Algorithmic Structure and Functional of «CutGlueApproximation» Software Complex

R. A. Neydorf[1], V. V. Polyakh[2], O. T. Yarakhmedov[3], I. V. Chernogorov[4]
Don State Technical University
Rostov-on-Don, Russia
ran_pro@mail.ru[1], silvervpolyah@gmail.com[2], orhashka@gmail.com[3], hintaivr@gmail.com[4]

***Abstract*— For the construction of mathematical models of essentially nonlinear technical objects, it was developed an alternative to the existing "Cut-Glue" approximation method, using the splitting of an array of experimental data into fragments having well described by polynomial functions data structure. The polynomials describing these fragments are multiply transformed, obtaining fragmentarily isolated impulse-like functions that describe the data within the boundaries of the fragments, and practically zero in the rest of the domain. These properties make it possible to obtain a mathematical model of the entire dependence by the additive union of fragments. In the presented method three problems are distinguished: fragmentation, approximation of fragments and their multiplicative-additive processing. The method of the task solving of data fragmentation is based on logical-combinatorial methods that allow us to investigate the domain of admissible solutions. In the problem of mathematical polynomial description of data fragments situated their structurally and parametrically optimal variants of approximating polynomials. In this case we use a hybrid of classical regression analysis and modification of the evolutionary-genetic algorithm. The solution of the resultant task of the method is carried out by two operations: 1) cutting along the boundaries with the help of special multiplicatively applied functions from the polynomials of their interval-isolated fragments, 2) the additive union of these fragments into a unified mathematical model of the investigated dependence. In this case, the model is minimized parametrically. The described method implements a fundamentally new approach, it is designed to solve problems of experimental description of nonlinear dependencies. "CutGlueApproximation" software package was developed for its implementation. The article describes the essence and possibilities of the functional implemented in the complex.**

Keywords— experimental data; modeling; nonlinearity; approximation; mathematical model; optimization; heuristic algorithms; combinatorics; software

I. INTRODUCTION

The construction of mathematical models (MM) of technical objects is often associated with the mathematical processing of experimental data (ED). For the experimental construction of such objects MM, many methods for approximating the data have been developed as like: polynomial regression [1-4], orthogonal expansions [5, 6], piecewise functions [7, 8], splines [9, 10], radial basis functions [11, 12] etc. But because of non-linearities, the description of such dependencies by known methods is difficult, and is associated with inaccuracies. In this connection, was created an alternative method of "Cut-Glue" approximation (CGA) [13-15]. It uses a decomposition of the modeled dependence into sections with a data structure that are well approximated by polynomial functions. Multiplicative "cutting" of fragments along the boundaries of sections is applied to them. And the procedure ends with the additive "gluing" of fragments into a single function that simulates the numerical structure of the original data array. The CGA method distinguishes three tasks: fragmentation, approximation and multiplicative-additive processing of the results of the first two problems [13-15]. The first task is the ED partitioning into fragments (EDF). The methods for solving it are considered in subsection *III.A*. The second task of the CGA is to describe the EDF with obtaining for the fragment the optimal version of the locally approximating function (LAF). It is investigated in subsection *III.B*. The resulting stage of the CGA method consists of two operations: transforming the LAF into a so-called. interval-isolated functions (IIF), which approximate the data within the boundaries of the EDF, and their additive composition in a single analytic function (GAF), which is an object MM. The essence and algorithms for performing these operations are described in subsection *III.C*. The CGA method implements fundamentally new approaches to approximation and is designed to solve problems of experimental description of nonlinear objects.

II. PROBLEM FORMULATION

In connection with the considerable diversity of the CGA method steps, both ideologically and algorithmically, it became necessary to develop promising approaches, both to theoretical and practical methods for their solution, to construct and test the main algorithms. Thus, in the article the following tasks should be solved:

1. to develop an algorithm for the variable decomposition of ED on EDF with respect to two-dimensional ED;
2. solve the problem of optimal polynomial EDF approximation;
3. algorithmically provide and study the stages of the multiplicative-additive transformation of the EDF with the production of MM-CAP.

Research supported by the Russian Fundamental Research Fund, project No. 18-08-01178A

978-1-5386-5824-6/18/$31.00 ©2018 IEEE

III. Algorithms of Implementation and Optimization of Stages of Non-Linear Objects Mathematical Modeling by «CGA» Method

A. *Logic-combinatorial algorithm for ED fragmentation in the CGA task*

The CGA method is designed to solve problems of approximating ED, represented by hyper-parallelepipeds in the *n*-dimensional factor space. To partition the ED array into a EDF, the factor space is covered by a grid constructed according to the coordinate-wise variation steps (CVS-grid) of the input variables chosen in the experiment planning step.

The main restriction on the fragmentation algorithm (FA) imposed by the CGA method [14–15] is the following 2 fundamental conditions:

- the union of all fragments is the whole investigated data space:

$$S_1 \cup S_2 \cup \ldots \cup S_n = S, \quad (1)$$

where: S_i – *i*-th fragment of a partition; n – number of fragments; S – the whole set of ED (S – «snippet»);

- the intersection of neighboring fragments is a subset of the elements that form their common boundary:

$$S_i \cap S_j = B_{ij} = B_{ji}, \quad (2)$$

where: S_i and S_j – contiguous faces B_{ij} and B_{ji} fragments, (B – «border»)

Condition (1) is obvious, and condition (2) is introduced in connection with the fact that in the future, when implementing the additive union of the final IIF, it is necessary to provide the "seamless" connection between them [16]. Since there are no prerequisites to require the EDF to completely coincide on the mutual boundaries, which is shown in Fig. 1, a fragmented set of ED can not be described by regular mathematical structures. We can conditionally assume that the EDF aggregate is "boundary-crossing irregular-cellular matrices". In this case, the result will be the set of EDFs related to the initial set of ED by the relations:

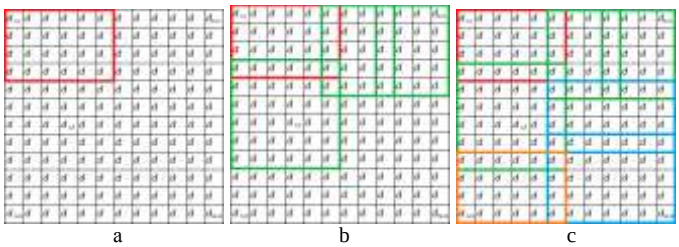
$$M = \prod_{j=1}^{n} m_j = \sum_{k=1}^{N} n_k \quad \text{(a)}, \qquad n_k = \prod_{j=1}^{n} m_{k_j} \quad \text{(b)}, \quad (3)$$

where: M – amount of data in ED; N – the number of fragments – hyper-parallelepipeds, each of which have common values for faces that are not boundary; m_j, m_{kj} – the number of factor values along the *j*-axis in ED and in the *k*-м EFD; n_k – the amount of data in the partition structure.

Fragmentation leads to the separation of the CVS-grid into a set of hyper-parallelepipeds with common adjacent data on non-boundary edges, and the implementation of the described algorithm is reduced to the sequential execution of the following operations:

- the range of permissible sizes of fragments in the CVS is defined: h_1 and h_2. The length of fragments h is generated in the interval $h_1 \le h \le h_2$ by the random distribution law;
- CVS-grid is filled with fragments in a random sequence with a logical check of restrictions on the size of EDF, until the entire space is filled.

Fig. 1 shows the process of filling the EDF in a simple test problem with a dimension of 12x12 step by step. The letter *d* denotes data (*d* – "data").



a b c

Fig. 1. Stages of the program: a – selection of the first fragment, b – intermediate stage, c – final result

As the first variant of the solution of problems of bounded dimension, the logical-code AF (LCFA) was developed, which showed rather high efficiency. A software tool (ST) «Experimental Data Fragmentation Method» (EDFM) has been designed and developed. Currently, the study and further modification of LCFA is being conducted to improve its effectiveness, as well as to assess the applicability limits within the framework of the task.

B. *Structural-parametric optimization of polynomial EDF approximation*

The methodological basis for the second stage of the CGA is the author's algorithms based on polynomial regression [8-11]. Power polynomials having a regular structure for any order are adopted as a universal form of LAF for EDF. This makes it possible to evaluate the complexity of the LAF critically by the presence and number of terms of a polynomial of arbitrary order

$$\begin{aligned} Y(x) &= b_0 + b_1 x_1 + \ldots + b_n x_n + b_{11} x_1^2 + b_{12} x_1 x_2 + \ldots \\ &+ b_{1n} x_1 x_n + b_{22} x_2^2 + \ldots + b_{2n} x_2 x_n + b_{(n-1)n} x_{n-1} x_n + \\ &+ b_{nn} x_n^2 + b_{111} x_1^3 + \ldots + b_{11n} x_1^2 x_n + \ldots b_{122} x_1 x_2^2 + \ldots \\ &+ b_{nnn} x_n^3 + b_{1111} x_1^4 + \ldots + b_{111n} x_1^3 x_n + b_{1122} x_1^2 x_2^2 + \ldots \end{aligned} \quad (4)$$

The regularity of the polynomial (4) simplifies the structural variation and the possibility of structural-parametric optimization (SPO) of each fragment. For an SPO approximating power polynomial (APS) of dimension n it is developed an algorithm based on its representation by a pseudo-linear polynomial of extended dimension n:

$$q(\tilde{x}) = \sum_{i=0}^{n} \tilde{b}_i \times \tilde{x}_i = \sum_{i=0}^{n} b_i \times x_i + \sum_{i=n+1}^{n} \hat{b}_i \times \hat{x}_i \,, \quad (5)$$

where: $\tilde{b}_i$ – coefficients of a pseudo-linear polynomial of $\tilde{n}$-th dimension and $\tilde{x}_i$ – generalized arguments of the investigated dependence, including both the original x_i, and the pseudo-arguments $\hat{x}_i$, These elements replace the nonlinearities of the power polynomial.

The second member of polynomial (5) contains, in fact, dependent arguments. Therefore, its coefficients, calculated from the classical matrix formula:

$$\tilde{b} = (X^T X)^{-1} X^T y, \quad (6)$$

are only suboptimal. Here y–N_k – is the vector of k-th EDF; X – matrix of inputs $\tilde{x}_{ji}$; $j \in [1, N_k]$ and i – are the numbers of the experiments and pseudo-variables; $\tilde{x}_{i0} = 1$. The interdependence of the part $\tilde{x}_i$ determines the sub-optimality of the calculated coefficients $\tilde{b}_i$, relative to the LSM criterion used in the CRA, but it itself is only heuristic.

The search for the SPO version of the APS option raises the quality of the EDF approximation. Often, complete polynomial does not give the best accuracy, because of the properties of some nonlinear terms that do not agree with the singularities of the curvature of the approximated dependences given by the EDF. The algorithm of SPO APS varies its structure, estimating its complexity and accuracy of EDF approximation. The problem of discrete SPO is solved on a set of such two-dimensional criteria [14]. If the dimension of the object is large and/or the amount of EDF is large too, combinatorial methods become resource-inefficient. Therefore, a modified evolutionary-genetic algorithm (MEGA) has been developed for SPO APS. Example of solved solution of the SPO APS problem for a particular EDF shown below by the matrices and Table 1. The upper row and the left column of the matrices (7) are the values of x_1 and x_2. The first matrix is the values of the EDF, and the second is its SPO LAF.

$$\begin{pmatrix} & 1500 & 2000 & 2500 & 3000 \\ 0 & 2438 & 4470 & 6632 & 8989 \\ 10 & 2238 & 4299 & 6550 & 8916 \\ 20 & 1732 & 3693 & 6060 & 8456 \\ 30 & 922,78 & 2911 & 5222 & 7884 \end{pmatrix} \Rightarrow \begin{pmatrix} & 1500 & 2000 & 2500 & 3000 \\ 0 & 2441 & 4449 & 6672 & 8978 \\ 10 & 2236 & 4229 & 6360 & 8928 \\ 20 & 1731 & 3687 & 6064 & 8470 \\ 30 & 923 & 2902 & 5213 & 7901 \end{pmatrix} \quad (7)$$

The top row of Table 1 contains the pseudo-variable codes indicating the numbers of the multiplied x_i. Below are the structural versions of the APS in order of increasing of the error of approximation. The columns of the table represent the members of the complete 4th order polynomial with the corresponding code specified in the top line. Their absence is indicated by the symbol "-". The last two columns give estimates of the absolute (**Δ**) and relative (**δ**) errors on the set of approximated data, showing the approximate accuracy of the variants of the polynomial. The results of Table 1 convincingly demonstrate the effectiveness of the structural variation of APS EDF in its MM optimizing.

TABLE I RESULT OF SPO APS X_2

0	1	2	11	12	22	111	112	122	222
0	1	2	11	12	--	111	--	122	--
0	1	2	11	12	--	--	112	122	--
0	1	2	--	--	--	--	112	--	222

Table 1 continue:

1111	1112	1122	1222	2222	Δ	δ
--	1112	--	1222	--	21,85576	0,00523
1111	--	--	1222	--	22,94357	0,00549
1111	--	1122	1222	2222	31,73257	0,00608

C. *The essence and parametric optimization of CGA method "Cut-Glue" phase*

The last CGA stage is realized by two operations: "Cut the fragments" (CF) and "Glue the fragments" (GF) [13, 14]. These operations with sufficient generality can be described by the following expressions:

$$\text{CF:} \quad \forall i = \overline{1, N} \rightarrow f_i(\vec{x}_i) = \phi_i(\vec{x}_i) \cdot \prod_{j=1}^{n} E_{i_j}(x_{i_j}, x_{i_{jL}}, x_{i_{jR}}, \varepsilon_{i_j}), \quad (8)$$

$$\text{GF:} \quad F(\vec{x}) = \sum_{i=1}^{N} f_i(\vec{x}_i), \quad (9)$$

where: $f_i(\vec{x}_i)$ – i-th IIF; N – fragments and their IIFs counts; $\phi_i(\vec{x}_i)$ – i-th LAF; n – factor dimension of object; $E_{i_j}(*)$ – j-th one-dimensional function for i-th LAF; x_{i_j} – i-th argument of j-th $E_{i_j}(*)$; $x_{i_{jL}}$, $x_{i_{jR}}$ – borders values of approximation range of i-th EDF by j-th variable; $\varepsilon_{i_{jL}}$ and $\varepsilon_{i_{jR}}$ – parameters of fronts steepness of i-th IIF by j-th variable.

In (8) CF operation implemented using special multiplicative allocation function (MAF or epsilon-function) $E_{i_j}(*)$. The basis of this function structure is one-dimensional ε-function or 1-MAF:

$$E(x, x_L, x_R, \varepsilon_L, \varepsilon_R) = \frac{0.25 \cdot \sigma_L(x, x_L, \varepsilon_L) \cdot \sigma_R(x, x_R, \varepsilon_R)}{\delta(x, x_L, x_R, \varepsilon_L, \varepsilon_R)}, \quad (10)$$

where: $\delta(x, x_L, x_R, \varepsilon) = \sqrt{\left[(x - x_L)^2 + \varepsilon_L^2\right] \cdot \left[(x_R - x)^2 + \varepsilon_R^2\right]}$,

$\sigma_R(x, x_R, \varepsilon) = x_R - x + \sqrt{(x_R - x)^2 + \varepsilon_R^2}$, $\sigma_L(x, x_L, \varepsilon) = x - x_L + \sqrt{(x - x_L)^2 + \varepsilon_L^2}$.

Property of this function at fragment borders are close to signature ones. It forms the IIF fronts $f_i(\vec{x}_i)$, separating the regions of neighboring EDFs by coordinates of factors (for multidimensionality $x \sim x_i$). 1-MAF "cuts out" from LAF $\phi_i(\vec{x}_i)$ its section at fragment interval by x_i and approximates to zero all values of IIF in (9) beyond the boundaries of "cutted" interval. This MAF property allows the next GF operation to additively integrate all IIFs $f_i(\vec{x}_i)$ into a single GAF function - MM of object.

However, MAF structure (10) not only ensures the effective integration of fragment models into a single model [13–15]. Its parameters ε_L, ε_R (in the multidimensional case $\varepsilon_{i_{jL}}$ and $\varepsilon_{i_{jR}}$) make it possible to reduce the overall error of ED approximation of the resulting GAF. The possibility of its minimization arises, due to alternative trends in the dependence of 1-MAF boundary values from ε_{i_j}. The effect investigations of MAF properties on GF results have shown that, in order to influence the accuracy of ED array approximation, it may be necessary to change $\varepsilon_{i_{jL}}$ and $\varepsilon_{i_{jR}}$ values over a wide range [13, 15].

To solve the tasks of error parametric minimization in ED approximation of the final GAF of any dimension, the heuristic method of swarming particles (MSP), modified under the problem, is used [16–20]. In connection with the multi-

factority of problem an algorithm of stepwise conditional optimization (SCO) of general methodological error CF-GF operations has been developed. The SCO idea is to find the components of ε vector, which close to the optimal ones as starting ones and to simplify as much as possible the problem of GAF global optimization. According to this, at first step the conditional parameters optima $\varepsilon_{i_{jL}}$ and $\varepsilon_{i-1_{jR}}$ are searched for pairs of neighboring IIFs with a common boundary, using fixed parameters values of other variables. Further, the perimeter of fragment is bypassed, using the replacement of conditionally assigned values ε_{i_j} by sub-optimized, if necessary, in the cycle.

To demonstrate the effectiveness of CF and GF operations, an experiment was conducted using 4 EDFs. The array of 2-dimensional ED is divided into 4 adjacent fragments and each fragment is approximated by LAF. The maximum relative error was ~7.84%. Table 2 shows the sections of relative error matrix (%) along the boundaries of IIFs gluing and boundaries values obtained during the stepwise optimization of fragments. In each EDF, that is close to the boundaries, the obtained sub-extremal vector ε values are indicated for optimized boundaries.

TABLE II CF AND GF OPERATIONS RESULT AND RESEVED EXTREMAL VECTOR E VALUES

...	ε1xR = 1,348	0,256	0,190	3,798	ε2xL = 1,440	...	...	...
ε1yL = 1,846	...	0,132	0,796	3,362	...	ε2yL = 1,720	...	...
2,003	1,347	0,591	1,012	1,581	2,74	0,016	4,41	4,496	2,721
1,051	1,105	0,511	1,017	1,319	0,919	1,419	4,447	4,366	0,599
0,717	1,507	1,254	0,135	0,470	1,414	1,709	2,629	2,831	1,365
ε3yR = 1,595	...		1,174	1,087	0,842	...	ε4yR = 1,352		...
...	ε3xR = 0,133		1,849	1,796	1,854	ε4xL = 0,364	...	...	...
...	...	...	2,082	4,291	6,760	...	...	...	...

The cyclic bypassing of EDFs boundaries are ended after repeated value of maximum relative approximation error ~4,496%. It proves the effectiveness of proposed algorithm, which simplified the problem of multivariate optimization to a cyclic two-dimensional one, because original maximum error of the first stages was reduced by 1.74 times.

IV. CONCLUSION

The developed software complex to solving the tasks of CGA stages implement created method. All stages are effectively algorithmized and implemented programmatically. A simple, but effective LCFA automates the most resourceful CGA stage. Structural-parametric optimization of the approximation stage of fragments structure minimizes the methodical error of the first two stages. At the same time, the proposed algorithm of initial pairwise 2-dimensional optimization of final GAF using final sub-optimization of the greatest error fragment in bypass process proved as an effective. The initial error of approximation can be substantially reduced.

REFERENCES

[1] Rawlings J.O., Pantula S.O., Dickey D.A. Applied Regression Analysis: A Research Tool, Second Edition, 1998, 671 p.

[2] Bates D.M., Watts D.G., Wiley J. & Sons Nonlinear regression analysis and its applications. New York, 1988, 365 p.

[3] Drapper N.R., Smith H., Wiley J. & Sons Applied regression analysis, vol. 1. New York, 1981, 366 p.

[4] Drapper N.R., Smith H., Wiley J. & Sons Applied regression analysis, vol. 2. New York, 1981, 351 p.

[5] Totik V. Orthogonal Polynomials. Surveys in Approximation Theory, vol. 1, 2005, PP. 70–125.

[6] Khrushchev S. Orthogonal Polynomials and Continued Fractions From Euler's Point of View. Atilim University, Turkey. Cambridge University Press, 2008, 495 p.

[7] Recent Developments In Generalized Analytic Functions And Their Applications. Ed. G.Giorgadze. Tbilisi, 2011, 158 p.

[8] Loran P. Aproximation and optimization: Trans. from English. "World" Publishing House, Moscow, Russia, 1975, 496 p.

[9] Ahlberg J.H., Nilson E.N., Walsh J.L. The Theory of Splines and Their Applications, 1st Edition. Mathematics in Science and Engineering: A Series of Monographs and Textbooks, vol. 38, 296 p.

[10] Boor D.C. A practical guide to splines. Springer, 1978, 348 p.

[11] Sarra S.A., Kansa E.J., Davis Multiquadric Radial Basis Function Approximation Methods for the Numerical Solution of Partial Differential Equations. Marshall University and University of California, 2009, 220 p.

[12] Buhmann, M.D. Radial Basis Functions: Theory and Implementations / Cambridge University Press, 2003, 272 p.

[13] Neydorf R., Bivariate "Cut-Glue" Approximation of Strongly Nonlinear Mathematical Models Based on Experimental Data. SAE Int. J. Aerosp. 2015. No. 8(1). Pp. 47-54. (In Russian).

[14] Neydorf R., Chernogorov I., Polyakh V., Yarakhmedov O., Goncharova J., Neydorf A. Formal Characterization and Optimization of Algorithm for the Modelling of Strongly Nonlinear Dependencies Using the Method "Cut-Glue" Approximation of Experimental Data. SAE Technical Paper. 2016. 2016-01-2033 issue. Pp. 1-12. (In Russian).

[15] Neydorf R., A. Neydorf. Technology of Cut-Glue Approximation Method for Modeling Strongly Nonlinear Multivariable Objects. Theoretical Bases and Prospects of Practical Application. SAE Technical Paper 2016-01-2035, 2016 (In Russian).

[16] Neydorf R., Chernogorov I., Polyakh V., Yarakhmedov O., Goncharova Y., Vucinic D. Comparative Analysis of Heuristic Algorithms for Solving Multiextremal Problems. International Journal on Advances in Systems and Measurements. vol. 10, No. 1&2. 2017. Pp. 86-99.

[17] Cieselski Z., Musielak J. Spline bases in function spaces. Approximation theory, Reidel, 1975, Pp. 49–54.

[18] Eberhart R.C., Kennedy J.A. New optimizer, using particle swarm theory. Proceedings of the 6th International Symposium on Micromachine and Human Science. Nagoya, Jap. 1995. Pp. 39-43.

[19] Kennedy J., Eberhart R. Particle Swarm Optimization. Proceedings of IEEE International Conference on Neural Networks. Piscataway, New Jersey. 1995. Pp. 1942-1948.

[20] Shi Y., Eberhart R.C. A modified particle swarm optimizer. Proceedings of the IEEE Congress on Evolutionary Computation. Piscataway, New Jersey. 1998. Pp. 69-73.