CONSTRUCTION OF APPROXIMATIONS FOR SCIENTIFIC AND TECHNICAL CALCULATIONS

V. V. Sychev, G. A. Spiridonov,
and Yu. I. Kas'yanov

UDC 518.5:536

Methods of constructing approximations for technical applications are outlined.

Formulation of the Problem

The function $u(x, y, z, \ldots)$ is specified in the form of values $u_1, u_2, \ldots, u_n$ at a discrete set of points $(x_1, y_1, z_1, \ldots), (x_2, y_2, z_2, \ldots), \ldots, (x_n, y_n, z_n, \ldots)$ with weights $W_1, W_2, \ldots, W_n$. It is required to construct a function $f(x, y, z, \ldots, a_1, a_2, \ldots, a_m)$ with the parameters $a_1, a_2, \ldots, a_m$, approximating $u(x, y, z, \ldots)$ with a mean-square error σ satisfying the condition

$$\sigma = 100 \sqrt{\sum_{k=1}^{n} (u_k - f(x_k, y_k, z_k, \ldots, \mathbf{a}))^2/(n-m)} \leqslant \sigma_0. \quad (1)$$

Here σ_0 is the permissible mean square error, %; a is a vector with the components $a_1, a_2, \ldots, a_m$.

A list of problems from thermophysics and energetics leading to the constructing of an approximation for functions specified in tabular form may be assembled. It may be arbitrarily divided into two groups.

The first group comprises problems associated with the development of tables and diagrams on the thermophysical properties of gases and liquids. This includes the construction of equations of state and equations for transfer coefficients, the description of properties at phase-equilibrium lines, the analytical representation of ideal-gas functions obtained from spectroscopic calculations, the approximation of virial coefficients, the properties of lines of extrema of thermodynamic quantities, potential curves, and many more. From the mathematical viewpoint, this class of problems reduces to the construction of approximations for functions specified in the form of a discrete set of values.

The second group of problems arises directly in the design and optimization of physical power equipment. Such problems always presume the introduction into a computer of information on the properties of working bodies and heat carriers, the characteristics of constructional materials, data on the operating conditions of equipment, etc. Here information must be fed to the computer in a form permitting its selection in any structural part of the problem with a minimum demand for machine time. This is due to the considerable fraction of computations required for the preparation and processing of the initial data for the optimizational part of the program. For example, in optimizing certain types of energy equipment, the machine time required solely for the calculation of the thermophysical properties of the working bodies and the heat carrier accounts for up to 80-90% of the total time for the solution of the problem [1].

In connection with this, very wide use is made in practice of the analytical representation of data using explicit approximations of the input variables. This approach makes economic use of the computer memory, reduces the machine time required, and significantly simplifies the algorithm for the calculation of the properties and other auxiliary data in the optimization process. However, outside the computer, the information which is to be input

Translated from Inzhenerno-Fizicheskii Zhurnal, Vol. 45, No. 5, pp. 855-860, November, 1983. Original article submitted November 2, 1982.

 0022-0841/83/4505-1342$07.50

is practically always specified in the form of tables or a set of experimental points. Therefore, the necessary approximations must be constructed preliminarily.

General theoretical questions of the approximation of functions specified in tabular form have been studied sufficiently well; however, from a practical viewpoint, two points are significant here: 1) the choice of approximating dependence; 2) the algorithm for finding the approximation coefficients.

The approximations are usually taken in the form of polynomials, piecewise-rational expansions, and various combinations of elementary functions. The coefficients of the approximations are found from the condition of a minimum of the sum

$$S = \sum_{k=1}^{n} W_k (u_k - f(x_k, y_k, z_k, \ldots, \mathrm{a}))^2 \tag{2}$$

(the least-squares method, or LSM).

To find the minimizing vector a, two fundamentally different schemes may be used.

1. Reduction to a System of Normal Equations

$$\partial S/\partial a_i = 0, \; i = 1, 2, \ldots, m, \tag{3}$$

$$\sum_{k=1}^{n} W_k (u_k - f(x_k, y_k, z_k, \ldots, \mathrm{a}))\, \partial f(x_k, y_k, z_k, \ldots, \mathrm{a})/\partial a_i = 0. \tag{4}$$

In the general case, the nonlinear system in Eq. (4) is solved by a numerical method (Jacobi, Newton, etc.). Questions of the selection of initial approximation and questions of the convergence of the iterative processes are known to constitute the problem here. On the whole, the above classical scheme is widely used in practice and is sufficiently effective in many cases, especially in the construction of approximations with linearly related parameters.

2. Direct Minimization of the Sum in Eq. (2). This scheme has been rarely used to date in approximation problems. In essence, this is a particular case of the problem of mathematical programming (absence of constraints on the desired parameters). At present, this trend is being vigorously developed in connection with the urgency of the general problem of mathematical programming. Some tens of algorithms reflecting a particular search strategy for the vector a have been proposed. It should be said that no single universal algorithm which is equally effective in all cases of minimization exists as yet, and possibly it cannot exist. Note, however, that the whole set of methods here is divided into two groups: with and without the calculation of derivatives.

Over many years, questions of the construction of effective approximation algorithms applicable to parameterization of different types have been investigated. The results for polynomials, rational fractions, and arbitrary combinations of elementary functions are given below.

Polynomials

Polynomial approximations are very popular with engineers and scientific workers, in view of their universality and expedience of programming, although in some cases other types of approximation are more effective. Three computational schemes are possible to find the coefficients of the polynomial approximations [2]. (For the sake of simplicity, polynomials of a single variable with positive powers will be considered. The results obtained below may also be extended to generalized polynomials of several variables.)

Nonorthogonal Polynomials in the LSM Scheme (Algorithm No. 1). In accordance with Eq. (4), a linear system of normal equations arises here; the specification of the system deteriorates here with increase in power of the polynomial. In connection with this, the solution of the system becomes steadily more sensitive to rounding errors and begins to depend strongly on the length of the pseudonumbers employed. Instability of "oscillating" type appears in the approximation, and the required accuracy of the approximation cannot be attained in this case. With increase in length of the mantissa of the number, the limit of instability is shifted toward increase in m, while the approximation itself improves (Table 1).

TABLE 1. Mean-Square Error of the Approximation of the Isobaric Specific Heat c_p^o/R of Nitrogen as a Function of the Temperature, Using Various Approximation Algorithms and Computers

m	Algorithm № 1			Algorithm № 2			Algorithm № 3		
	BESM-4	IRIS-80	V-6700	BESM-4	IRIS-80	V-6700	ES-1040	BESM-4	IRIS-80
5	0,228	0,228	0,228	0,228	0,228	0,228	0,228	0,228	0,228
6	0,090	0,090	0,090	0,090	0,090	0,090	0,090	0,090	0,090
7	0,054	0,054	0,054	0,054	0,054	0,054	0,054	0,054	0,054
8	0,048	0,048	0,048	0,048	0,048	0,048	0,048	0,048	0,048
9	0,052	0,024	0,024	0,024	0,024	0,024	0,031	0,024	0,024
10	0,043	0,009	0,009	0,009	0,009	0,009	0,015	0,009	0,009
11	0,037	0,009	0,009	0,010	0,009	0,009	0,020	0,009	0,009
12	0,040	0,008	0,007	0,010	0,007	0,007	0,012	0,007	0,007
13	0,044	0,008	0,003	0,136	0,003	0,003	0,043	0,003	0,003
14	0,034	0,009	0,002	0,256	0,002	0,002	0,200	0,002	0,002
15	0,028	0,006	0,002	9,016	0,002	0,002	0,615	0,002	0,001
16	0,044	0,006	0,001	10,44	0,001	0,001	0,286	0,002	0,001
17	0,033	0,006	0,001	5042	0,001	0,001	—	0,006	0,001

Quasiorthogonal Polynomials in an Orthogonal Scheme (Algorithm No. 2). In the case of a discrete set of arguments, particular orthogonal polynomials must be constructed in each specific case (for a continuous set, these are Legendre, Chebyshev, Hermite polynomials, etc.). With increase in power of the polynomial in any of the schemes developed to date, disruption of the orthogonality of the desired system of polynomials results. They will be called quasi-orthogonal. This disruption of orthogonality is reflected in the properties of the approximation, the coefficients of which are found from the usual orthogonal scheme. Ultimately, at some m, there is a breakdown in the approximation of "avalanche" type. The required accuracy of the approximation may also be unachievable here. With increase in length of the mantissa of the numbers employed, the instability boundary shifts toward increase in m, while the degree of approximation is improved (Table 1).

Quasiorthogonal Polynomials in a LSM Scheme (Algorithm No. 3). An algorithm using quasiorthogonal polynomials in a LSM scheme has been proposed. In comparison with the two preceding algorithms, this one is of higher stability with respect to the rounding errors, and allows good approximations to be obtained on a computer, working with sufficiently "short" numbers.

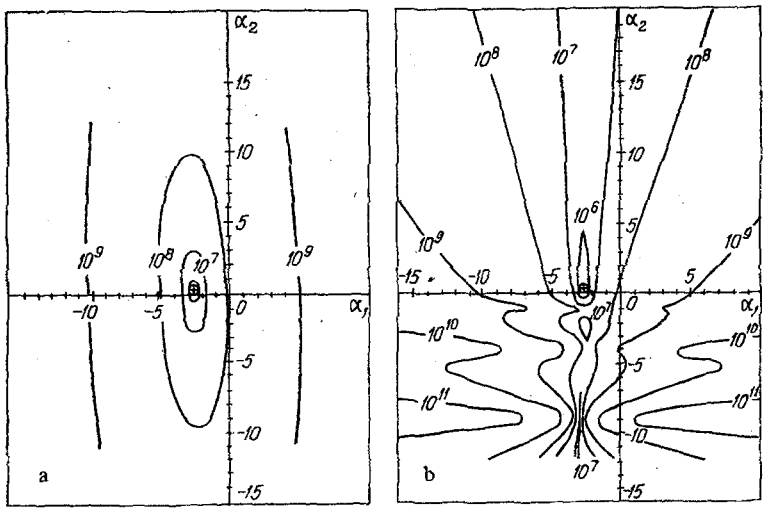

Fig. 1. Relief of the function $S = S(\alpha_1, \alpha_2)$ for linear (a) and nonlinear (b) variants of the LSM scheme in an approximation by rational fractions.

TABLE 2. The Number of Calculations of the Target Function in Using Certain Minimization Methods. The Values of the Target Function at the End of Search Are Shown in Parentheses

Type of target func.	Nelder–Mid method [7]	Rosenbrock method [7]	Powell method [9]	Newton method [6]
$S=S(\alpha_1, \alpha_2)$ (11)	398 ($0{,}56\cdot10^{-17}$)	1580 ($0{,}72\cdot10^{-17}$)	170 ($0{,}27\cdot10^{-26}$)	129 ($0{,}75\cdot10^{-28}$)
$S=S(\alpha_1, \alpha_2)$ (12)	170 ($0{,}26\cdot10^{-3}$)	303 ($0{,}26\cdot10^{-3}$)	37 ($0{,}22\cdot10^{-3}$)	18 ($0{,}22\cdot10^{-3}$)
$S=S(\alpha_1, \alpha_2)$ (13)	264 ($0{,}18\cdot10^{-3}$)	594 ($0{,}17\cdot10^{-3}$)	405 ($0{,}17\cdot10^{-3}$)	195 ($0{,}18\cdot10^{-3}$)
$S=S(\alpha_1, \ldots, \alpha_4)$ (14)	800 ($0{,}13\cdot10^{-3}$)	3160 ($0{,}13\cdot10^{-3}$)	456 ($0{,}13\cdot10^{-3}$)	163 ($0{,}13\cdot10^{-3}$)
$S=S(\alpha_1, \alpha_2)$ (15)	196 ($0{,}94\cdot10^{-3}$)	235 ($0{,}95\cdot10^{-3}$)	Diverges	72 ($0{,}94\cdot10^{-3}$)
$S=S(\alpha_1, \ldots, \alpha_5)$ (16)	623 ($0{,}35\cdot10^{-3}$)	18273 ($0{,}35\cdot10^{-3}$)	Diverges	90 ($0{,}34\cdot10^{-3}$)

Table 1 gives the results of approximating the isobaric specific heat of nitrogen in an ideal-gas state in the temperature range 10-2000°K. In the approximation, polynomials in direct powers of the temperature are used. The calculations are performed with numbers of different length: ES-1040 (seven decimal places), BÉSM-4 (10-11 decimal places), IRIS-80 (14 decimal places), V-6700 (22 and 11 decimal places).

Rational Fractions

With all their advantages, polynomials poorly transmit, and are sometimes not at all in a state to transmit, the behavior of functions with sharply expressed extrema [3]. In many cases, polynomials have poor extrapolational properties and do not satisfactorily describe derivatives at the boundaries of the approximate region. Rational fractions are found to be very effective here (for simplicity, the case of a single variable is considered):

$$y=\left(\sum_{j=0}^{m}a_jx^j\right)\Big/\left(1+\sum_{j=1}^{n}b_jx^j\right). \tag{5}$$

Two computational schemes may be used to find the coefficients $\{a_j\}$ and $\{b_j\}$.

Linear Variant of the LSM Scheme. The coefficients of the expansion in Eq. (5) are determined from the condition of a minimum of the sum

$$S=\sum_{k=1}^{N}W_k\left(y_k\left(1+\sum_{j=1}^{n}b_jx_k^j\right)-\sum_{j=0}^{m}a_jx_k^j\right)^2. \tag{6}$$

In accordance with Eq. (4), a linear system of equations which is solved by a numerical method appears here.

Nonlinear Variant of the LSM Scheme. The coefficients $\{a_j\}$ and $\{b_j\}$ are determined from the condition of a minimum of the sum

$$S=\sum_{k=1}^{N}W_k\left(y_k-\left(\sum_{j=0}^{m}a_jx_k^j\right)\Big/\left(1+\sum_{j=1}^{n}b_jx_k^j\right)\right)^2. \tag{7}$$

Minimization of Eq. (7) according to the normal-equation scheme produces a complex and unstable algorithm; successful work with this algorithm depends on the closeness of the initial vector to the global-minimum point. In such cases, it is more expedient to use methods of direct minimization without calculating the derivatives (so-called logical-search methods). Note, however, that, in the case of a linear variant of the LSM scheme, it is possible to use methods of direct minimization. However, on account of the convexity of the relief, the sum in Eq. (6) corresponds to a more optimal algorithm than Eq. (7).

This may be explained for the example of the approximation of a tabulated function of a rational fraction of the form

$$y = \left(1 + \frac{\alpha_1}{x}\right) \Big/ \left(1 + \frac{\alpha_2}{x^2}\right). \tag{8}$$

In Fig. 1a the relief of the sum of the type in Eq. (6) for the approximation in Eq. (8) constructed for a set of ten points when $1 \leqslant x \leqslant 10$ is shown:

$$S = S(\alpha_1, \alpha_2) = \sum_{k=1}^{10} W_k (y_k (1 + \alpha_2/x_k^2) - (1 + \alpha_1/x_k))^2. \tag{9}$$

The relief of a sum of the type in Eq. (7) for the same conditions as in Eq. (9) is shown in Fig. 1b:

$$S = S(\alpha_1, \alpha_2) = \sum_{k=1}^{10} W_k (y_k - (1 + \alpha_1/x_k)/(1 + \alpha_2/x_k^2))^2. \tag{10}$$

It follows from Fig. 1 that, in the first case, practically any method of direct minimization rapidly converges to the solution even if the initial vector $\alpha^{(0)}$ is sufficiently far from the minimizing vector α^*. For the sum in Eq. (10), the relief is very complex (absence of convexity, presence of local minima, etc.). Successes in minimization depend here largely on the choice of the initial approximation and the method of search, determining the search trajectory.

Remember that, in the terminology of mathematical programming, the sum S = S(a) is called the target function.

Arbitrary Combination of Elementary Functions

An extensive range of algorithms based on the use of direct minimization is investigated in this context. Special attention is paid here to the method of search without calculating the derivatives and to the question of choosing the initial approximation. The minimization procedure practically always consists of two stages. In the first, the initial approximation is chosen by the scanning method and, in the second, the iterative process determining the search strategy is undertaken.

The methods of Danilin and Pshenichnyi [4, 5], of coordinate decline [6], of Hooke and Jeeves [7], Nelder–Mid [7], Rosenbrock [7, 8], and Powell [9], and the Newton linearization method have been investigated. Calculations were performed for functionals with different degree of complexity of the relief. The main aim of the investigations was to establish more effective algorithms, on the basis of which a packet of applied approximation programs intended for a wide range of uses may be developed.

From the viewpoint of stability of the process of decline with respect to the relief, the Nelder–Mid method and the Rosenbrock method are the most suitable, while the Newton is the most rapid. The results of the investigation for some types of target functions are shown in Table 2; analytic expressions for these functions are given below:

$$S = 100(\alpha_2 - \alpha_1)^2 + (1 - \alpha_1)^2, \tag{11}$$

$$S = \sum_{k=1}^{10} W_k (y_k - (\alpha_1/x_k^{12} + \alpha_2/x_k^6))^2, \tag{12}$$

$$S = \sum_{k=1}^{10} W_k (y_k - (1/x_k^{\alpha_1} - 1/x_k^{\alpha_2}))^2, \tag{13}$$

$$S = \sum_{k=1}^{10} W_k (y_k - (\alpha_1/x_k^{\alpha_3} + \alpha_2/x_k^{\alpha_4}))^2, \tag{14}$$

$$S = \sum_{k=1}^{10} W_k (y_k - (1 + \alpha_1/x_k)/(1 + \alpha_2/x_k^2))^2, \tag{15}$$

$$S = \sum_{k=1}^{10} W_k (y_k - (\alpha_1 + \alpha_2 x_k + \alpha_3 x_k^2)/(1 + \alpha_4 x_k + \alpha_5 x_k^2))^2. \quad (16)$$

On the basis of more effective methods, a packet of applied programs for Fortran has been developed, including approximation programs: generalized polynomials of a single variable (algorithm No. 3); rational fractions (a linear variant of the scheme); functions of arbitrary form using the Nelder–Mid method; functions of arbitrary form using the Rosenbrock method; functions of arbitrary form using the Newton method. In working with approximation programs for functions of arbitrary form, users also employ the OB"EKT subprogram, in which the form of the approximating dependence and the means of input of the initial information are indicated.

LITERATURE CITED

1. L. S. Popyrin, "Determining the physical properties of the heat carriers and working bodies of thermal power plants," in: Methods of Calculating the Thermophysical Properties of the Working Bodies and Heat Carriers with Complex Optimization of Thermal Power Plants [in Russian], SÉI Sib. Otd. Akad. Nauk SSSR, Irkutsk (1979), pp. 6-17.
2. G. A. Spiridonov and Yu. I. Kas'yanov, "Approximation of functions that are specified in tabular form by the least-squares method using a quasiorthogonal basis," in: Thermophysical Properties of Substances and Materials [in Russian], No. 14, GSSSD, Moscow (1980), pp. 128-136.
3. G. A. Spiridonov and Yu. I. Kas'yanov, "Approximation of functions that are specified in tabular form by means of piecewise-rational expressions," Tr. Mosk. Energ. Inst., No. 424, 15-19 (1979).
4. Yu. M. Danilin and B. N. Pshenichnyi, "Minimization methods with accelerated convergence," Zh. Vychisl. Mat. Mat. Fiz., 10, No. 6, 1341-1354 (1970).
5. Yu. M. Danilin and B. N. Pshenichnyi, "Method of minimization without calculating derivatives," Zh. Vychisl. Mat. Mat. Fiz., 11, No. 1, 12-21 (1971).
6. N. S. Bakhvalov, Numerical Methods [in Russian], Part 1, Nauka, Moscow (1975).
7. D. M. Himmelblau, Applied Nonlinear Programming, McGraw-Hill (1972).
8. H. H. Rosenbrock and C. Storey, Computational Methods for Chemical Engineers [Russian translation], Mir, Moscow (1968).
9. M. I. D. Powell, "A method for minimizing a sum of squares of nonlinear functions without calculating derivatives," Comput. J., 7, No. 4, 303-307 (1965).