

ABOUT AN ALGORITHM OF FUNCTION APPROXIMATION BY THE LINEAR SPLINES

B.BAYRAKTAR¹, V.KUDAEV², §

ABSTRACT. The actual application for the problem of best approximation of grid function by linear splines was formulated. A mathematical model and a method for its solution were developed. Complexity of the problem was that it was multi - extremal and could not be solved analytically. The method was developed in order to solve the problem of dynamic programming scheme, which was extended by us. Given the application of the method to the problem of flow control in the pressure-regulating systems, the pipeline network for transport of substances (pipelines of water, oil, gas, and etc.) that minimizes the amount of substance reservoirs and reduces the discharge of substance from the system. The method and the algorithm developed here may be used in computational mathematics, optimal control and regulation system, and regressive analysis.

Keywords: grid functions, the best approximation, minimal deviation, linear splines, dynamic programming, optimal regulation.

AMS Subject Classification: 65D15, 49L20, 49J35

1. INTRODUCTION

As it is known functions may be approximated with linear splines. It is one of the main directions in computational mathematics and has a number of applications in systems and process modeling.

Problems of asymptotically optimal nodes where splines may be coupled with selection and determine the order of the best spline approximations with increasing number of nodes are solved for different kinds of splines [1-4]. Optimal nodes selection not only decreases approximation error but also allows making qualitative conclusions on behavior in modeling process. For instance, it helps to determine the transition point in researches of transitional process. However, until now, algorithms of optimal choice of gluing unit splines have not been developed.

Linear splines are used in systems of robotized design, optimal control and regulation in systems, and statistics [5, 6]. But there is no algorithm for optimal nodes selection to glue splines together though it is the simplest kind of splines. The algorithm to choose optimal nodes for interpolation type problem, i.e., the case when nodes of linear spline coupling are set in grid function diagram, is represented in the article [7]. The method

¹ Uludag University, Education Faculty, Gorukle Campus, Bursa, Turkey.
e-mail: bbayraktar@uludag.edu.tr;

² Institute of Computer Science and Problems of Regional Management of Kabardino-Balkar Scientific Centre of RAS, Russia.
e-mail: iipru@hotmail.com;

§ Manuscript received: February 07, 2016; accepted: May 05, 2016.

TWMS Journal of Applied and Engineering Mathematics, V.6, No.2; © Işık University, Department of Mathematics, 2016; all rights reserved.

of solution to the problem of the best approximation for grid function with linear splines with minimal deviation is presented in our article.

2. FORMULATION OF THE PROBLEM

The problem. Let us give the discrete set $T = \{t_i \in R \mid t_i < t_{i+1}, i = 0, 1, 2, \dots, n-1\}$ and function $f: T \rightarrow R$. It requires function $f(t)$ approximate by piecewise linear function $l_m(t, x_0, y_0, \dots, x_m, y_m)$ with a given number of lines segments $m \leq n$, where $x_k \in T$, $y_k \in R$, for $k=0, 1, \dots, m$ desired coordinates of its nodes, $t_0=x_0 < \dots < x_m=t_n$, so that the deviation l_m from f must be minimal:

$$\rho(t_i, x_0, y_0, \dots, x_m, y_m) = \max_{0 \leq i \leq n} |f(t_i) - l_m(t_i, x_0, y_0, \dots, x_m, y_m)| \rightarrow \min \quad (1)$$

Let us prove the existence of the decision (1).

There is finite number of sets (x_0, x_1, \dots, x_m) as the scale t is discrete and $\forall i \in (0, m)$, $\exists k \in (0, n)$, $x_i = t_k$. Let $(\hat{x}_0, \hat{x}_1, \dots, \hat{x}_m)$ be any among any sets, and $\hat{\rho}$ is deviation l_m from f for the set, for instance, $\hat{y}_0 = f(\hat{x}_0)$, $\hat{y}_1 = f(\hat{x}_1), \dots, \hat{y}_m = f(\hat{x}_m)$. Then it is evident, at optimal decision of the problem (1), if it is exist, $\rho_{opt} \leq \hat{\rho}$. Thus, the problem (1) is equivalent to this one:

$$\begin{cases} \rho = \max_{0 \leq i \leq n} |f(t_i) - l_m(t_i, x_0, y_0, \dots, x_m, y_m)| \rightarrow \min \\ \rho \leq \hat{\rho} \end{cases} \quad (2)$$

Now let $(\hat{x}_0, \hat{x}_1, \dots, \hat{x}_m)$ be an arbitrary fixed set. Multiplicity of values y_0, y_1, \dots, y_m that satisfies

$$f(\hat{x}_j) - \hat{\rho} \leq y_j \leq f(\hat{x}_j) + \hat{\rho}, j = 0, 1, \dots, m \quad (3)$$

is limited and covered as it represents m dimensional cube, and the function (1) is continuous in it. That is why the problem (2) on this set has the decision or its multiple solutions are empty. Having enumerated all the possible sets (x_0, x_1, \dots, x_m) and having compared decisions on those among them where multiplicity of decisions is not empty, we will find optimal decision for initial problem.

3. THE PROBLEM SOLUTION ALGORITHM

We will now give definitions of some terms and concepts that are used in the article.

Definition 3.1. Vertical or layer is called an ordered set of numbers of possible discrete values of the function at a given point.

Definition 3.2. The smallest possible deviation of a given function from a straight line is called the potential or an error.

Definition 3.3. An arc is called an ordered pair of numbers of vertices, respectively, from the previous and subsequent layers, the relationship between which is possible.

The solution of the problem consists of two phases in each of which the constructed expanded scheme of the Dynamic Programming (DP) is used in a special way. In contrast to the standard scheme DP in the proposed scheme, the intersection of points on the verticals is not empty.

At the first phase we solve the problem of the best interpolation of the given function. This decision is made for the initial solution of a problem of approximation. At the second stage, verticals are built in the received optimum interpolation points and the problem of the best approximation of function is solved.

The detailed description of the solution algorithms of the problem is given below.

3.1. Algorithm of the solution of a problem of interpolation. For convenience, we introduce the following notation:

- $y(t)$ – straight line equation passing through two given points $(i, f(i))$ and $(q, f(q))$;
 - δ_{iq} – maximum deviation of a line segment connecting the points with coordinates $(i, f(i))$ and $(q, f(q))$ from a given function, where $i = 0, 1, 2, \dots, n - 1, q = i + 1, i + 2, \dots, n$;
 - P_q^j – potential q vertex j layer, where $j = 1, 2, \dots, m, q = i + 1, i + 2, \dots, n$;
 - G_q^j – discrete set of vertices of the previous $(j - 1)$ layer incident to q vertex of j layer;
- Then,

$$y(t) = \frac{(t - i)f(q) - (t - q)f(i)}{i - q} \tag{4}$$

$$\delta_{iq} = \max_{\forall t \in [x_i, x_q]} |f(t) - y(t)| \tag{5}$$

Then Bellman-Dijkstras recurrent process [8] for points potentials calculation in layers occurs:

$$P_q^{j+1} = \min_{i \in G_q^j} \max\{P_i^j, \delta_{iq}\}, q = 0, 1, 2, \dots, m - 1 \tag{6}$$

1. We set the number of linear splines - m .
2. For all i and q let us calculate errors δ_{iq} forming a triangular Matrix of Errors (ME) of interpolation:

$$\begin{pmatrix} \delta_{01} & \delta_{02} & \delta_{03} & \delta_{04} & \dots & \delta_{0n} \\ & \delta_{12} & \delta_{13} & \delta_{14} & \dots & \delta_{1n} \\ & & & & \dots & \dots \\ & & & & \delta_{n-2,n-1} & \delta_{n-2,n} \\ & & & & & \delta_{n-1,n} \end{pmatrix}$$

3. Using this matrix we develop the scheme DP.
 - 3.1 For this purpose in the 1st layer of the scheme, all elements are marked all elements (that are led round by circle, see figure 1) accessible from the zero vertex, that is vertex corresponding to the 1st line ME. In the 2nd layer, all those vertices are placed, which are reachable at least from one marked vertex of the first layer and so on. In the last layer, all those vertices are placed, which are reachable from the marked vertices of a penultimate layer and also those vertices from which the last vertex N is achievable.

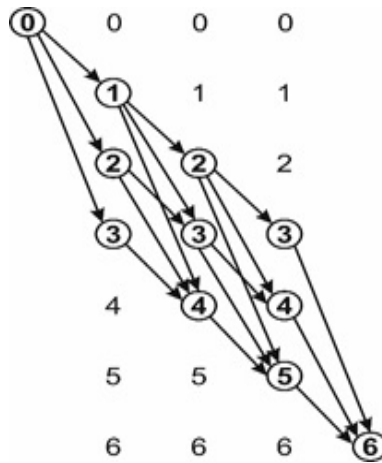


Fig. 1. Scheme DP with $n = 6, m = 4$.

3.2 Affixing corresponding weights (errors) to the arcs connecting the vertices of adjacent layers.

4. Potential of zeros vertex, apparently, is equal to zero.
5. We count potentials of the subsequent layer on the basis of the potentials of vertices of the previous layer and errors corresponding to the arches connecting the previous and subsequent layers on formula (6). The arc corresponding to value P_p^{j+1} is remembered. Calculations of potentials proceed until we reach the last, m layer with n vertex.
6. After that we move from the last vertex by the beginning the column DP on the allocated arcs, remembering the vertices corresponding to them. The sequence of these vertices also give the best interpolation of the given function to the linear splines with a minimum potential $\hat{\delta} = P_{n-1}^m$.

3.2. Algorithm of the solution to a problem of approximation. We take result of the solution to a problem of the best interpolation for the initial solution to a problem of approximation. To solve this problem, we do not need to build a matrix of error. Formation of the best path from the starting point to the end was made sequentially, starting with the first layer, then the second, etc. up to and including the last layer.

Let us enter the following designations:

k – predetermined positive integer;

s – integer, and $s = -k, \dots, -1, 0, 1, \dots, +k$;

j – number of a layer and $j = 0, 1, \dots, m - 1$;

$\hat{\delta}$ – interpolation error;

$f(x_j + s\hat{\delta}) = f_s^j$; for example $f(x_3) - 2\hat{\delta} = f_{-2}^3$

(x_j, f_s^j) – coordinates of points which numbers form verticals of each of m of layers of the scheme DP;

$y_{is}^j(t)$ – equation of a line segment connecting the i -th point of j -th layer and s -th point $(j + 1)$ -th layer;

δ_{is}^j – maximum deviation of a straight line segment $y_{is}^j(t)$ from a function graph $f(t)$ on a piece $[x_j, x_{j+1}]$.

P_s^j – potential of s -th point, that is, the point (x_j, f_s^j) of j -th layer.

Then

$$y_{is}^j(t) = \frac{f_s^{j+1} - f_i^j}{x_{j+1} - x_j} t - \frac{f_i^{j+1} x_j - f_s^j x_{j+1}}{x_{j+1} - x_j} \quad (7)$$

$$\delta_{is}^j = \max_{\forall t \in [x_j, x_{j+1}]} |f(t) - y_{is}^j(t)| \quad (8)$$

Potential of i – th point of $(j + 1)$ –th layer, that is, the point (x_i, f_i^{j+1}) , will be

$$P_i^{j+1} = \min_{-k \leq s \leq k} \max\{P_s^j, \delta_{is}^j\} \quad (9)$$

1. We take the best solution to a problem of interpolation for the initial decision:

$x_0 \rightarrow x_1 \rightarrow \dots \rightarrow x_m \rightarrow x_n$ with a margin error $\hat{\delta}$;

2. We form scheme DP layers which each layer contains points (x_i, f_s^i) .

3. Sequentially we calculate potentials of vertices of layers.

3.1 Potential of vertices of the 1st layer – are values δ_{si}^0 – that is, evasion of straight lines connecting a point x_0 with vertices of the 1st layer $(s, i = -k, \dots, +k)$.

3.2 Let all the potentials be already calculated P_s^j , vertices of j -th layer $(s = -k, \dots, +k)$.

3.3 Calculation of potentials of $(j + 1)$ -th layer is made sequentially by formulas (7) (9) for all i and s . Except the potential, each vertex is supplied with number of vertex of i

of the previous layer proceeding from, which potential i -th vertex of $(j + 1)$ -th layer was received.

4. In the last layer we choose vertex with a minimum potential $p = \min_{v_s} \{P_s^m\}$, which gives the optimum path from the last layer into the first layer. On this trajectory nodes define the best approximation of a given function.

4. RESULTS OF NUMERICAL EXPERIMENT OF THE PROBLEMS OF INTERPOLATION AND APPROXIMATION OF FUNCTION

The developed algorithms are realized in programming language C++ in the environment of Builder 6.0. On the example of periodic function of $f(x) = \sin x$ on an interval $[0, 2\pi]$ numerical experiments of interpolation and approximation for $n = 72$ and $m = 13$ are made. Results of numerical experiment are presented on screen forms (Fig. 2 and 3) and are summarized in Table 1-3.

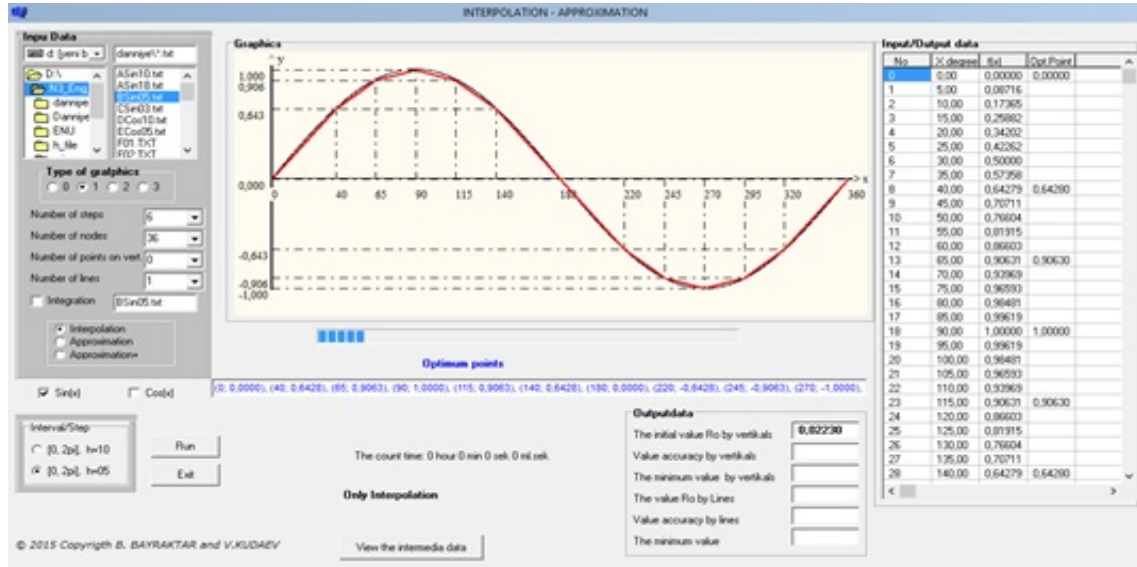


Fig. 2. Screen shot for optimal interpolation
 Displayed results: $n=73$, layers quantity $m=13$, lines quantity-12, minimal deviation $\hat{\delta} = 0.02230$.

TABLE 1. Optimal coordinates of the interpolation

N	α (deg)	$\sin(\alpha)$	x_j	N	α (deg)	$\sin(\alpha)$	x_j
1	0	0.0000	0.0000	8	220	-0.6428	-0.6428
2	40	0.6428	0.6428	9	245	-0.9063	-0.9063
3	65	0.9063	0.9063	10	270	-1.0000	-1.0000
4	90	1.0000	1.0000	11	295	-0.9063	-0.9063
5	115	0.9063	0.9063	12	320	-0.6428	-0.6428
6	140	0.6428	0.6428	13	360	0.0000	0.0000
7	180	0.0000	0.0000				

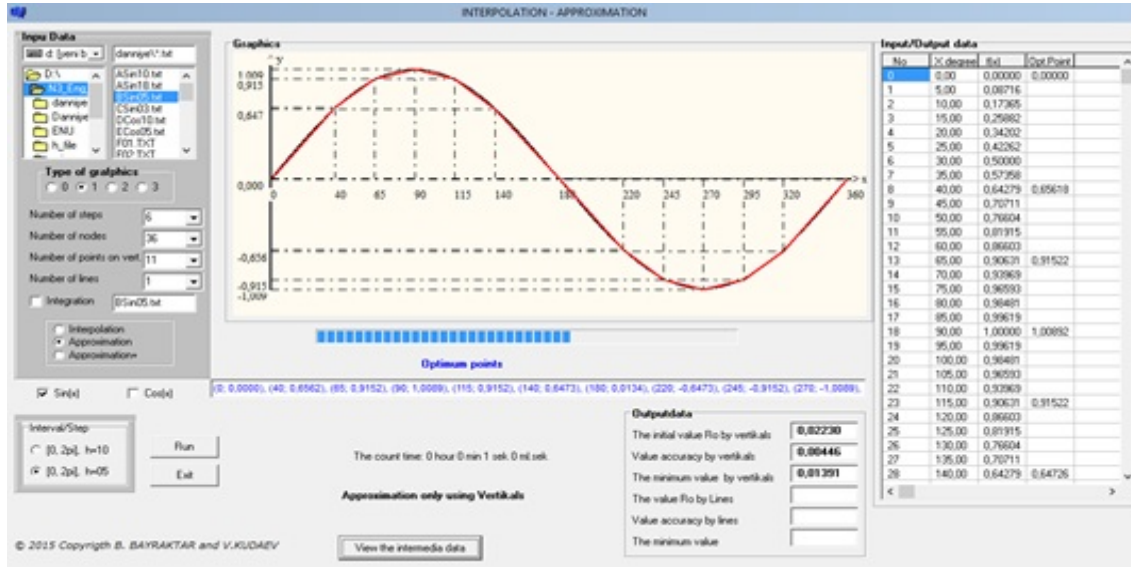


Fig. 3. Screen shot for optimal approximation

Displayed results: $n=73$, layers quantity $m=13$, lines quantity-12, initial value $\hat{\delta} = 0.02230$, minimal deviation 0.01391.

TABLE 2. Optimal coordinates of the approximation for $k = 5$

N	α (deg)	$\sin(\alpha)$	x_j	N	α (deg)	$\sin(\alpha)$	x_j
1	0	0.0000	0.0000	8	220	-0.6428	-0.6473
2	40	0.6428	0.6562	9	245	-0.9063	-0.9152
3	65	0.9063	0.9152	10	270	-1.0000	-1.0089
4	90	1.0000	1.0089	11	295	-0.9063	-0.9152
5	115	0.9063	0.9152	12	320	-0.6428	-0.6562
6	140	0.6428	0.6473	13	360	0.0000	0.0000
7	180	0.0000	0.0134				

TABLE 3. Results of approximation at various quantities of points on verticals (here $\hat{\delta} = 0.02230$ and received from interpolation)

Quantity points on the verticals(2k+1)	δ_{opt}	Computing Time(sec)
3	0.02228	0
5	0.01503	0
7	0.01487	0
9	0.01503	0
11	0.01391	1
13	0.01487	1
15	0.01423	1

5. OPTIMIZATION OF THE FLOW CONTROL IN THE PRESSURE-REGULATING SYSTEM

In pipeline systems, an obligatory element is the tanks used for storage, clearing and further transportation of substance (water, gas and oil pipelines) and pump stations. The regulating capacity of tanks depends only on the hourly schedule of consumption and the schedule of work of pump stations of system. The working principle of the system consists of the followings; substance at pump station insufficiently given to the pipeline, deficiency

of substance moving in system from the tank, and surplus giving from pump station - surplus arrival in the tank.

One needs to minimize the volume of the reservoir for a given number of switching pump stations and substance consumption schedule. When the optimal solution of the problem is reached, the amount of substance that consumers receive hourly is very close to the required level.

In the article [9] the problem was solved by local variations method and the solution was specified in the vicinity of the locally optimal trajectory. The method provided in the article [10] is in the following. The descent to the optimum is carried out simultaneously from several initial solutions built by direct search on the rough grid. Then construct an iterative process whereby "narrowing of neighborhoods" is clarified, each of these solutions with the prescribed accuracy. Among these solutions, the best solution is chosen. The method is effective in a small number of segments.

Let $g = g(t)$ be a discrete step-function of time the setting consumption of amount of substance by system in each hour days percentage of total of substance, then accumulation of substance in the tank can be expressed by integral function:

$$f(t) = \int_0^t g(x)dx, \text{ where } t = 1, 2, 3, \dots, 24.$$

For the solution of a problem of minimization of volumes of tanks it is required to approximate integral function of $f(t)$ the given quantity of the linear splines the least evading from function $f(t)$ and to construct a step-function. This step-function will also be the schedule of work of the pump station.

The method provided here allows us to solve the problem with accuracy (defect) δ without restrictions on number of branches.

The screenshot (Fig. 4) and Table 5 provide information for management manager of the pipeline system. There is a recommendation for (optimal) 5-mode schedule pumping station (thin line) and a maximum of 24-hour substance supply schedule (bold line) in its center:

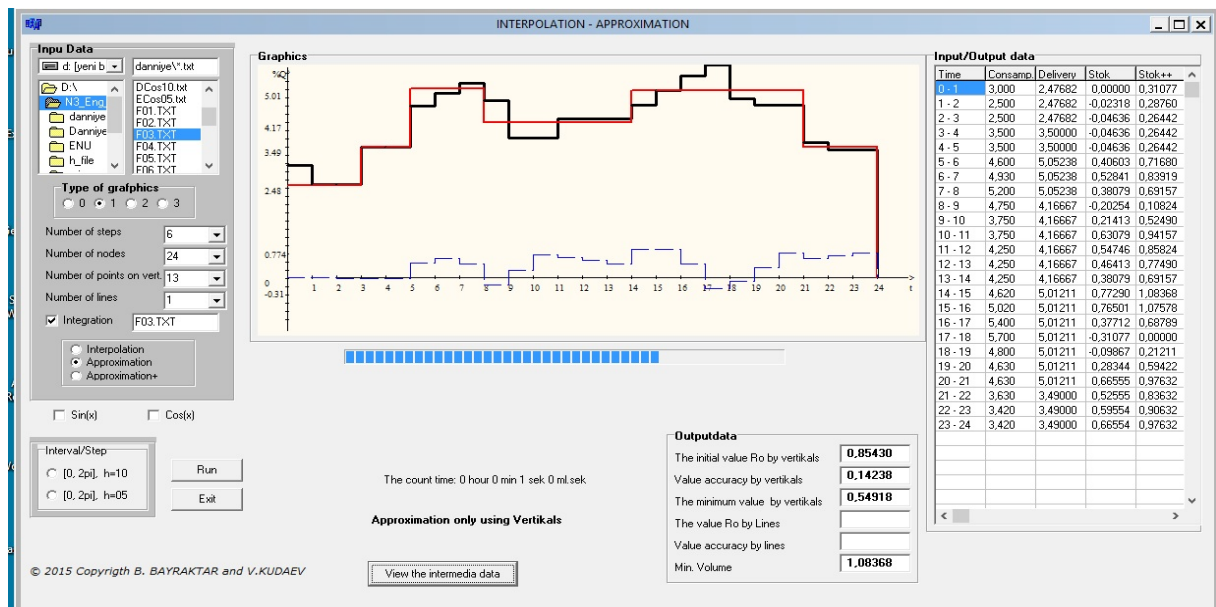


Fig.4 Screen shot to determine the optimal 6 - step schedule pumping station

In the table of graph to the right: Consumption by an hour a day (percentage of total consumption), Stock change by an hour a day, Stock ++ -total accumulation. The minimum volume of the regulatory 1.08368% from the total consumption.

TABLE 4. Information of control system for pipeline system manager

The period of time (h)	0–3	3–5	5–8	8–14	14–21	21–24
Delivery for network(%)	2.47682	3.50000	5.05238	4.16667	5.01211	3.49000

6. CONCLUSION

We formulated the actual for application a mathematical model of the problem of best approximation of the grid functions by linear splines and the method of its solution.

To solve the problem, we developed a method based on the idea of the DP method. The expanded scheme DP, which differed from the reference scheme DP in that the point set on verticals could not be empty, was offered.

The problem was solved in two stages: At the first stage, the problem of the best interpolation of function was solved. At the second stage, the problem of the best approximation was solved. The problem of approximation for the initial solution to the problem was taken as a base for the solution to the interpolation problem.

The application was implemented to the problem of control optimization of flows in the pressure-regulating systems, supply and consumption of substances.

7. ACKNOWLEDGMENT

Researched with the support of Russian Fond of Basic Researches grants 13-01-00929-a, 13-07-01002-a, 15-01-05844-a which is gratefully acknowledged.

REFERENCES

- [1] Boor,K., (1985), A practical guide to splines, Radio and Communication.
- [2] Grebennikov,A.I., (1976), The choice of nodes in the spline approximation of functions., Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 16 (1), pp. 219-223.
- [3] Ligun,A.A. and Shumeiko,A.A., (1997), Asymptotic methods for rebuilding curves, Institut Matematiki NAN Ukrainy, Kiev.
- [4] Walters,H.J., (2004), A Newton-type method for computing best segment approximations, Communications on Pure and Applied Analysis, 3 (1), pp. 133-149.
- [5] Shumeiko,A.A. and Shumeiko,E.A., (2011), On the construction of asymptotically optimal piecewise-linear regression, Informatics and Mathematical Methods in Simulation, 1 (2), pp. 99-106.
- [6] Thamaratnam,K., Claeskens,G., Croux,C. and Salibian-Barrera,M., (2010), S-estimation for penalised regression splines, Journal of Computational and Graphical Statistics, 19 (5), pp. 609-625.
- [7] Pakhnoutov,I.A., (2011), Vybor uzlov sglazhivaniya lineynymi splaynami, Izvestia Kaliningradskogo Tekhnicheskogo Universiteta, 23, pp. 122-126,
- [8] Christofides,N., (1978), Graph theory. An algorithmic approach, Mir, Moskva.
- [9] Kudaev,V.C. and Bayraktarov,B.R., (2013), Regulation optimization in network systems., News of Kabardin-Balkar Scientific Center of Russian Academy of Sciences, 6 (56), pp. 33-38.
- [10] Bayraktar,B., (2013), Problem of constructing a step function fluctuating least around a given function, TWMS J. Pure Appl. Math., 4(2), pp. 131-145.



Bahtiyar BAYRAKTAR graduated from the Faculty of Mathematics of Kabardino - Balkarian State University (USSR) in 1976. From 1977 to 1989 he worked in the Scientific Research Institute of Applied Mathematics and Mechanics(Nalchik, Russia). In 1992 he got his Ph.D. degree from Rostov State University. His research areas are calculus of variations and optimal control. He is an Assistant of Prof. of Pedagogical Faculty of Uludag University in Turkey.



Kudaev Valery Cherimovich graduated from the Faculty Electronic Engineering of Novocherkassk Polytechnic Institute in 1969. From 1977 to 1989 he worked in the Scientific Research Institute of Applied Mathematics and Mechanics(Nalchik, Russia). In 1993 he got his Ph.D. degree from Rostov State University. He is a reseacher in the Institute of Computer Science and Problems of Regional Management of the Kabardin-Balkar Scientific Center of the Russian Academy of Sciences. His research interests are mathematical modeling and optimization of network systems.
