, and consequently as a segment boundary. By doing so, the forecast may be performed only for vectors included into a single segment, otherwise values from the previous video segments may give a negative impact on the whole process of segmentation. Thereby, extrapolation intervals should be shifted according to segment boundaries while making forecasts.

Fig. 1 shows an example of video application for the proposed extrapolation method. It can be seen clearly that the moment when characteristic values change a lot over a period, the forecast also changes greatly in some time. In other words, the difference between an actual value and its forecast defines moments of changes in shots. In addition, if short-term random outliers appear, they are smoothed, and they do not influence shot boundary detection.



**Fig. 1 – Experimental Result**

- a) initial video data;**
- b) multidimensional linear extrapolation;**
- c) difference between initial data and its multidimensional linear extrapolation**

Using this line of reasoning, one may come to the conclusion that the proposed extrapolation approach can be used for «rough» temporal partitioning in terms of spatio-temporal segmentation.

## V. CONCLUSION

The results of the study are evaluated and assessed in the light of the problems of temporal video analysis in an appropriate feature spaces. The major points covered by this paper may be summarized as follows. A method for multi-dimensional non-stationary time series forecast, which is based on the spatial extrapolation paradigm, has been proposed. The method does not require large volumes of a priori information (it is sufficient to have a small learning sample), it is not necessary to solve additional optimization problems or

to make pseudoinversion. It is rather simple from computational point of view.

However, when choosing forgetting parameter, smoothing effect can affect forecasting validity, and it is required to take into consideration assessment of both threshold and length of 'sliding window'. It is known with certainty that there is a good reason to create procedures of finding these parameters which ought to be included in adaptive forecasting based on multidimensional linear extrapolation.

### References

- [1] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis: Forecasting and Control. 4th Edition* (Inbunden: Wiley Series in Probability and Statistics, 2008).
- [2] S. Makridakis, S. Wheelwright, R.J. Hyndman, *Forecasting: methods and applications* (New York: John Wiley & Sons, 1998).
- [3] D.C. Lewis, *Industrial and Business Forecasting Methods*. (London: Butterworths Scientific, 1982)
- [4] T. Masters, *Neural, Novel & Hybrid Algorithms for Time Series Prediction* (N.Y.: John Wiley & Sons, Inc., 1995).
- [5] Yu.P. Lucashin, *Adaptive methods of short-term prognosis of time series* (M.: Finance and statistics, 2003). /In Russian/
- [6] A.G. Ivachnenko, J.A. Muller, *Selbstorganization von Vorherzagemodellen* (Berlin: VEB Verlag Technik, 1984).
- [7] D.T. Pham, X. Liu, *Neural Networks for Identification, Prediction and Control* (London: Springer – Verlag, 1995).
- [8] S. Kingdom, *Intelligent Systems and Financial Forecasting*. (Berlin: Springer – Verlag, 1997).
- [9] J.S. Zirilli, *Financial Prediction Using Neural Networks* (London: Int. Thomson Computer Press, 1997).
- [10] D.P. Mandic, J.A. Chambers, *Recurrent Neural Networks for Prediction* (Chichester: John Wiley & Sons, Ltd., 2001).
- [11] L.A. Rastrygin, Yu.P. Ponomariov, *Extrapolation methods in planning and management* (M.: Engineer, 1986). /In Russian/
- [12] A. Albert, *Regression and the moor-penrose pseudoinverse* (N.Y. and London: Academic Press, 1972).
- [13] Ye. Bodyanskiy, O. Rudenko, *Artificial neural networks: architectures, educating, application* (Kharkov. TELETECH, 2004). /In Russian/
- [14] D. Kinoshenko, S. Mashtalir, V. Shlyakhov Temporal video segmentation via spatial image segmentation *International journal information technologies & knowledge* 7(3), 2013, 212-219.