

# Local approximation of discrete processes by interpolation polynomials

A A Kolpakov<sup>1</sup>, Yu A Kropotov<sup>1</sup>

<sup>1</sup>Vladimir State University named after Alexander and Nicholay Stoletovs, Orlovskaya street, 23, Murom, Vladimir Region, 602264

e-mail: desT.087@gmail.com

**Abstract.** This paper discusses the structure of the devices and their defining formulas used for local approximation using power-algebraic polynomials when the observed data are known exactly. A multichannel system for processing discrete sequences is considered. On the basis of the considered system the research of acceleration of calculations in the system from specialized computational modules is carried out. The carried out researches have shown, that the developed model of multichannel data processing system allows to reduce essentially time for data processing.

## 1. Introduction

Organization of calculations or processing of data on the observed samples of the process carried out with the help of the representation of algebraic polynomials refers to the class of methods of analytical representation and signal processing. This paper discusses the structure of the devices and their defining formulas used for local approximation with the help of power algebraic polynomials, when the observed data are known exactly.

## 2. Multi-channel system for processing discrete sequences

The representation of the functions of the observed data can be referred to the issues of interpolation with the help of polynomials on linearly independent systems of functions and are considered later as the issues of coordination of local interpolation functions on smoothness in the nodes of their conjugation [2, pp. 146-194].

The approximation of the processes on the observed samples can be carried out with the help of algebraic polynomials, for example, the point method of least squares or methods of local interpolation, which belong to the class of methods of analytical representation of signals. Here we consider algebraic models and methods of the need to solve local satisfaction with the conditions of smooth coupling, approximation by discrete data.

Thus, at first it is assumed that all sample values of the process and its derivatives in the interpolation and coupling nodes are known.

Let's denote an interpolation polynomial through  $P_i(t)$ , which provides the approximation of the process  $y(t)$  at the interval  $[\bar{t}_i, \bar{t}_{i+1})$ ,  $\bar{t}_i < \bar{t}_{i+1}$ ,  $i = 0, 1, \dots$ , by some number of known values of this

process  $y(t_{ij}) = y_{ij}$ ,  $j = 0, 1, \dots, n$ , and, possibly, its derivatives  $y^{(r)}(t_{ij}) = y_{ij}^{(r)}$ ,  $r = 0, 1, \dots, m_j$  and  $y_{ij}^{(0)} \equiv y_{ij}$ .

The moments  $\bar{t}_i$  of the pairing of polynomials  $P_{i-1}(t)$  and  $P_i(t)$  will be referred to points or nodes in the pairing, and the moments  $t_{ij}$ , at which sample values  $y_{ij}$  and  $y_{ij}^{(r)}$  are taken, will be referred to interpolation nodes. The intervals  $[\bar{t}_{i-1}, \bar{t}_i]$  and  $[\bar{t}_i, \bar{t}_{i+1}]$ , called interpolation intervals, are obviously areas of definition for polynomials  $P_{i-1}(t)$  and  $P_i(t)$ . Interval boundaries, within which interpolation nodes  $t_{ij}$  are located, are denoted as moments  $\bar{\bar{t}}_i$  and  $\bar{\bar{t}}_{i+1}$ , so interval  $[\bar{\bar{t}}_i, \bar{\bar{t}}_{i+1}]$  is the area where the observed data used in interpolation formula construction or mapping  $P_i(t) : [\bar{\bar{t}}_i, \bar{\bar{t}}_{i+1}] \rightarrow [\bar{t}_i, \bar{t}_{i+1}]$  are defined.

Interpolation polynomial  $P_i(t)$  by the system of algebraic functions  $t^k$ ,  $k = 0, 1, \dots, N$ , i.e. polynomial  $P_i(t) = a_{i0} + a_{i1}t + a_{i2}t^2 \dots + a_{iN}t^N$ , can be written in the form

$$P_i(t) = \sum_{j=0}^n \sum_{r=0}^{m_j} L_{ij}^r(t) y_{ij}^{(r)}, \quad (1)$$

where  $L_{ij}^0(t) \equiv L_{ij}(t)$  and  $L_{ij}^r(t)$  is polynomials satisfying the conditions in the interpolation nodes

$$L_{ij}(t) \Big|_{t=t_{ik}} = \delta_{jk}, \quad \frac{d^l}{dt^l} L_{ij}^r(t) \Big|_{t=t_{ik}} = \delta_{jk} \delta_{lr}, \quad l = 0, 1, \dots, m_j, \quad (2)$$

where  $\delta_{jk}$  is Kroneker's symbol.

It should be taken into account that the system of functions  $t^k$  refers to the class of generalized Chebyshev systems, which by definition should have a non-zero determinant

$$U^* \begin{pmatrix} 0 & 1 & 2 & \dots & N \\ t_0 & t_1 & t_2 & \dots & t_N \end{pmatrix} = \begin{vmatrix} 1 & 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ t_0 & 1 & \dots & 0 & \dots & t_n & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ t_0^N & t_0^{N-1} & \dots & t_0^{N-m_0} & \dots & t_n^N & \dots & t_n^{N-m_n} \end{vmatrix} \neq 0. \quad (3)$$

In this case, we can conclude that there is an interpolation polynomial at any location of interpolation nodes and, in particular, the existence of fundamental polynomials  $L_{ij}^r(t)$ , and hence the possibility of approximation of observed data using the (1).

The essential point of this conclusion is that in each node of interpolation there should be known the values of both the process and all its derivatives up to a given maximum order. However, if the physical features of the problem as the initial data in the construction of the interpolation polynomial are sample values of the process and its derivatives, measured at different moments of time, and interpolation polynomial instead of (1) is given by the expression

$$P_i(t) = \sum_{j=0}^n \sum_{r \in M_j} L_{ij}^r(t) y_{ij}^{(r)}, \quad (4)$$

where  $M_j$  is a subset of multiple integers  $(0, 1, \dots, m_j)$ . The fundamental polynomials satisfy the conditions

$$\frac{d^l}{dt^l} L_{ij}^r(t) \Big|_{t=t_{ik}} = \delta_{jk} \delta_{lr}, \quad l, r \in M_j. \quad (5)$$

It can be shown that the problem of constructing polynomials (4) and (5) in some combinations of interpolation nodes may not have a solution.

This can be shown in a simple example of building a second-degree polynomial  $P(t) = a_0 + a_1t + a_2t^2$  using the values specified in the interpolation nodes  $P(t_0) = y_0$ ,  $P(t_1) = y_1$ ,  $P(t_2) = y_2$ . The coefficients of this polynomial, taking into account the given conditions, should satisfy the system of linear algebraic equations

$$\begin{pmatrix} 1 & t_0 & t_0^2 \\ 0 & 1 & 2t_1 \\ 1 & t_2 & t_2^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix}. \quad (6)$$

The determinant of this system is  $(t_2 - t_0)(t_2 + t_0 - 2t_1)$ . In case of equidistant nodes, when  $t_2 + t_0 = 2t_1$ , it is equal to zero. At the same time, the solution of the system and, accordingly, the solution of the interpolation problem does not exist, except for the unlikely case of a multi-digit solution, when the ranks of the main and extended matrices coincide.

Such a situation, although not necessarily, may arise in general. Therefore, if there is a need to use formulas of the type (4), (5), the selected interpolation scheme should be checked for the existence of a solution to the problem of interpolation polynomial construction. Apart from this condition, no other restrictions are imposed on the application of these formulas.

Using formula (1), the approximation  $\hat{y}(t)$  of process  $y(t)$  on the entire time axis can be written as

$$\hat{y}(t) = \sum_{i=0}^{\infty} P_i(t) I_i(t) = \sum_{j=0}^n \sum_{r=0}^{m_j} \sum_{i=0}^{\infty} I_i(t) L_{ij}^r(t) y_{ij}^{(r)}. \quad (7)$$

Sampling values  $y_{ij}$  and  $y_{ij}^{(r)}$ , which clearly define interpolation and fundamental polynomials (1; 4), can be divided into two classes: the actual measured values, the values of process  $y(t)$  and its derivatives  $y^{(r)}(t)$  in the interpolation nodes  $t_{ij}$ , and the values that represent their estimation from indirect data. This estimation can be obtained, for example, by calculating the values of polynomial  $P_{i-1}(t)$  representing the process at the previous step of interpolation, and the values of its derivatives  $P_{i-1}^{(r)}(t)$  in the required interpolation nodes. On the interpolation schemes, such values will be marked with an asterisk in the future so that they can be distinguished from the measured values.

To ensure the required smoothness of the interpolation formula (7) in the interval nodes  $[\bar{t}_i, \bar{t}_{i+1})$ , nodes  $\bar{t}_i$  should obviously be included in the set of interpolation nodes  $t_{ij}$  for the given index value  $i$ , if they are no longer included in their number by definition. In this case, nodes  $\bar{t}_{i+1}$  do not have to belong to this set.

When considering the case where the layout of the interpolation nodes and the length of the segments  $[\bar{t}_i, \bar{t}_{i+1})$  do not depend on the number  $i$  of the interpolation step, i.e.

$$\bar{t}_{i+1} - \bar{t}_i = \Theta \quad \text{and} \quad t_{ij} = t_j + i\Theta, \quad (8)$$

and if you enter the discrete time  $t = \tau + k\Theta$ , the interpolation formula (7) can be written as

$$\hat{y}(t) \equiv \sum_{k=0}^{\infty} \hat{y}_k(\tau) = \sum_{k=0}^{\infty} \sum_{j=0}^n \sum_{r=0}^{m_j} I(t - k\Theta) L_j^r(t - k\Theta) y_{kj}^{(r)}, \quad (9)$$

where  $L_j^r(\tau) \equiv L_{0j}^r(\tau)$  – the fundamental polynomials defined in interval  $[\bar{t}_0, \bar{t}_1)$ ,  $L_j^0(\tau) \equiv L_j(\tau)$  and the time window  $I(t - k\Theta) \equiv I_k(t)$ . Further on, for the sake of certainty, it is also believed that  $\bar{t}_0 \equiv t_{0q} = 0$ .

The interpolation of processes in the multidimensional space represented by a discrete function is done in a similar way, if the formula (9) is applied separately to each vector component. When interpolating the trajectories or boundaries of objects in a multidimensional space, the question arises

of choosing the most natural coordinate system, in which the curve under consideration takes a smoother form or is described by simpler equations.

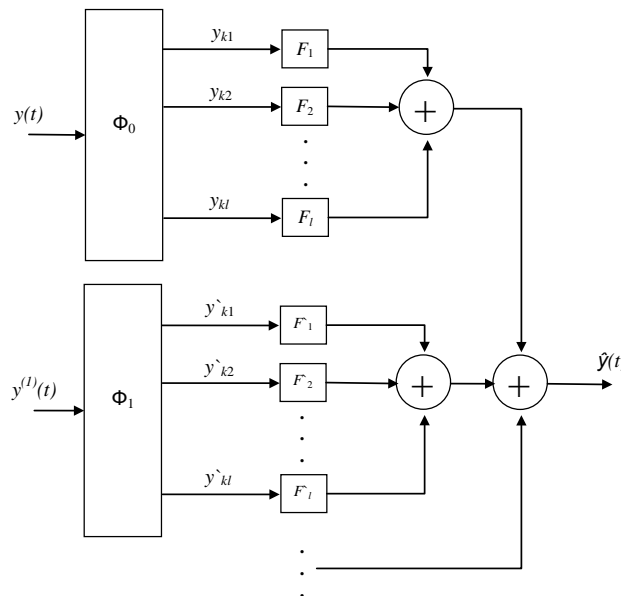
From the above studies and according to [8], it is clear that the use of dot MNAs is fully justified, since this method allows us to obtain the lowest number of coefficients compared to other interpolation methods, such as spline functions, interpolation polynomial Lagrange, etc., with the same value of error.

### 3. Study of acceleration of calculations in the system of specialized computational modules

The expression (9) corresponds to the function of the multichannel data processing system model, which looks like

$$\hat{y}(t) = \sum_{l=1}^l F_l^r y(t) \Phi_0 = \sum_{l=1}^l F_l^r y_{kl}, \quad (10)$$

where  $l$  is a number of specialized calculators with processor function  $F_l^r$ . Structural scheme, which implements the model of multichannel system of processing discrete data sequences, is shown in Figure 1.



**Figure 1.** Multi-channel model of data processing system.

As can be seen from Figure 1 and in accordance with (10), each channel of the data processing system is characterized by the function of a processor  $F_l$ . At the same time, input information in the form of a sequence  $y_{kl}^{(r)}$  on a finite interval of  $N$  samples is used as input actions in these channels, which is obtained as a result of sampling of the process by switching devices  $\Phi_0, \Phi_1, \dots$ . Taking into account that the specified sequences are defined by the expression

$$y_{kl}^{(r)} = y^{(r)}(t_k + \Theta), \quad k = 0, 1, 2, \dots, N-1, \quad (11)$$

where  $r \in \{0, 1, 2\}$ . In the case of equidistant interpolation nodes forming a periodic sequence of reference points with the period  $T$  and taking into account  $\Theta = dT$ , expression (11) takes the form

$$y_{kj}^{(r)} = y^{(r)}((j + dk)T), \quad k = 0, 1, 2, \dots, N-1. \quad (12)$$

where  $d$  is thinning,  $d \in \{0, 1, 2, \dots, N-1\}$ ,

$j$  is sequence offset,  $j \in \{0, 1, 2, \dots, N-1\}$ .

An evenly distributed sequence can be represented depending on offset  $j$  in the form of

$$y_{kj}^{(r)} = \begin{cases} y^{(r)}[d(k-j)T], & j < k, \\ y^{(r)}[dkT], & j = k, \\ y^{(r)}[d(k+j)T], & j > k. \end{cases} \quad k = 0, 1, 2, \dots, N-1. \quad (13)$$

In practical terms, the expressions (12) and (13) describe situations, where the nodes of the interface respectively coincide with the reference points of process  $y(t)$  and its derivatives.

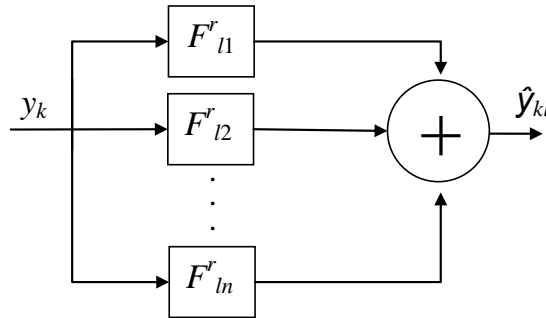
It should also be noted that the  $l$ -th channel of the model of multichannel data processing system with the function of the data generator  $F_l$  is a set of filters  $F_{li}^r(S)$  in the form of

$$F_l = \sum_{i=1}^n F_{li}(S),$$

which form the structural scheme presented in Figure 2. with the transfer function in the form of

$$F_{li}^r(S) = \sum_{j=1}^l \sum_{i=1}^n \frac{1}{S_{li}} (a_{jil}^r - c_{jil}^r e^{-s\Theta}), \quad (14)$$

where  $c$  is series coefficients.



**Figure 2.** Filter block diagram.

After transformation of expression (14) relative to  $S_i$  with regard to the limitations of the equality type

$$r = 0, d = 1, j = 0, c = 0, \quad (15)$$

the solution of determining the acceleration of computational operations depending on the number of specialized calculators and on the volume coefficient of parallel calculations  $a$  is carried out by the formula

$$S_l = \frac{l}{l(1-a) + a}. \quad (16)$$

We can see from expression (16) that the increase of calculation efficiency depends on the algorithm of the task, when restricting the type  $a < 1$ , which allows us to estimate the efficiency of parallelization of the algorithm and the conclusion about the necessary number of specialized calculators [1, 5].

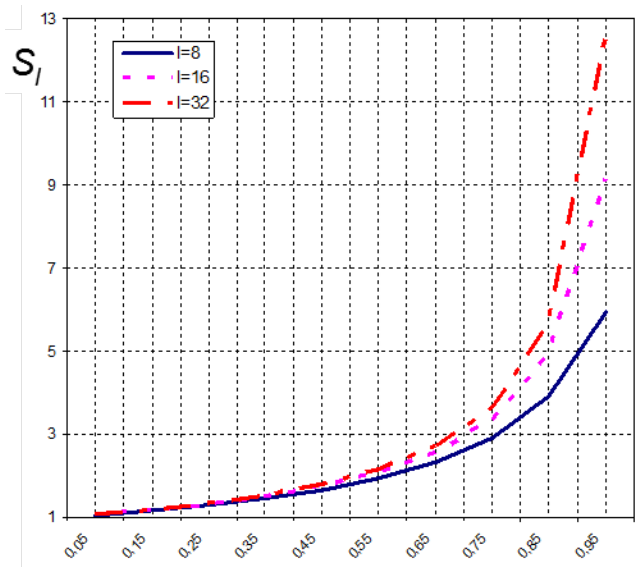
Dependence of the growth of computational operations' acceleration on different number of specialized calculators on the size coefficient of parallelized calculations  $a$  is shown in Figure 3.

Figure 3 shows that the increase of computation speed for the computer system from  $l$  specialized calculators can be obtained at the value of parameter  $a$  according to the condition  $0.25 \leq a \leq 0.75$ .

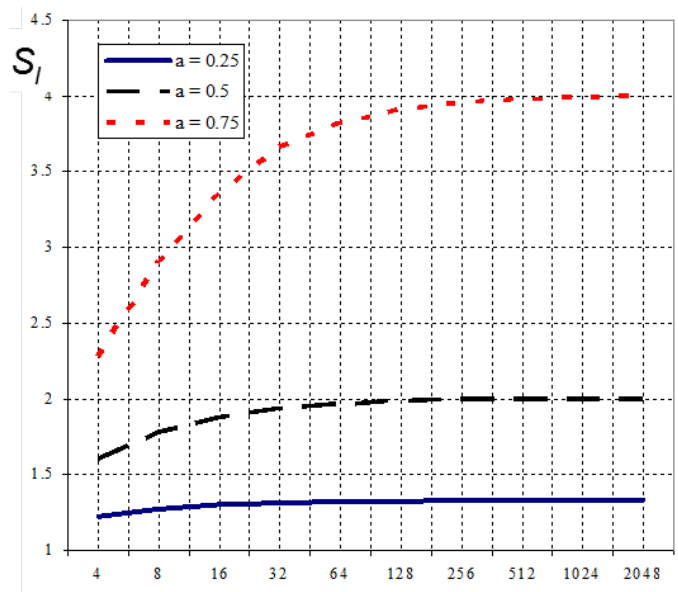
The results of the research of the dependence of the acceleration of calculations on the number of specialized calculators  $l$  are shown in Figure 4.

Figure 4 shows that with the coefficient  $a = 0.75$  in the algorithm, the increase in the number of parallelized cores of specialized calculators up to the value of  $l \geq 16$  leads to a significant increase in

performance, in particular, the use of a single quad-core GPU as a specialized calculator increases the performance by more than 2 times.



**Figure 3.** Dependence of calculations acceleration in the system from  $l$  specialized calculators on the value  $a$ .



**Figure 4.** Dependence of calculations acceleration  $S_l$  in the system from specialized calculators on the value of  $l$ .

The developed model of multichannel data processing system was used for research and development of the algorithm of advanced audio stream mixing in telecommunication systems [10]. Application of the developed algorithm in a heterogeneous computer system reduces the time for data processing to  $0.2226 \times 10^{-3}$  sec. instead of  $1.351 \times 10^{-3}$  sec. – the time of data processing by the base algorithm.

#### 4. Conclusions

Studies have shown that the increase in calculation efficiency depends on the algorithm of the task when limiting the  $a < 1$  type. It allows you to estimate the efficiency of algorithm parallelization and make a conclusion about the necessary number of specialized calculators. Thus, the increase of calculation acceleration for a computer system from 1 specialized calculators can be obtained with the

value of parameter  $a$  in accordance with the condition  $0,25 \leq a \leq 0,75$ . At the same time, the increase of the number of parallelized cores of specialized calculators up to the value of  $l = 16$  at the value of the coefficient  $a = 0,75$  in the algorithm leads to a significant increase in performance, but with a further increase up to  $l > 256$  it does not bring any significant results, which is consistent with the main provisions of Amdal and Graham's researches.

## 5. References

- [1] Bakhvalov N S and Voevodin V B 2005 *Modern problems of computational mathematics and mathematical modeling T. 1: Computational mathematics* (Moscow: Science) p 342
- [2] Kulbak S 1967 *Information theory and statistics* (Moscow: Science) p 408
- [3] Ljung L 1991 *System identification. Theory for the user* (Moscow: Science) p 432
- [4] Brillinger D R 1990 A study of second- and third-order spectral procedures and maximum likelihood in the identification of bilinear systems *IEEE Trans. on Acoustics, Speech and signal processing* **38(7)** 1238-1245
- [5] Pupkov K A, Kapalin and V I Yushchenko A S 1976 *Functional rows in non-linear systems theory* (Moscow: Science) p 448
- [6] Bakhvalov N S, Zhidkov N P, and Kobelkov G M 2008 *Numerical methods* (Moscow: BINOM Knowledge Lab) p 640
- [7] Helman, D R and JaJa J 1999 Designing Practical Efficient Algorithms for Symmetric Multiprocessors *Lecture Notes in Computer Science, International Workshop ALENEX'99* **1619** 37-56
- [8] Formalev V F and Reviznikov D L 2004 *Numerical methods* (Moscow: FISMATLYT) p 400
- [9] Kropotov Y A and Proskuryakov A Y 2007 Mathematical model of probability law for the amplitudes of speech waveforms in the exponential basises *17th International Crimean Conference - Microwave and Telecommunication Technology, CRIMICO* 364-366
- [10] Kolpakov A A and Kropotov Y A 2017 Advanced mixing audio streams for heterogeneous computer systems in telecommunications *CEUR Workshop Proceedings* **1902** 32-36
- [11] Kropotov Y A, Belov A A and Proskuryakov A Y 2018 Method for forecasting changes in time series parameters in digital information management systems *Computer Optics* **42(6)** 1093-1100  
DOI: 10.18287/2412-6179-2018-42-6-1093-1100