Innovative Method to Improve of Experimental Data Approximation Accuracy

P.S. Vasiliev[a], A.A. Shagarova[b] and L.A. Ilina[c]
Volgograd State Technical University, Lenin Ave., 28, Volgograd, Russia
nestorvv@mail.ru, shagarang@mail.ru, vgtupaph@gmail.com

Keywords: Mathematical model, Least squares method, Optimization, Influencing factors, Approximation accuracy, Computing power.

Abstract: The paper deals with the experimental data approximation accuracy improvement to obtain stochastic mathematical models for various fast-flowing processes control. The main advantages and disadvantages of the known improving the approximation accuracy methods are considered. The mathematical foundations of the modified least squares method, which allows determining the regression coefficients values as a result of the optimization problem solving – minimizing the objective function, represented as the sum of experimental and the theoretical function's values absolute deviations are described in detail. The mathematical foundations of the modified least squares method, which allows determining the regression coefficients values as a result of the optimization problem solving – minimizing the objective function, represented as the sum of experimental and the theoretical function's values absolute deviations are described in detail. In addition to the increased of experimental data the approximation accuracy, the proposed method for any number influencing factors makes it possible to determine the regression coefficients values at a constant cycles number of their calculation, which considerably saves computing power.

1 INTRODUCTION

The practical value of mathematical models describing various phenomena in science, technology and human life is determined by the accuracy of their equations. This circumstance is especially relevant for real-time management responsible processes, including fast-flowing physical and biological processes in electronics, radio engineering, medicine, etc.

For example, the quality of various tracking systems operation, determined mainly by stochastic mathematical models, depends on the approximation accuracy of experimental data performing in real time.

In electronics and radio engineering, this makes it possible to increase the transmitted information reliability and guarantee various technical systems automatic control without human intervention, which increases their competitiveness (Oishi, 2006; Raisch, 1999).

In medicine, this makes it possible to perform endoscopic operations on the heart without stopping it with the help of a robot surgeon who moves the scalpel in time with heart contractions, which increases the survival rate of patients and reduces their rehabilitation period (Harky, 2020).

It should be noted that from a practical point of view, the most valuable is always the simplest result. In this regard, theoretical results obtained with the help of complex deterministic mathematical models are often first presented in the form of tabular data (numerical experiment data), which are then approximated in simpler terms.

Thus, improving experimental data approximation accuracy is an urgent task of mathematical modeling, which is of great practical importance.

One relatively simple method of improving experimental data approximation accuracy is to increase the member's number of the approximating function (in particular, increasing the degree of its polynomial).

[a] https://orcid.org/0000-0003-0262-686X
[b] https://orcid.org/0009-0009-0057-9263
[c] https://orcid.org/0000-0002-3119-0738

However, in this case, the function is often presented as a bulky and inconvenient expression for practical use.

A similar method is an experimental data array division into several parts, each of which is approximated by a separate function. However, this increases the number of approximating expressions, which reduces the practical value of this method (Petryanin, 2016).

The refined approximation method, performed using the method of least squares (Petryanin, 2016) and allowing reducing the average relative error by about 10 % of its value for classical approximation is more advanced.

According this method, first, for the initial array of experimental data (x_i; y_i), i=1…n the coefficients and values of the approximating function $f_i(x)$ are determined using the method of least squares.

Then the specified values of the random variable are calculated using formulas:

$$\begin{cases} y_i^* = y_i + |y_i - f_i(x)| + \left|\dfrac{y_i}{f_i(x)}\right|, & y_i > f_i(x) \\ y_i^* = y_i - |y_i - f_i(x)| - \left|\dfrac{y_i}{f_i(x)}\right|, & y_i < f_i(x). \\ y_i^* = y_i, \quad y_i = f_i(x) \end{cases} \quad (1)$$

Next, according to least squares method, the precise coefficients and values of the approximating function $f_i^*(x)$ are determined for the data array $(x_i; y_i^*)$.

In case of technical or budgetary constraints, in practice it is managed to obtain a relatively small experimental data array then in order to improve their approximation accuracy, the method of repeated sampling is used, which is formed artificially based on the condition of an equally likely materialization of the observed phenomenon (Walter, 2014). This method has found wide application in biotechnology.

The most complex methods for increasing experimental data approximation accuracy include computational optimization methods based on perturbation theory using (Kuz'min, 2017) and on the artificial neural networks apply (Meade, 1993).

In spite of the fact that these methods allow to guarantee maximum experimental data approximation accuracy, from the viewpoint of computing power using they are extremely costly and cannot currently form the basis of mathematical models for controlling fast-flowing processes.

2 METHODS

An approximating function obtaining is one of the main regression analysis tasks. Therefore, it is first necessary to confirm the hypothesis of a random variable normal distribution using the Shapiro-Wilk consent criteria (Shapiro, 1965) and Epps-Palli (Epps, 1983), which have the greatest statistical power.

Afterward it is necessary to exclude sharply standout values from the experimental data array (x_i; y_i), i=1…n using the Grubbs criterion (Grubbs, 1950; Kobzar, 2006). Then, you need to choose the approximating function's type (regression equations) and present the experimental data in linearized form (X_i; Y_i).

Further mathematical processing is carried out using the modified least squares method in accordance with the condition:

$$\min\left[\sum_{i=1}^{n}\left(\frac{Y_i - F_i(X)}{Y_i^p}\right)^2\right], \quad (2)$$

where $F_i(X)$ – linearized values of the approximation function (regression equations); p – coefficient that determines the regression line position in the k-dimensional factor space.

For example, a power-law approximating function (regression equation) is considered in the form:

$$f(x) = a \cdot x_1^b \cdot x_2^c,$$

where k=2. It will be presented in linearized form, as:

$$F(X) = A + b \cdot X_1 + c \cdot X_2,$$

where $F(X)$=ln [$f(x)$]; A=ln a; X_1=ln x_1; X_2=ln x_2.

To determine the values of the regression coefficients a, b, c according to condition (2), the following transformations must be performed:

$$\frac{\partial}{\partial A}\sum_{i=1}^{n}\left(\frac{Y_i - F_i(X)}{Y_i^p}\right)^2 =$$

$$= \sum_{i=1}^{n}\left[2\left(\begin{matrix} Y_i^{1-p} - AY_i^{-p} - \\ -bX_{1i}Y_i^{-p} - cX_{2i}Y_i^{-p} \end{matrix}\right)\left(-Y_i^{-p}\right)\right] = 0,$$

$$\frac{\partial}{\partial b}\sum_{i=1}^{n}\left(\frac{Y_i - F_i(X)}{Y_i^{\,p}}\right)^2 =$$

$$= \sum_{i=1}^{n}\left[2\begin{pmatrix} Y_i^{1-p} - AY_i^{-p} - \\ -bX_{1i}Y_i^{-p} - cX_{2i}Y_i^{-p}\end{pmatrix}\left(-X_{1i}Y_i^{-p}\right)\right] = 0,$$

$$\frac{\partial}{\partial c}\sum_{i=1}^{n}\left(\frac{Y_i - F_i(X)}{Y_i^{\,p}}\right)^2 =$$

$$= \sum_{i=1}^{n}\left[2\begin{pmatrix} Y_i^{1-p} - AY_i^{-p} - \\ -bX_{1i}Y_i^{-p} - cX_{2i}Y_i^{-p}\end{pmatrix}\left(-X_{2i}Y_i^{-p}\right)\right] = 0.$$

Then the linear algebraic equations system is solved:

$$\begin{cases} A\sum\limits_{i=1}^{n}Y_i^{-2p} + b\sum\limits_{i=1}^{n}X_{1i}Y_i^{-2p} + c\sum\limits_{i=1}^{n}X_{2i}Y_i^{-2p} = \\ = \sum\limits_{i=1}^{n}Y_i^{1-2p}, \\ A\sum\limits_{i=1}^{n}X_{1i}Y_i^{-2p} + b\sum\limits_{i=1}^{n}X_{1i}^2Y_i^{-2p} + c\sum\limits_{i=1}^{n}X_{1i}X_{2i}Y_i^{-2p} = \\ = \sum\limits_{i=1}^{n}X_{1i}Y_i^{1-2p}, \\ A\sum\limits_{i=1}^{n}X_{2i}Y_i^{-2p} + b\sum\limits_{i=1}^{n}X_{1i}X_{2i}Y_i^{-2p} + c\sum\limits_{i=1}^{n}X_{2i}^2Y_i^{-2p} = \\ = \sum\limits_{i=1}^{n}X_{2i}Y_i^{1-2p}. \end{cases}$$

To determine the value corresponding to the maximum accuracy of experimental data approximation, unconditional nonlinear optimization (minimization) of the objective function is carried out:

$$\frac{1}{n}\cdot\sum_{i=1}^{n}\Delta_i = \frac{1}{n}\cdot\sum_{i=1}^{n}\left|Y_i - F_i(X)\right| = f(p). \qquad (3)$$

For example, while searching the minimum of the objective function (3) with the dichotomy method on the interval $p \in [p_1;\ p_2]$, in the first step its values at two points are calculated (Gill, 1981):

$$p_{\min 1} = \frac{p_1 + p_2}{2} - \frac{\varepsilon}{4},$$

$$p_{\max 1} = \frac{p_1 + p_2}{2} + \frac{\varepsilon}{4},$$

where ε – solving error problem's optimization.

Then the uncertainty interval is constricted as follows:

$$\begin{cases} a_1^* = p_1, & \sum\limits_{i=1}^{n}\Delta_{1i} < \sum\limits_{i=1}^{n}\Delta_{2i} \\ a_1^* = p_{\min 1}, & \sum\limits_{i=1}^{n}\Delta_{1i} \ge \sum\limits_{i=1}^{n}\Delta_{2i} \end{cases},$$

$$\begin{cases} b_1^* = p_{\max 1}, & \sum\limits_{i=1}^{n}\Delta_{1i} < \sum\limits_{i=1}^{n}\Delta_{2i} \\ b_1^* = p_2, & \sum\limits_{i=1}^{n}\Delta_{1i} \ge \sum\limits_{i=1}^{n}\Delta_{2i} \end{cases}.$$

From the second step, the optimization problem's solution is defined by the following expressions:

$$p_{\min j} = \frac{a_{j-1}^* + b_{j-1}^*}{2} - \frac{\varepsilon}{4},$$

$$p_{\max j} = \frac{a_{j-1}^* + b_{j-1}^*}{2} + \frac{\varepsilon}{4},$$

$$\begin{cases} a_j^* = a_{j-1}^*, & \sum\limits_{i=1}^{n}\Delta_{1i} < \sum\limits_{i=1}^{n}\Delta_{2i} \\ a_j^* = p_{\min j}, & \sum\limits_{i=1}^{n}\Delta_{1i} \ge \sum\limits_{i=1}^{n}\Delta_{2i} \end{cases},$$

$$\begin{cases} b_j^* = p_{\max j}, & \sum\limits_{i=1}^{n}\Delta_{1i} < \sum\limits_{i=1}^{n}\Delta_{2i} \\ b_j^* = b_{j-1}^*, & \sum\limits_{i=1}^{n}\Delta_{1i} \ge \sum\limits_{i=1}^{n}\Delta_{2i} \end{cases}.$$

where $j=2\ldots m$.

The steps number is calculated by the equation:

$$m \approx \log_2\left(2\cdot\frac{p_2 - p_1}{\varepsilon} - 1\right).$$

After complete m steps, the desired solution to the optimization problem is obtained:

$$p_m = \frac{a_m^* + b_m^*}{2}.$$

3 RESULTS

Let's approximate a small arbitrary data massive with one influencing factor (k=1) (Table 1).

The hypothesis of the normal random variable distribution y_i for the significance level of 0.05 is confirmed by the fact that the calculated value of the Shapiro-Wilke consent criterion is 0.914, and the critical value is 0.842 (Shapiro, 1965), the calculated value of the Epps-Palley consent criterion is 0.117, and the critical value is 0.357 (Epps, 1983).

There are no sharply distinguished values in the data massive for the significance level of 0.05. While the calculated value of the Grubbs criteria is 2.166 and the critical values for the smallest and largest deviations have values of 2.294 and 2.441, respectively (Grubbs, 1950; Kobzar, 2006).

While the calculated value of the Grubbs criteria is 2.166, and the critical values for the smallest and largest deviations are 2.294 and 2.441, respectively (Grubbs, 1950; Kobzar, 2006).

As a result of performing a classical linear approximation of a data array using the method of least squares, the equation is obtained:

$$f(x) = 1262.197 + 63.955 \cdot x, \quad (4)$$

the average relative error of which is δ_{me}=37.85 %.

As a result of performing a linear approximation using the modified method of least squares according to condition (2) at p=-14.439, the equation is obtained:

$$f(x) = -909.251 + 113.638 \cdot x, \quad (5)$$

the average relative error of which is δ_{me}=20.06 %.

And at p=2.566 – the equation is obtained:

$$f(x) = -306.722 + 86.911 \cdot x, \quad (6)$$

the average relative error of which is δ_{me}=17.98 %.

Relative errors at each point for the classical and proposed approximation methods are presented in Table 1.

The dependencies of the average absolute and relative approximation errors of the data array Table 1 from the coefficient $p \in [-20; 20]$, respectively, constructed on the basis of solving the objective function optimization problem results (3) by the dichotomy method (Gill, 1981) are shown in Figure 1 and Figure 2.

The margin of solution error was chosen $\varepsilon = 1 \cdot 10^{-3}$. The dotted line shows the average errors values corresponding to the classical approximation method at p=0.

Table 1: Comparison of classical and proposed methods of approximating an array of data with one influencing factor (k=1).

i	x	y	Relative approximation error δ, %		
			equation (4)	equation (5)	equation (6)
1	15	1000	+122.15	-20.47	-0.31
2	20	1400	+81.52	-2.61	+2.25
3	30	2500	+27.23	-0.01	-7.98
4	40	3500	+9.15	+3.89	-9.44
5	50	4700	-5.11	+1.55	-14.07
6	60	5800	-12.08	+1.88	-15.38
7	80	7950	-19.77	+2.92	-16.40
8	100	8500	-9.91	+22.99	-1.36
9	140	15000	-31.89	0.00	-20.93
10	180	8000	+59.68	+144.32	+91.72

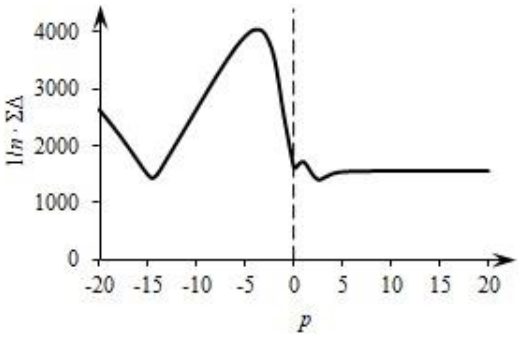

Figure 1: The average absolute approximation error as function of the data array (Table 1) from the coefficient p.

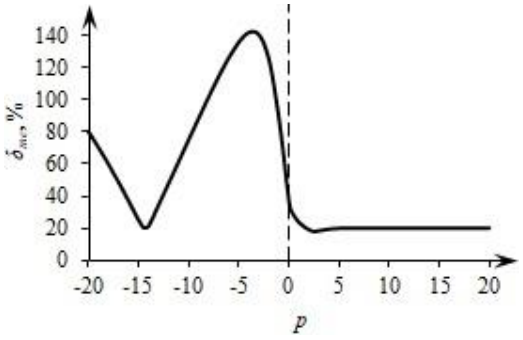

Figure 2: The average absolute approximation error as function of the data array (Table 1) from the coefficient p.

The regression coefficients values a and b of linear equation $f(x)=a+b\cdot x$ obtained as a result of the data array (Table 1) approximation by modified least squares method according to the condition (2) at $p \in [-20; 20]$ are shown in Table 2.

Table 2: The regression coefficients values of a linear equation obtained by the proposed approximation method at $p \in [-20; 20]$.

p	a	b
-20	-3602.750	132.876
-18	-2729.203	126.637
-16	-1744.857	119.606
-14	-665.330	111.895
-12	482.822	103.694
-10	1662.280	95.269
-8	2822.709	86.974
-6	3879.253	79.361
-4	4603.403	73.414
-2	4274.422	68.072
0	1262.197	63.955
2	-245.474	83.352
4	-230.681	82.015
6	-202.722	80.181
8	-200.257	80.017
10	-200.025	80.002
12	-200.002	80.000
14	-200.000	80.000
16	-200.000	80.000
18	-200.000	80.000
20	-200.000	80.000

Let's expand the data massive under consideration, increasing the number of influencing factors to four (k=4), and conduct its linear and power approximations (Table 3).

As a result of performing the classical approximation of the extended data array by the least squares method, linear and power equations respectively are obtained:

$$f(x) = -72848.942 + 573.180 \cdot x_1 + \\ + 32481.523 \cdot x_2 - 86.454 \cdot x_3 + 585.990 \cdot x_4, \quad (7)$$

the average relative error of which is δ_{me}=20.47 %, and

$$f(x) = 0.116 \cdot x_1^{1.543} \cdot x_2^{5.530} \cdot x_3^{0.898} \cdot x_4^{0.572}, \quad (8)$$

the average relative error of which is δ_{me}=11.28 %.

As a result of the approximation using the modified least squares method according to condition (2) at p=1.645, a linear equation is obtained:

$$f(x) = -4831.793 + 144.238 \cdot x_1 + \\ + 3496.753 \cdot x_2 - 29.856 \cdot x_3 - 50.551 \cdot x_4, \quad (9)$$

the average relative error of which is δ_{me}=4.60 %, and at p=7.010 – power-law equation is obtained:

$$f(x) = 0.841 \cdot x_1^{1.082} \cdot x_2^{2.799} \cdot x_3^{0.694} \cdot x_4^{0.742}, \quad (10)$$

the average relative error of which is δ_{me}=8.29 %.

Relative errors at each point for the classical and proposed approximation methods in Table. 3 are presented.

Table 3: Comparison of classical and proposed methods of approximating the data array with four influencing factors (k=4).

i	x_1	x_2	x_3	x_4	y	Relative approximation error δ, %			
						equation (7)	equation (8)	equation (9)	equation (10)
1	15	1.5	2	30	1000	+87.78	-7.25	+0.07	-1.05
2	20	1.4	3	29	1400	-41.20	-0.45	-0.51	+2.78
3	30	1.3	5	27	2500	-21.52	+5.07	+1.08	-1.94
4	40	1.2	8	26	3500	+2.86	+12.17	+2.31	+2.98
5	50	1.1	13	24	4700	-4.69	+7.62	-1.59	+1.00
6	60	1.0	21	22	5800	-12.08	-0.16	-3.79	-0.16
7	80	0.9	35	17	7950	+15.41	-13.45	0.00	-12.83
8	100	0.8	90	11	8500	+7.29	+8.41	+7.60	+4.06
9	140	0.7	222	5	15000	-7.53	-29.30	-27.14	-39.12
10	180	0.6	500	3	8000	+4.29	+28.92	+1.87	+17.02

Table 4 shows data array taken from (Petryanin, 2016), as a result of which a power equation by the least squares method was obtained:

$$f(x) = 6.176 \cdot x^{2.971}, \quad (11)$$

the average relative error of which is δ_{me}=8.31 %.

As a result of using the refined approximation method according to formulas (1), a power equation was obtained (Petryanin, 2016):

$$f^*(x) = 6.531 \cdot x^{2.906}, \quad (12)$$

the average relative error of which is δ_{me}=6.97 %.

As a result of the approximation using the modified least squares method according to condition (2) at p=1.998, a power equation is obtained:

$$f(x) = 6.474 \cdot x^{2.890}, \quad (13)$$

the average relative error of which is δ_{me}=5.36 %.

Relative errors at each point for the classical, refined and proposed approximation methods are presented in Table. 4.

Table 4: Comparison of the classical, refined (Petryanin, 2016) and proposed data array approximating method.

i	x	y	Relative approximation error δ, %		
			equation (11)	equation (12)	equation (13)
1	1.0	6.50	-4.98	+0.48	-0.41
2	1.5	20.38	+1.07	+4.11	+2.55
3	2.0	46.40	+4.34	+5.49	+3.45
4	2.5	88.63	+5.99	+5.63	+3.23
5	3.0	151.10	+6.86	+5.24	+2.56
6	3.5	237.90	+7.29	+4.62	+1.71
7	4.0	535.00	-29.07	-31.43	-33.47
8	4.5	500.30	+7.63	+3.26	0.00
9	5.0	684.50	+7.57	+2.51	-0.89

For those presented in Table 4 data, the hypothesis of a normal random variable distribution y_i for a significance level of 0.05 is confirmed by the fact that the calculated value of the Shapiro-Wilk concord criterion is 0.860 and the critical value is 0.829 (Shapiro, 1965), the calculated value of the Epps-Palley consent criterion is 0.280, and the critical value is 0.350 (Epps, 1983).

There are no sharply distinguished values in the data massive (Table 4) for the significance level of 0.05. While the calculated value of the Grubbs criteria is 1.692 and the critical values for the smallest and largest deviations have values of 2.237 and 2.392, respectively (Grubbs, 1950; Kobzar, 2006).

4 DISCUSSION

The need to test the hypothesis of the normal distribution of a random variable y_i in an experimental data array (x_i; y_i), $i=1\ldots n$ is determined by the central limit theorem existence in probability theory.

The significance level chooses is based on the amount of experimental data n. And the absence of sharply distinguished values in the array (x_i; y_i) is a mathematical confirmation of the experimental conditions stationary, which guarantees with a given probability a significant influence absence on the observed unaccounted factors phenomenon (Kobzar, 2006).

The modified least squares method apply in accordance with condition (2) makes it possible to select such a position of the regression line in the k-dimensional factor space and the corresponding values of the regression coefficients that provide increased accuracy of experimental data approximation in comparison with the classical approximation method.

In this case, the position of the regression line in the k-dimensional factor space according to condition (2) is determined by only one single coefficient p, which value is found as a result of solving the unconditional nonlinear objective function optimization problem of (3).

It should be noted that this fact allows you to solve the optimization problem with any number of influencing factors, i. e. at $k \in [1; n]$, and a given solution error ε=const for the same number of steps m.

Analysis of linear approximation data array (Table 1) results with one influencing factor shows that the mean relative error of the equation (6) obtained by the modified least squares method according to condition (2) is 19.87 % (2.10 times) less than the average relative error of the equation (4) obtained using the classical approximation method. At the same time, the maximum relative error of the equation (6) is 30.44 % (1.33 times) less than the maximum relative error of the equation (4).

Analysis of linear approximation data array (Table 3) results with four influencing factors shows that the average relative error of the linear equation (9) and the power-law equation (10) obtained using the modified least squares method according to condition (2), respectively, is 15.87 % (4.45 times) and 2.99 % (1.36 times) less than the average relative error of the linear equation (7) and power-law equation (8) obtained by the classical approximation method.

At the same time, the maximum relative error of the linear equation (9) is 60.64 % (3.23 times) being smaller than that for the linear equation (7), and the maximum relative error of the power-law equation (10) is 9.82 % (1.34 times) being greater than that for the power-law equation (8).

It may be noted that from the viewpoint of the classical method, the data array (Table 3) is most accurately approximated by a power-law function, since the average relative error of the power-law equation (8) δ_{me}=11.28 % is less than the average relative error of the linear equation (7) δ_{me}=20.47 %.

And from the proposed method perspective, the greatest accuracy of the approximation is provided just by the linear function, since the average relative error of the linear equation (9) δ_{me}=4.66 % is being smaller than that for the power-law equation (10) δ_{me}=8.29 %.

Thus, for the proposed method, the preliminary selection of approximating function (regression

equations) type, based on geometric considerations, is not as important as for the classical method.

Analysis of approximation data array (Table 4) results shows that the average relative error of the power-law equation (13) obtained by the modified least squares method according to condition (2) is 1.61 % (1.30 times) being smaller than that for the power equation (12) obtained by the refined approximation method (Petryanin, 2016), and 2.95 % (1.55 times) less than the average relative error of the power-law equation (11) obtained by means of the classical approximation method.

At once, the maximum relative error of the power-law equations (12) and (13), respectively, is 2.36 % (1.08 times) and 4.40 % (1.15 times) being greater than that for the power-law equation (11). Thus, the proposed method provides increased accuracy of approximation of experimental data in comparison with the refined approximation method (Petryanin, 2016).

Most solutions to unconditional nonlinear optimization problem of the objective function (3) belong to the interval $p \in [-20; 20]$ and can be found with an error $\varepsilon=1\cdot10^{-3}$ that corresponds to an acceptable accuracy for most practical purposes.

Three objective function (3) minimums on the interval $p \in [-20; 20]$, found as a result of data array (Table 1) mathematical processing according to condition (2): local minimum at p=-14.439 according to equation (5), local minimum at p=0 in relation to equation (4), global minimum at p=2.566 in relation to equation (6) are demonstrated on the Figure 1.

The dependencies of the average relative approximation errors of the data array (Table 1) from the coefficient p showed on the Figure 2 is almost similar to the dependencies on the Figure 1.

At the same time, the increase and decrease in the objective function (3) in Figure 1 corresponds almost synchronously to the ascending and decreasing values of the regression coefficient in Table 2 and the regression coefficient values move from descending to increasing at p=0, which corresponds to the classical approximation method.

Thus, the fulfillment of condition (2) provides the best position of the regression line in the k-dimensional factor space at any value, which allows for high approximation accuracy.

Let's estimate the required number of calculation cycles the regression coefficients values of the equation with k influencing factors using the modified least squares method according to condition (2).

In order to find the optimization problem solution, it is necessary that the objective function (3) on the interval $p \in [p_1; p_2]$ is unimodal, and therefore we will choose the width of the interval equal to one. In this regard, to completely cover the most likely range $[-20; 20]$ in which the desired solution is located 40 steps will be required.

Let's choose the error of the solution $\varepsilon=1\cdot10^{-3}$. Then, by means of the dichotomy method, it is necessary to do m=11 steps for two cycles in each, since at each step the objective function (3) values are calculated at two points (Gill, 1981). As a result, the required number of calculation cycles will be 880, which is very acceptable from the point of view of using modern computing power.

5 CONCLUSIONS

The proposed method can significantly improve the accuracy of the experimental data approximation compared to most well known methods and can be applied to any kind of approximating function (regression equations).

The maximum achievable approximation accuracy of the proposed method is commensurate with the accuracy due to the use of artificial neural networks, and the number of computational operations performed can significantly save computing power. Thus, the proposed method of approximating experimental data allows in real time to construct stochastic mathematical models for controlling fast-flowing processes, which form the basis of various, including responsible, tracking systems work.

ACKNOWLEDGEMENTS

The work was carried out with the financial support of the Russian Federation President grant MK-1603.2022.4 "Development of energy-efficient equipment for the implementation heat and mass exchange processes using multiple droplets boiling and highly porous materials based on experimental and theoretical research".

REFERENCES

Epps, T.W., Pulley, L.B., 1983. *The article*, Biometrika, 70.

Gill, P.E., Murray, W., Wright, M.H., 1981. *The book*, Practical optimization. London, New York, Toronto, Sydney, San Francisco.

Grubbs, F.E., 1950. *The article*, Annals of Mathematical Statistics, 21.
Harky, A., Chaplin, G., Chan, J.S.K., Eriksen, P., etc., 2020. *The article*, Heart, Lung and Circulation, 29.
Kobzar, A.I., 2006. *The book*, FIZMATLIT. Moscow.
Kuz'min, A.N., 2017. *The article*, Atomic Energy, 122.
Meade, A.J., 1993. The conference paper. *31st Aerospace Sciences Meeting & Exhibit*.
Oishi, J., Asakura, K., Watanabe, T., 2006. The conference paper. *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications & Conference on Real-Time Computing Systems and Applications*.
Petryanin, D.L., 2016. *The article*, Reliability and quality of complex systems, 3.
Raisch, J., Klein, E., Meder, C., Itigin, A., 1999. The conference paper. *Lecture Notes in Computer Science*.
Shapiro, S.S., Wilk, M.B., 1965. *The article*, Biometrika, 52.
Walter, W., Striberny, B., Gaquerel, G., Baldwin, I.T., etc, 2014. *The article*, BMC Bioinformatics, 15.