

Polynomial Optimization in Mathematical Models Defining Experimental Data Dependencies

Rudolf Neydorf*, Victor Poliakh**

* Don State Technical University,
Russia, Rostov-on-Don,
e-mail: ran_pro@mail.ru

** Don State Technical University,
Russia, Rostov-on-Don,
e-mail : silvervpoliah@gmail.com

Dean Vucinic***

***Vesalius College (VeCo)
Vrije Universiteit Brussel (VUB)
e-mail : dean.vucinic@vub.ac.be

Faculty of Electrical Engineering, Computer
Science and Information Technology (FERIT)
Josip Juraj Strossmayer University of Osijek
e-mail : dean.vucinic@ferit.hr

Abstract - In this paper, the algorithm to mathematically model fragments, which are extracted from non-linear experimental dependencies, is developed, and represents the key steps within the Cut-Glue approximation method. The hybrid search algorithm is based on the classical regression analysis, which takes into account the polynomial structures implemented through the combinatorial laws, and low dimensionality. In the case when the direct search is resource-impossible, the modified evolutionary-genetic algorithm (EGA) is applied. The advantage of the developed algorithm is the guarantee that the optimal polynomial structure exists and can be found. The proposed approach carries out the structural-parametric optimization for each of the studied fragments to define its experimental data dependence. The validation of the polynomial structural-optimization is performed by applying a specially developed software tool, which, in theory, makes possible to approximate fragments of any dimension.

Keywords – optimization; approximation; regression analysis; mathematical model; experimental data; combinatorics.

I. INTRODUCTION

The technical processes occurring in real life are essentially nonlinear and frequently governed by unknown laws [1][2]. Therefore, their direct mathematical modeling is a complex or impossible task, which can be overcome by constructing the Mathematical Models (MM) based on the experimental data, whose input-output dependences are found to be nonlinear [3]. The approximation of such dependencies is a difficult problem to solve and can lead to significant errors, thus special methods are required for effectively solving these problems [4]-[7].

In particular, the presented method is generalized and developed in [8][9]. It is based on the multiplicative "excision" of the modeled dependence sections that are sufficiently and accurately approximated with the analytic functions applying their additive "gluing" characteristics, which allows to have a single analytic function as final result. This data processing approach is called the "Cut-Glue" Approximation (CGA) [5]-[8], and this associative term, in some other related works, is called "multiplicative-additive approximation", as being more mathematically oriented.

The implementation of the CGA method is based on the successive execution of the following operations:

- Cutting out the Experimental Data (ED) array into fragments that can be successfully approximated by appropriate analytic functions (the fragments must have common boundaries, so that their union gives the original array);
- Approximation of fragments by applying the most suitable analytical functions for their data profile to minimize the approximation error and the complexity of the approximation functions;
- Formation of new interval-isolated functions, as different from the functions approximating fragments, which are spatially separated by their coordinate boundaries (special nonlinear multiplicative "cutting" functions [4]-[9] are suggested);
- Assembling the cut-out interval-isolated fragments into a single analytic function (i.e., implementation of the additive operation "glue"), defining the mathematical model of the investigated dependence.

Each CGA method operation determines the quality of the final result. However, the error in the ED mathematical description is, to the greatest extent, determined by the quality of the constructed analytic functions, which isolate the initial experimental data [8][9] that approximates fragments. Therefore, to minimize the approximation error, the received MM can be considered as the key step in the CGA method. This paper is devoted to the development of the methodology and its approximation algorithm, with the focus on the MM regression approximation of data fragments applying the pseudo-linear polynomial combinatorial model. The computed regression polynomial is the approximation function for the applied gene-chromosome model. In addition, the performed computational tests are using the structural-parametric optimization algorithms applying the experimental data regression. The paper goal is to define the analytical function that approximates an arbitrary ED fragment that does not contain any discontinuities in the defined function and its respective derivatives. Such defined dependency function and its structure can be varied and thus, parametrically optimized.

II. REPRESENTING MM REGRESSION APPROXIMATION OF DATA FRAGMENTS BY A PSEUDOLINEAR POLYNOMIAL

Research in the field of the problem, given in [4]-[9], has shown that the method of execution stage approximation fragmented ED treatment should be functionally flexible, combining the flexibility and structure, and analytical and structural diversity. All these requirements are best-served by a well-proven machine classical regression analysis (CRA), focused on the construction of polynomial models [10]-[13]. Polynomial degree has universal design and is suitable for both structural variations within the polynomial members and to effectively optimize the parametric regression coefficients by the method of least squares. As a result, in the polynomial CRA polymer, there is the possibility of effective structural and efficient parametric optimization, which approximates the models for the considered MM fragment.

Structure of $Y(x)$ polynomial of arbitrary m -th degree at the n -th dimension may be represented as follows:

$$\begin{aligned}
 Y(x) = & b_0 + b_1x_1 + \dots + b_nx_n + b_{11}x_{12} + b_{12}x_1x_2 \\
 & + b_{13}x_1x_3 + \dots + b_{1n}x_1x_n + b_{22}x_2x_2 \\
 & + b_{23}x_2x_3 + \dots + b_{2n}x_2x_n + \dots + b_{(n-1)n}x_{n-1}x_n \\
 & + b_{mn}x_{n2} + b_{111}x_{13} + b_{112}x_{12}x_2 + \dots \\
 & + b_{11n}x_{12}x_n + \dots + b_{122}x_1x_2x_2 + \dots + b_{1mn}x_1x_nx_2 \\
 & + b_{222}x_{23} + b_{223}x_{22}x_3 + \dots \\
 & + b_{22n}x_{22}x_n + \dots \\
 & + b_{nnn}x_{n3} + b_{1111}x_{14} + b_{1112}x_{13}x_2 + \dots + b_{111n}x_{13}x_n + \\
 & b_{1122}x_{12}x_{22} + b_{1123}x_{12}x_2x_3 + \dots \quad (1)
 \end{aligned}$$

where $b_{ijk\dots}$ are the coefficients of the n -th and m -th degree polynomials whose composite indices indicate the variables that are multiplied when forming the polynomial term (for example, b_{1123} is the multiplier for the product $x_1 x_2 x_3$); x_i are the indexed independent (input) variables of the experimental dependence being investigated.

One of the simplest and most well-known methods of simplifying the algorithm for finding regression coefficients is the representation of its nonlinear terms as additional arguments of the pseudo-linear factor space of a new vector of variables \tilde{x} of extended dimension, where

$$\begin{aligned}
 \forall i = \overline{1, n} \rightarrow \tilde{x}_i = x_i; \tilde{x}_{n+1} = x_1 \cdot x_1; \tilde{x}_{n+2} \\
 = x_1 \cdot x_2; \dots \tilde{x}_{n+n} = x_1 \cdot x_n; \tilde{x}_{2n+1} \\
 = x_2 \cdot x_2; \dots \\
 \tilde{x}_{3n-1} = x_2 \cdot x_n; \tilde{x}_{3n} = x_3 \cdot x_3; \tilde{x}_{3n+1} = \\
 x_3 \cdot x_4; \dots \tilde{x}_{4n-3} = x_3 \cdot x_n; \dots, \quad (2)
 \end{aligned}$$

\tilde{x}_i are generalized arguments of the dependency, including the original arguments x_i , and also the pseudo arguments \tilde{x}_i , which are all possible products of the original arguments.

In this case, the nonlinear polynomial (1) takes the form

$$Y(\tilde{x}) = \sum_{i=0}^{\tilde{n}} \tilde{b}_i \cdot \tilde{x}_i \quad (3)$$

where \tilde{b}_i are the coefficients of the pseudo-polynomial of the \tilde{n} -th dimension, b_i and \tilde{b}_i - coefficients of any variant of the linear polynomial (3) describing the modeled dependence and are calculated from the well-known matrix formula:

$$\tilde{b} = (X^T X)^{-1} X^T Y, \quad (4)$$

where Y is the vector of values of the dependent variable, and X is the matrix of inputs to be examined, which consists of lines \tilde{x}_i , the line numbers correspond to the number of the experiment, the column numbers correspond to the term of the polynomial. It is worth noting that the values of the column of conditionally introduced variables \tilde{x}_0 of the matrix X are taken in calculations to be equal to unity.

The non-linear polynomial (1) in the form of a pseudo-linear polynomial (3) is well structured, which makes it possible to create a convenient encoding of its terms for computer implementation of combinatorial variation of its structure. This makes it possible to organize a computer search for a structurally and parametrically optimal variant of an approximating polynomial. The need for this is due to the fact that, often, a complete polynomial does not guarantee the best accuracy. This is because the properties of certain nonlinear terms contradict the nature of the approximated dependence. However, which members will be able to describe the model in the best way we do not know in advance. This is due to the peculiarities of the curvature of the hypersurfaces approximating the experimental data of each individual fragment.

III. COMBINATORIAL MODEL OF THE REGRESSION POLYNOMIAL (CMRP)

Form (1) is convenient to organize the successive shifts combinations for the nonlinear terms defined with the polynomial power. In the program, the encoded algorithm mapping is done for a complete multifactor polynomial of any degree using the natural numbers. The increasing series structure is constructed from the indices of its polynomial coefficients and has the following form:

$$\begin{aligned}
 0, 1, \dots, n, 11, 12, \dots, 1n, 22, 23, \dots, 2n, \dots, 33, 34, \dots \\
 3n, \dots, nn, \dots 111, 112, \dots, 11n, \dots, 122, \\
 123, \dots, 12n, \dots, 222, 223, \dots, 22n, \dots, nnn, 1111, \dots \\
 111n, \dots \quad (5)
 \end{aligned}$$

The combinatorial analysis of the polynomial structure (5) shows that the number of its variants is determined by the number of combinations of the polynomial terms indices for the variables included in the formula. Consequently, the determination of the structural variants number can be performed using the well-known combinatorial formula:

$$C_k^n = \frac{n!}{(n-k)!k!} \quad (6)$$

where n is the number of independent variables of the polynomial and k is the order of the power

polynomial. If the value of C_k^n turns out to be acceptable for calculating all polynomial variants of the selected order within a reasonable time, then the structure-parametric search can be performed by a deterministic algorithm.

Since the algorithm for enumerating all possible combinations of a polynomial is NP, then for the unrealizable value (5) it becomes necessary to use heuristic algorithms for solving this type of problem. In this paper, we use the evolutionary-genetic algorithm (EGA), which is developed as a tool for searching the polynomial terms variation. To find an effective solution, the EGA paradigm is considered in this work, which implies the development of the convenient gene chromosome model for this object-oriented algorithm.

IV. POLYNOMIAL APPROXIMATION FUNCTION FOR GENE-CHROMOSOME MODEL

The essence and originality of the proposed model consists of 2 chromosomes with variable polynomial structure. The main chromosome of the polynomial (meaningful) is given by the sequence of terms of the complete polynomial of the m-th power, ordered by the mnemonic rule (1).

$$Ch=(ch_1, ch_2, \dots, ch_n) \tag{7}$$

where N is the number of terms of a complete polynomial of dimension n and of order m .

These terms $ch_i (i = \overline{1, N})$ form the genome of the main chromosome. The binary model is based on the auxiliary (structural) chromosome that forms the structure of the polynomial; its genome is given by a deuce (0, 1). The description of the structure of the polynomial is made by way, where "1" on the next line of the line means using the corresponding polynomial term in the final structure obtained by the merging of the chromosomes. The presence of "0" at some position means the exclusion of this term - the gene of the main chromosome, from the finite structure of the polynomial. In the algorithm, this means that the regression coefficient is considered zero and is not calculated. Thus, the structural chromosome has the form:

$$E = (e_1, e_2, \dots, e_n), e_i \in (0,1) \tag{8}$$

With such gene-chromosome scheme for the formation and inheritance of the properties of a polynomial, the mathematical model for the transfer of genetic information from generation to generation is built on a multiplicative basis, and is given by the expression of the following form:

$$P = Ch \cdot E^T \tag{9}$$

An obligatory condition for such a coding is strict observance of the correspondence between the position of the term in the formula and its structure proposed by the expression (1).

V. SEARCH ALGORITHM FOR OPTIMIZING A POLYNOMIAL APPROXIMATING FUNCTION

The proposed modification of the EGA, like its classic prototype, involves the use of crossing-over operator, mutation and selection. Moreover, the structural chromosome of EGA is used as the transformed chromosome.

The structure of the adjusting parameters of the EGA includes the number of populations, the population size in one generation, the probability of crossing-over and the likelihood of a mutation. At the first stage of the EGA, the initial population is formed, when the mutation operator (Fig. 1a) and single-point crossover are applied to it (Fig. 1b). Further, with the help of the selection operator for each individual received, a decision is made whether to include it or not to include it in the next generation of EGA.

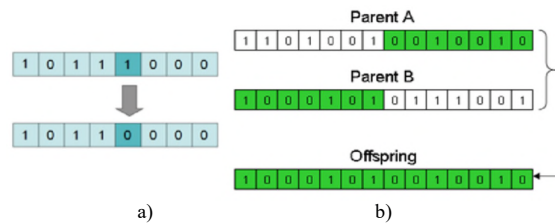


Figure 1. Illustration of mutation (a) and crossover (b) operators execution.

The binary chromosomes obtained with the help of crossing-over and mutation operators, which carry the structures information are formulated according to formula (9), where the individuals are defined as the EGA-variants of polynomials.

For selecting the next generation individuals, those already formed, with the repeated polynomial structure, are eliminated with the aid of the EGA. As a result, only the unexplored individuals are passing into the next generation.

The algorithm for calculating the coefficients of linear regression uses the specially developed software.

For the resulting regression equation, the value of the simulated dependence is calculated for the corresponding experimental points, together with the error of the approximating polynomial at each calculated point.

VI. STRUCTURAL-PARAMETRIC OPTIMIZATION ALGORITHMS FOR THE EXPERIMENTAL DATA REGRESSION DESCRIPTION

To obtain the variant, which represents the most accurate sampled data description, the calculated results are ranked in descending accuracy order. Therefore, the first output is the parametric optimum of the problem being solved.

However, there are many results that often satisfy the permissible approximation error. For example, the calculation of all structural variants of a polynomial of order 4 over an array of $m22$ defined from 20 experimental data (see Figure 2) leads, according to (5),

	500	1000	1500	2000	2500	3000
0	440	1148	2438	4470	6632	8989
10	330	944,9	2238	4299	6550	8916
20	-100	472,9	1732	3693	6060	8456
30	-700	-216,9	922,8	2911	5222	7884
40	-1550	-1037	70,5	1860	4179	6899
50	-2700	-1859	-1090	592,8	2877	5637
60	-4200	-2709	-2313	-864,5	1339	4081
70	-6300	-3515	-3562	-2445	-379	2288

Figure 2. The data array of the investigated dependence.

to the receipt of 77,811 solutions. At the same time, from this total number of variants of the polynomial (4), only a few fall under the chosen error limit. Therefore, this is quite accessible for sorting and deciding on the best structural choice by using a conventional PC.

In a particular case, a complete search of all investigated dependence variants is an acceptable approach for identifying the optimal terms combination of the describing polynomial. However, in the case when investigating the large multidimensional dependencies, a full search due to NP-completeness becomes an unsolvable problem. Therefore, it is proposed to solve it with EGA, which performs the suboptimal structure search of the describing polynomial in an acceptable time. However, it cannot guarantee the finding of the optimal polynomial structure.

In this paper, the set of polynomial variants for the 2 approaches is presented, which is characterized both by errors and by structural-parametric estimates of the complexity of the description. For such estimates, different polynomial parameters are considered, such as its order, the total number of its terms, and so on. The regularization of this, largely informal, task is not considered here. The possibilities of this approach are well traced in the example below.

VII. AN EXAMPLE OF APPLYING THE PROPOSED METHODS IN THE FRAGMENTS APPROXIMATION OPTIMIZATION.

As an example of the proposed approximation approach, the m22 matrix is examined (see Figure 2), and is defined as a fragment of an experimental data array obtained when describing the dependence of the

aerodynamic moment values of an airship with the velocity v and the angle of its roll α taken from [5][6].

At the first stage, the investigated dependence (fragment) is described with the help of a complete 4th order polynomial. Using the classical regression analysis, the coefficients of the polynomial are calculated, from which, by using the obtained coefficients, the analytical estimates of the investigated fragment experimental values for the matrix are found. On the basis of this analytical data and the available experimental data, the regression estimate errors with the greatest relative error for the experimental values are calculated. If the describing relationship accuracy of the complete n-order polynomial is insufficient, the dependence is further investigated by looking through all possible variants of the n -th order polynomial in order to identify the optimal combination of polynomial terms, as well as with EGA.

The total number of variants of polynomial (4) is 21,209, which fulfill the 5% approximation error. The study of the dependence by means of EGA is using the following parameters structure:

- Number of individuals in the population = 100
- Number of generations = 100
- The probability of crossing-over = 60%
- Probability of mutation = 30%
- Selected approximation error of 5%, for which there are 1059 variants.

When studying the structure by means of the 2 presented methods, it was noted that the data description by the complete 4-order polynomial has the maximum absolute error value of ~ 291.43, and relative error value of ~ 0.0573 (5.73%), which did not fall within the set error limit.

Among the selected variants, the correlation between the complexity of the structure of a polynomial and the approximation accuracy is provided.

In the columns of Table I and Table II, the members of the complete 4th order polynomial appearing in (4) with the corresponding code indicated in the upper line are represented. Their absence is indicated by the symbol "-". The last 2 columns give the maximum estimates for the approximated absolute and relative errors, showing the approximation accuracy for the polynomial variants.

TABLE I. POLYNOMIAL STRUCTURE RESULTS BY MEANS OF EGA

0	1	2	11	12	22	111	112	122	222	1111	1112	1122	1222	2222	Max absolute error	Relative error
0	1	2	11	12	22	111	--	--	--	1111	1112	1122	--	--	47,93095	0,0196
0	1	2	11	12	22	--	--	--	222	--	--	--	--	--	173,10336	0,02196
0	1	2	11	12	22	--	--	--	222	--	--	1122	--	--	178,19125	0,0226
0	1	2	11	12	22	--	--	122	222	--	1112	--	--	--	57,50005	0,02486

TABLE II. POLYNOMIAL STRUCTURE RESULTS BY MEANS OF A COMPLETE COMBINATORIAL SEARCH FOR ALL VARIANTS

0	1	2	11	12	22	111	112	122	222	1111	1112	1122	1222	2222	Max absolute error	Relative error
0	1	2	11	12	22	111	112	--	222	1111	1112	1122	--	--	25,32678	0,0103
0	1	2	11	12	22	111	112	--	222	--	1112	1122	--	--	25,53261	0,0104
0	1	2	11	12	22	111	--	122	222	1111	1112	1122	--	--	26,89585	0,011
0	1	2	11	12	22	111	--	122	222	--	1112	1122	--	--	27,10168	0,0110

Among all the options presented in Table 1, the top row can be considered as optimum, because it has the smallest error. However, despite the fact that the second combination has a large error of about 2%, its structure is more economical than the structure of the first, third and fourth rows. Its combination is the structurally optimal polynomial variant for investigating $m22$ dependence (see Figure 2).

In Table 2, the first line represents the best polynomial combination found by means of EGA, which can be considered as an absolute structural parametric optimum, since it has the smallest error for all orders. The second combination has more than 100% greater error and has a more economical structure than the first, but less precise for the data approximation.

For comparing the results of the developed methods, it is worth noting that the results obtained with EGA are not inferior in accuracy to the results obtained by a full search. However, the amount of time that was spent to study the same data structure, in the EGA case is 9 sec, significantly reduce the time of a full search to 65 seconds.

VII. CONCLUSION

This study has shown that the use of different criteria for estimating the accuracy of constructing the regression equation (MNC) and for estimating the accuracy of the mathematical description of data leads to a significant ambiguity between structural complexity and error, which opens the possibility for investigating the solution based on the structural parametric optimization of the created experimental data mathematical model.

The developed algorithm and its respective software implementation make this approach efficient to find the structure and parameters of the suboptimal polynomial variant for the investigated dependence fragment.

It is well known that, when the dimension and the order of the complete polynomial are sufficiently small, the search for all its possible variations is feasible. However, when the dependence dimension and approximation polynomial order increase, finding its best structure by direct selection is not possible, since it is NP-complete. Thus, this identified complexity fully justifies the proposed EGA application in order to make it solvable.

REFERENCES

[1] A. Isidori, *Nonlinear Control Systems, third edition*, Springer Verlag, London, 1995.

[2] H. K. Khalil, "Nonlinear Systems", *third edition*, Prentice Hall, Upper Saddle River, New Jersey, 2002.

[3] R. Neydorf, and Y. Sigida, "Identification of Traction and Power Characteristics of Air-Screw Propulsors in Mathematical Description of Airship," SAE 2014 Aerospace Systems and Technology Week, September 23 – 25 2014 – Cincinnati, OH, USA// SAE Technical Paper 2014-01-2134, 2014.

[4] R. Neydorf "Approximation building of mathematical models by experimental data with using of "Cut-Glue" method", Vestnik DSTU, 2014, T14, № 1 (76). - pp. 45-58.

[5] R. Neydorf, "Cut-Glue Approximation In Problems On Static And Dynamic Mathematical Model Development", Proceedings of the ASME 2014 International Mechanical Engineering Congress and Exposition, Montreal, Quebec, Canada, November 14-20, 2014.

[6] Neydorf, R., "Bivariate "cut-gluе" approximation of strongly nonlinear mathematical models based on experimental data," SAE International Journal of Aerospace, № 1, vol. 8, 2015, pp.47-54, doi:10.4271/2015-01-2394.

[7] R. Neydorf and 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, doi:10.4271/2016-01-2035

[8] R. Neydorf, I. Chernogorov, V. Polyakh and O. Yarakhmedov "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, doi:10.4271/2016-01-2033.

[9] A. Sen and M. Srivastava, "Regression Analysis — Theory", Methods, and Applications, Springer-Verlag, Berlin, 2011.

[10] D. Serber, "Linear regression analysis" Mir, 1980, pp. 450-456.