NANOSYSTEMS: PHYSICS, CHEMISTRY, MATHEMATICS, 2014, 5 (3), P. 384–390

# RELATIONSHIP PLURALITY APPROXIMATION

V. D. Lukyanov

JSC Avangard, Saint Petersburg, Russia

lukyanovvd@rambler.ru

**PACS 02.30.Mv**

An algorithm for the approximation of relationship pluralities is set by linear combinations of functions with unknown coefficients, which in part coincides in all relationship pluralities having been built using ordinary least squares. Examples of the algorithm's realization, when finding particular solutions plurality of linear nonhomogeneous differential equations, have been given.

**Keywords:** approximation, ordinary least squares, dependences plurality.

*Received: 4 June 2014*

## 1. Introduction

During experimental data processing, there arises a problem of analytical dependence recovery being a linear combination of basic functions with unknown coefficients. It is possible to face this situation when the data obtained in the experiment is not for one relationship, but for several of such equations of curves, that for instance describe the same physical process at different values of external parameters. We assume that these differences in the context of the experiment provide data that lead to analytical dependences, differing from each other by values of some linear combination of coefficients; all other coefficients for relationship pluralities are the same. In this case we shall talk about dependences plurality approximation.

In order to evaluate relationship pluralities, we may face approximation of the solution of linear nonhomogeneous differential equation (LNDE) of $N$-th order with constant coefficients with special right-hand side:

$$\frac{d^N y}{dx^N} + b_{N-1}\frac{d^{N-1}y}{dx^{N-1}} + \ldots + b_0 y = \sum_{m=1}^{M} a_m f_m(x). \tag{1}$$

Let $R$ plurality of particular solutions be built when working out $R$ various Cauchy problems for LNDE (1):

$$y_r = y_r(x) = \sum_{n=1}^{N} C_n^{(r)}\varphi_n(x) + \sum_{m=1}^{M} A_m f_m(x), \tag{2}$$

where $r$ is a serial number of a particular solution, $r = 1, 2, \ldots, R$; functions $\varphi_1(x)$, $\varphi_2(x)$, $\ldots$, $\varphi_N(x)$ are linearly independent solutions of homogeneous equation, the second summand is LNDE particular solution. Coefficients $A_1$, $A_2$, $\ldots$, $A_M$ are calculated by finding a particular solution for the nonhomogeneous differential equation and they are the same for all solutions. Coefficients $C_1^{(r)}$, $C_2^{(r)}$, $\ldots$ , $C_N^{(r)}$ are calculated from the initial conditions of the Cauchy problem or from boundary conditions and determine LNDE particular solution.

Function (2) is the expansion of $y_r(x)$ functions on a given basis, consisting of a plurality of linearly independent functions $\{\varphi_n(x)\}_1^N$ and $\{f_m(x)\}_1^M$.

The set of equations given by representation of function (3) shall be called the relationship plurality, if some of the coefficients of linear functions combination — here they are $A_1$, $A_2$, …, $A_m$ coefficients — are equal for all relationships of the plurality, and the other coefficients $C_1^{(r)}$, $C_2^{(r)}$, …, $C_N^{(r)}$ for varied relationships differ in at least one value. The number of coefficients in function (2) equals $N_R = M + RN$.

We have $R$ equations of dependences plurality for type (2). For each equation of this plurality, we know several pairs of $x$ argument and $y(x)$ function values, however, the function values are known approximately. For the $r$-th relationship, we have $S_r > 0$ of $x_{rs}$ argument values with $s = 1, 2, \ldots, S_r$ and approximate $y_{rs}$ function values. For each relationship argument, values may not match, as well as the number of these values.

In total, we have $S$ values for the entire relationship plurality, $S = \sum\limits_{r=1}^{R} S_r$, while $S > N_R$.

We pose the task to determine all $N_R$ of analytical dependence coefficients (2) with the help of these data. Briefly, we shall name the task as $(M, N, R)$ task, where $M$ is the number of coefficients whose relationship plurality values are the same, $N$ — number of coefficients whose values are different and $R$ is the number of relationships in the plurality.

To solve the problem, we shall use ordinary least squares (OLS) [1–4]. For the $s$-th curve of the plurality, with every $x_{rs}$ argument value, we shall calculate a deviation square $\delta(x_{rs})$ function preset value $y_{rs}$ of the set function (2) value, while argument value equals: $\delta(x_{rs}) = y_{rs} - y_r(x_{rs}$. Then, we shall calculate $F$ value — mean square deviation $\delta(x_{rs})$ for all argument values:

$$F = \frac{1}{S}\sum_{r=1}^{R}\sum_{s=1}^{S_r}\delta^2(x_{rs}) = \frac{1}{S}\sum_{r=1}^{R}\sum_{s=1}^{S_r}\left(y_{rs} - \sum_{n=1}^{N} C_n^{(r)}\varphi_n(x_{rs}) - \sum_{m=1}^{M} A_m f_m(x_{rs})\right)^2. \qquad (3)$$

The $F$ function depends on $N_R$ arguments: $A_1$, $A_2$, …, $A_M$ coefficients, as well as on all $C_1^{(r)}$, $C_2^{(r)}$, …, $C_N^{(r)}$ coefficients with $r = 1, 2, \ldots, R$.

For brevity, we shall use the designation for two types of function $G(x)$ averaging: data averaging for $r$-th dependence $\overline{G_r} = \frac{1}{S_r}\sum\limits_{s=1}^{S_r} G(x_{rs})$ and averaging over all data $\langle G\rangle = \frac{1}{S}\sum\limits_{r=1}^{R}\sum\limits_{s=1}^{S_r} G(x_{rs}) = \frac{1}{S}\sum\limits_{r=1}^{R} S_r\overline{G_r}$. In this designation, it will be $F = \left\langle\overline{\delta^2}\right\rangle$ function.

According to ordinary least squares, values of unknown coefficients can be obtained from minimum $F$ function condition. Setting all the partial derivatives of all function arguments to zero is a necessary condition for the function's minimum.

Calculating partial derivatives gives:

$$\frac{1}{2\,S_r}\frac{\partial F}{\partial A_i} = \left\langle\sum_{n=1}^{N} C_n^{(r)}\overline{\varphi_n(x_r)f_i(x_r)}\right\rangle + \sum_{m=1}^{M} A_m\left\langle\overline{f_m(x_r)f_i(x_r)}\right\rangle - \left\langle\overline{y_r f_i(x_r)}\right\rangle,$$

when $i = 1, 2, \ldots, M$,

$$\frac{S}{2S_r}\frac{\partial F}{\partial C_j^{(r)}} = \sum_{n=1}^{N} C_n^{(r)}\overline{\varphi_j(x_r)\varphi_n(x_r)} + \sum_{m=1}^{M} A_m\overline{\varphi_j(x_r)f_m(x_r)} - \overline{y_r\varphi_j(x_r)},$$

when $j = 1,\, 2,\, \ldots,\, N$.

Equating partial derivatives $\partial F/\partial A_i$ and $\partial F/\partial C_j^{(r)}$ to zero, and performing elementary transformations for every $r$ value, we obtain a nonhomogeneous system of linear equations (SLE), which we write in matrix form:

$$\begin{cases} \left\langle \overline{\mathbf{Q}^{(r)}}\, \mathbf{C}^{(r)} \right\rangle + \left\langle \overline{\mathbf{P}^{(r)}} \right\rangle \mathbf{A} = \left\langle \overline{\mathbf{B}_1^{(r)}} \right\rangle, \\ \overline{\mathbf{R}^{(r)}}\, \mathbf{C}^{(r)} + \overline{\mathbf{T}^{(r)}}\mathbf{A} = \overline{\mathbf{B}_2^{(r)}},\ r = 1,\, 2,\, \ldots,\, R, \end{cases} \tag{4}$$

where matrix columns are introduced $\mathbf{C}^{(r)} = \left(C_1^{(r)},\, C_2^{(r)}, \ldots,\, C_N^{(r)}\right)^T$, $\mathbf{A} = (A_1,\, A_2, \ldots,\, A_M)^T$, $\overline{\mathbf{B}_1^{(r)}} = \left(\overline{y_r f_1(x_r)}, \overline{y_r f_2(x_r)}, \ldots,\, \overline{y_r f_M(x_r)}\right)^T$, $\overline{\mathbf{B}_2^{(r)}} = \left(\overline{y_r \varphi_1(x_r)}, \overline{y_r \varphi_2(x_r)}, \ldots,\, \overline{y_r \varphi_N(x_r)}\right)^T$, $T$ is the operation of matrix transposition, $\overline{\mathbf{Q}^{(r)}} = (q_{mn})_{M,N} = \left(\overline{f_m(x_r)\, \varphi_n(x_r)}\right)_{M,N}$ – matrix of $M\times N$ size with $q_{mn} = \overline{f_m(x_r)\, \varphi_n(x_r)}$, $\overline{\mathbf{P}^{(r)}} = \left(\overline{f_m(x_r) f_j(x_r)}\right)_{M,M}$, $\overline{\mathbf{R}^{(r)}} = \left(\overline{\varphi_n(x_r)\varphi_j(x_r)}\right)_{N,N}$, $\overline{\mathbf{R}^{(r)}} = \left(\overline{\varphi_n(x_r)\varphi_j(x_r)}\right)_{N,N}$ elements, in this designations it is believed that during matrix averaging, the averaging of its elements also takes place.

In order to solve SLE (4), $\mathbf{C}^{(r)}$ column matrices shall be excluded from its first matrix equation with the help of other $R$ matrix equations. Taking into account the nonsingularity of $\overline{\mathbf{Q}^{(r)}}$ matrix, at all $r = 1,\, 2,\, \ldots,\, R$, we obtain:

$$\mathbf{C}^{(r)} = \left(\overline{\mathbf{R}^{(r)}}\right)^{-1} \overline{\mathbf{B}_2^{(r)}} - \left(\overline{\mathbf{R}^{(r)}}\right)^{-1} \overline{\mathbf{T}^{(r)}}\mathbf{A}. \tag{5}$$

Substitution of proportion (5) into the first matrix equation in SLE (4) gives a linear matrix equation for the unknown $\mathbf{A}$ matrix, which can be rewritten as

$$\left(\left\langle \overline{\mathbf{P}^{(r)}} \right\rangle - \left\langle \overline{\mathbf{Q}^{(r)}} \left(\overline{\mathbf{R}^{(r)}}\right)^{-1} \overline{\mathbf{T}^{(r)}} \right\rangle\right) \mathbf{A} = \left\langle \overline{\mathbf{B}_1^{(r)}} \right\rangle - \left\langle \overline{\mathbf{Q}^{(r)}} \left(\overline{\mathbf{R}^{(r)}}\right)^{-1} \overline{\mathbf{T}^{(r)}} \right\rangle.$$

Assuming nondegeneracy of this equation, we find the required $\mathbf{A}$ matrix whose elements are $A_m$ coefficients in function (2)

$$\mathbf{A} = \left(\left\langle \overline{\mathbf{P}^{(r)}} \right\rangle - \left\langle \overline{\mathbf{Q}_1^{(r)}} \left(\overline{\mathbf{R}^{(r)}}\right)^{-1} \overline{\mathbf{T}^{(r)}} \right\rangle\right)^{-1} \left(\left\langle \overline{\mathbf{B}_1^{(r)}} \right\rangle - \left\langle \overline{\mathbf{Q}^{(r)}} \left(\overline{\mathbf{R}^{(r)}}\right)^{-1} \overline{\mathbf{T}^{(r)}} \right\rangle\right).$$

After finding $\mathbf{A}$ column matrix by means of formulas (7) we shall calculate $\mathbf{C}^{(r)}$ column matrices.

Let us consider the implementation of this technique for solving of some particular $(M,\, N,\, R)$ tasks. It is necessary to note that the solution of $(1,\, 1,\, R)$ task at approximation of $R$ linear function pluralities $y_r = C^{(r)}x + A$ is given in [2].

## 2. Function plurality approximation in (1, 1, *R*) task

Let's define the relationship plurality $y_r = y_r(x) = C^{(r)}\varphi(x) + A\, f(x)$.

SLE (4) takes the form of:

$$\begin{cases} \left\langle \overline{f^2(x_r)} \right\rangle A + \left\langle \overline{\varphi(x_r) f(x_r)}\, C^{(r)} \right\rangle = \left\langle \overline{y_r\, f(x_r)} \right\rangle, \\ \overline{f(x_r)\varphi(x_r)}A + \overline{\varphi^2(x_r)}\, C^{(r)} = \left\langle \overline{y_r\, \varphi(x_r)} \right\rangle,\ r = 1,\, 2,\, \ldots,\, R. \end{cases} \tag{6}$$

$C^{(r)}$ values shall be excluded from the first equation by means of solving of the following $R$ equations:

$$C^{(r)} = \frac{1}{\overline{\varphi^2(x_r)}} \left( \overline{y_r\,\varphi(x_r)} - A\overline{f(x_r)\varphi(x_r)} \right), \tag{7}$$

with $r = 1, 2, \ldots, R$. Exclusion of $C^{(r)}$ values leads to a linear equation for the $A$ value. The solution of this equation has the following form: $A = \Delta_1/\Delta$, where

$$\Delta_1 = \left\langle \overline{y_r f(x_r)} \right\rangle - \left\langle \frac{\overline{y_r\,\varphi(x_r)}}{\overline{\varphi^2(x_r)}} \overline{f(x_r)\varphi(x_r)} \right\rangle, \Delta = \left\langle \overline{\psi^2(x_r)} \right\rangle - \left\langle \frac{\overline{f(x_r)\varphi(x_r)}^2}{\overline{\varphi^2(x_r)}} \right\rangle.$$

After finding $A$ value, we shall calculate $C^{(r)}$ coefficients using function (7).

## 3. Functions plurality approximation in (2, 1, *R*) task

Having the relationship plurality, $y_r = y_r(x) = C^{(r)}\varphi(x) + A_1\,f_1(x) + A_2\,f_2(x)$, SLE (4) takes the form:

$$\begin{cases} \left\langle \overline{\varphi(x_r)f_1(x_r)}\,C^{(r)} \right\rangle + \left\langle \overline{f_1^2(x_r)} \right\rangle A_1 + \left\langle \overline{f_1(x_r)f_2(x_r)} \right\rangle A_2 = \left\langle \overline{y_r\,f_1(x_r)} \right\rangle, \\ \left\langle \overline{\varphi(x_r)f_2(x_r)}\,C^{(r)} \right\rangle + \left\langle \overline{f_1(x_r)f_2(x_r)} \right\rangle A_1 + \left\langle \overline{f_2^2(x_r)} \right\rangle A_2 = \left\langle \overline{y_r\,f_2(x_r)} \right\rangle, \\ \overline{\varphi^2(x_r)}\,C^{(r)} + \overline{\varphi(x_r)f_1(x_r)}\,A_1 + \overline{\varphi(x_r)f_2(x_r)}\,A_2 = \overline{y_r\,\varphi(x_r)},\ r = 1, 2, \ldots, R. \end{cases} \tag{8}$$

From the last $R$ equations, we obtain the following:

$$C^{(r)} = \frac{1}{\overline{\varphi^2(x_r)}} \left( \overline{y_r\,\varphi(x_r)} - A_1\overline{\varphi(x_r)f_1(x_r)} - A_2\overline{\varphi(x_r)f_2(x_r)} \right), \tag{9}$$

with $r = 1, 2, \ldots, R$. After exclusion of $C^{(r)}$ value from the first two equations in SLE (10) we move to SLE of the second order for matrix $\mathbf{A} = (A_1,\ A_2)^T$: $\mathbf{VA} = \mathbf{D}$, where $\mathbf{V}$ matrix and $\mathbf{D}$ column matrix have the following elements:

$$V_{ij} = \left\langle \overline{f_i(x_r)f_j(x_r)} \right\rangle - \left\langle \overline{\varphi(x_r)f_i(x_r)} \frac{\overline{\varphi(x_r)f_j(x_r)}}{\overline{\varphi^2(x_r)}} \right\rangle, D_i = \left\langle \overline{y_r\,f_i(x_r)} \right\rangle - \left\langle \overline{\varphi(x_r)f_i(x_r)}\,\frac{\overline{y_r\,\varphi(x_r)}}{\overline{\varphi^2(x_r)}} \right\rangle,$$

where $i,\ j = 1, 2$.

Let's find the required $\mathbf{A} = \mathbf{V}^{-1}\mathbf{D}$ matrix. With the help of $A_1$ and $A_2$ coefficients obtained from the function (9), we shall find $C^{(r)}$ coefficients with $r = 1, 2, \ldots, R$.

## 4. Second order LNDE partial solutions plurality approximation according to experimental data for partial solutions (task (1, 2, *R*))

For a second order LNDE, we have a general solution in the form of:

$$y_r(x) = C_1^{(r)}\varphi_1(x) + C_2^{(r)}\varphi_2(x) + Af(x). \tag{10}$$

Coefficients $C_1^{(r)}$, $C_2^{(r)}$ and $A$ in function (10) shall be obtained from SLE (4):

$$\begin{cases} \left\langle C_1^{(r)}\overline{\varphi_1(x_r)f(x_r)}\right\rangle + \left\langle C_2^{(r)}\overline{\varphi_2(x_r)f(x_r)}\right\rangle + A\left\langle \overline{f^2(x_r)}\right\rangle = \left\langle \overline{y_r(x_r)f(x_r)}\right\rangle, \\ C_1^{(r)}\,\overline{(\varphi_1(x_r))^2} + C_2^{(r)}\,\overline{\varphi_1(x_r)\varphi_2(x_r)} + A\,\overline{\varphi_1(x_r)f(x_r)} = \overline{y_r(x_r)\varphi_1(x_r)}\ , \ r = 1,\, 2,\, \ldots R, \\ C_1^{(r)}\,\overline{\varphi_1(x_r)\varphi_2(x_r)} + C_2^{(r)}\overline{(\varphi_2(x_r))^2}\ + A\,\overline{\varphi_2(x_r)f(x_r)} = \overline{y_r(x_r)\varphi_2(x_r)},\ r = 1,\, 2,\, \ldots R. \end{cases} \tag{11}$$

From every $r$-th pair, consisting of the second and third equation of function (11), we shall express the unknown $C_1^{(r)}$ and $C_2^{(r)}$ values, through the sought quantity $A$:

$$C_1^{(r)} = \frac{T_{12}(x_r) - A\,P_{12}(x_r)}{\Delta(x_r)}, \qquad C_2^{(r)} = \frac{T_{21}(x_r) - A\,P_{21}(x_r)}{\Delta(x_r)}, \tag{12}$$

where

$$T_{ij}(x_r) = \overline{y_r(x_r)\varphi_i(x_r)} \cdot \overline{\varphi_j^2(x_r)} - \overline{\varphi_1(x_r)\varphi_2(x_r)} \cdot \overline{y_r(x_r)\varphi_j(x_r)},$$
$$P_{ij}(x_r) = \overline{f_r(x_r)\varphi_i(x_r)} \cdot \overline{\varphi_j^2(x_r)} - \overline{\varphi_1(x_r)\varphi_2(x_r)} \cdot \overline{f_r(x_r)\varphi_j(x_r)}.$$

Using formulas (12), we shall exclude $C_1^{(r)}$ and $C_2^{(r)}$ values from the first equation of function (11). Thus we shall obtain a linear equation for $A$ value, whose solution has the form:

$$A = \frac{\Delta_1}{\Delta}, \tag{13}$$

where

$$\Delta = \left\langle \overline{f^2(x)}\right\rangle - \left\langle \overline{\varphi_1(x_r)f(x_r)}\frac{P_{12}(x_r)}{\Delta(x_r)}\right\rangle - \left\langle \overline{\varphi_2(x_r)f(x_r)}\frac{P_{21}(x_r)}{\Delta(x_r)}\right\rangle,$$
$$\Delta_1 = \left\langle \overline{y(x)f(x)}\right\rangle - \left\langle \overline{\varphi_1(x_r)f(x_r)}\frac{T_{12}(x_r)}{\Delta(x_r)}\right\rangle - \left\langle \overline{\varphi_2(x_r)f(x_r)}\frac{T_{21}(x_r)}{\Delta(x_r)}\right\rangle.$$

## 5. Example

Figure 1 shows the result of LNDE partial solutions approximation

$$y'' + y + 9.250y = -0.275\sin(4x).$$

Its general solution has the form:

$$y = C_1e^{-0.5x}\cos(4x) + C_2e^{-0.5x}\sin(4x) + 0,3\,\sin(4x). \tag{14}$$

In order to set the initial data, values of $x$ arguments in function (14) have been obtained using a generator of uniformly distributed random numbers in the interval [0,3] for curves 1 and 3, and in the interval [1,2] for curve 2. Initial data for approximation, simulating experimental values were obtained after addition of calculated values, according to function (14) with the values of normally distributed random variable with zero mean of distribution and mean-square deviation $\sigma = 0.1$.

Initial data are marked at Fig. 1: crosses — for the first solution, points — for the second, squares — for the third one. Solid lines — graphics of functions approximating initial data, where the coefficients are calculated according to functions (12) and (13). Dashed lines shows a dependence diagram of type (14) with coefficients $C_1 = 1$, $C_2 = 1$ (curve 1), $C_1 = -1$, $C_2 = 1$ (curve 2), $C_1 = 0$, $C_2 = 1$ (curve 3).

Figure 2 compares the proposed OLS method of relationship plurality recovery and a traditional OLS for one relationship. We approximate the data for curve No. 2 at Fig. 1. Dashed and solid lines are taken from Fig. 1 and are constructed by method of relationship plurality approximation. Dash-dot line is the reconstructed relationship of a private solution

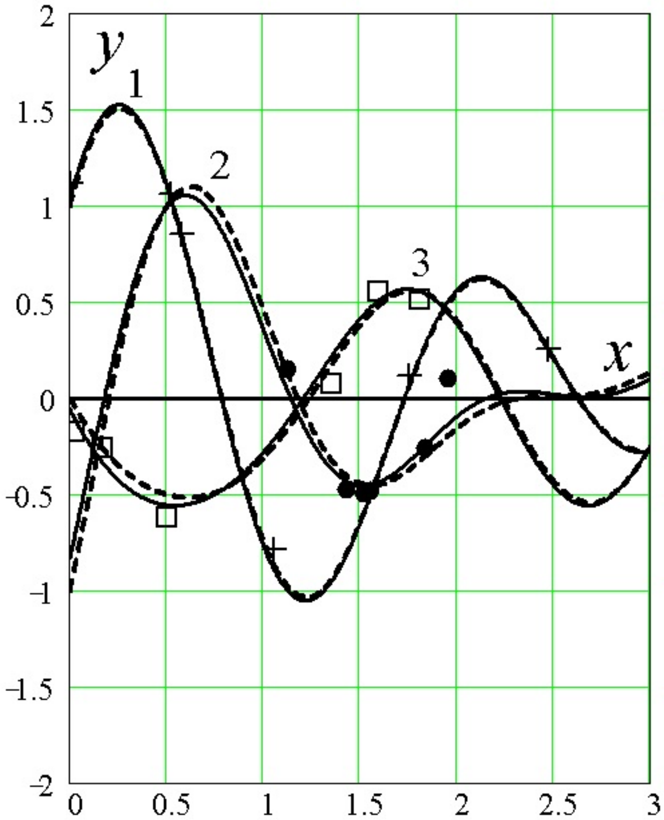


FIG. 1. Results of approximation of family of private solutions of LNDE

only according to data for curve No. 2. The calculated formulae are obtained by solving a SLE, if we assume that there is only one relationship ($R = 1$). An unsatisfactory result on the last relationship recovery at [0, 3] is explained by small data quantity for this curve, grouped near the middle of the gap. However, curve approximation as a representative of curves plurality gives quite a satisfactory result.

The proposed OLS usage for analytical relationship plurality approximation allows consideration of a certain function feature, combining them into a plurality. Relationships are set by a linear combination of known functions with the desired coefficients. Analytic plurality properties are given by the fact that a part of linear combination coefficients are the same for all the relationship pluralities. Coefficients are calculated using OLS, which provides a minimum of the average of data deviation square for all functional relationship pluralities. Combining data for all functional relationships leads to good results for their approximation, even when only having a small amount of data for some of the relationships.

Practical application of the proposed approximation algorithm of function plurality showed the recovery efficiency for the analytic dependence, included into the relationship plurality, with insufficient quantity of measured values or with unsuccessful location of interpolation point of some curves in this plurality. Availability of data for other plurality relationships allows quite satisfactorily the approximation of the equation and these curves.

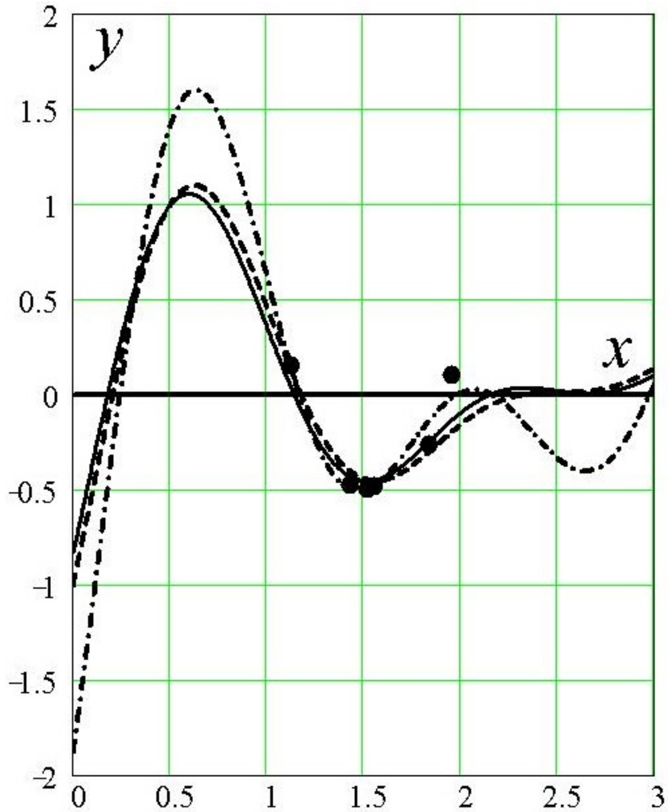


FIG. 2. Comparison of two approximations of one of solutions of LNDE. The first approximation (the continuous line) is received with use of all data for family of curves. The second approximation (the dash-dotted line) — according to the data noted by points, only for this decision. The shaped line — the exact solution of LNDE

## References


[1] Kobzar' A.I. *Applied Mathematical Statistics. For Engineers and Scientists*. Moscow, Fizmatlit, 2006.
[2] Seber G.A.F. *Linear Regression Analysis*. John Wiley and Sons. New York, London, Sydney, Toronto, 1977.
[3] Williams E.J. *Regression Analysis*. New York, Wiley, 1959.
[4] Sprent P. *Models in Regression and Related Topics*. London, Methuen, 1969.
[5] Lukyanov V.D. On the Issue of Interpolation Approximation Polynomial Construction. *Nanosystems: Physics, Chemistry, Mathematics*, **3** (6), P. 5–15 (2012).