CSCanada
Studies in Mathematical Science
Vol. 4, No. 2, 2012, pp. 17-24
DOI:10.3968/j.sms.1923845220120402.165

ISSN 1923-8444 [Print]
ISSN 1923-8452 [Online]
www.cscanada.net
www.cscanada.org

Using Rounding Function in the Problems of Finite-Element Analysis

V.S. Kosachev[a]; E.P. Koshevoy[a]; S.A. Podgornyy[a],*

[a] Kuban State Technological University, Russia.
*Corresponding author.
Address: Flat 13, 9 Karyakina Str, 350072 Krasnodar, Russia.

Received 7 January, 2012; accepted 20 April, 2012

Abstract
This work offers to supplement polynomial elements used by the step function of Heaviside and rounding functions which allows to simplify and formalise the record of test piecewise continuous function applying it for continual problems solution by the method of Galerkin.

Key words
Finite element method; Galerkin; Rounding function

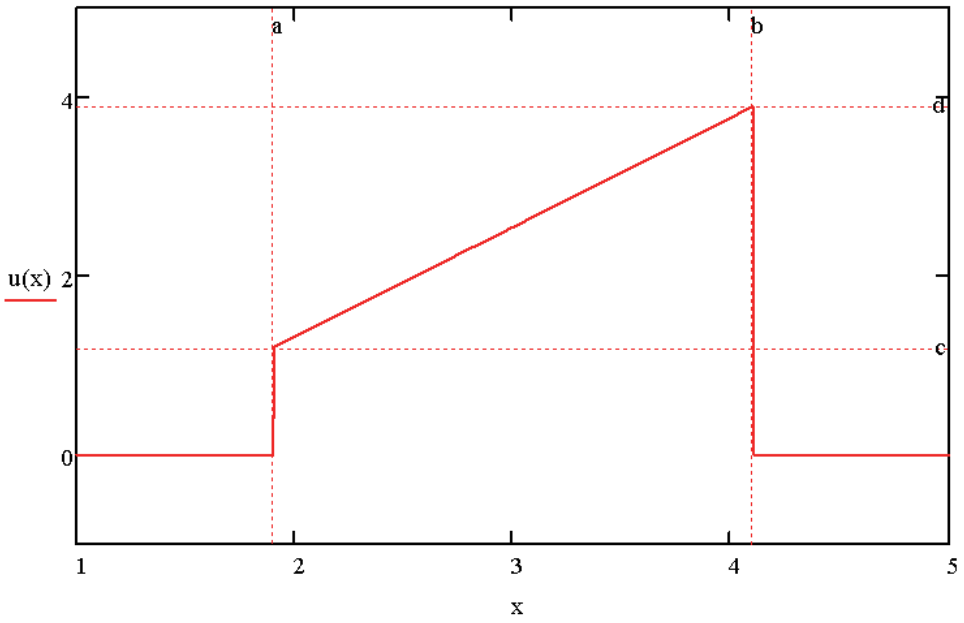
V.S. Kosachev, E.P. Koshevoy, S.A. Podgornyy (2012). Using Rounding Function in the Problems of Finite-Element Analysis. *Studies in Mathematical Science, 4*(2), 17-24. Available from URL: http://www.cscanada.net/index.php/sms/article/view/j.sms.1923845220120402.165
DOI: http://dx.doi.org/10.3968/j.sms.1923845220120402.165

Today a finite element method plays a leading role and is widely used to solve a range of mathematical physics problems [1-5]. It allows to arrange algebraic equations into an adequate scheme in relation to pivotal values of unknown functions. In this case approximation of solution is done at standard polynomial element. This work offers to supplement polynomial elements used by the step function of Heaviside and rounding functions which allows to simplify and formalise the record of test piecewise continuous function applying it for continual problems solution by the method of Galerkin.

If the problem is quite complicated, analytical solution is impossible or is so complex that it is inapplicable. In this case a family of functions defined by a finite number of parameters is considered. There is no accurate solution of the problem among such functions although the selection of parameters can roughly meet the problem equations and thus draw up its approximate solution. A specific feature in the method of finite elements is drawing up a family of functions determined by a finite number of parameters. Let us choose such family of functions u(x) where $X_{min} \le x \le X_{max}$. The interval $X_{min}\ldots X_{max}$ is one-dimensional region of the problem being solved where the existing solution is split into the finite number of parts (elements), interconnected and connected with interval range at points X_i. Within the range of each element the function is specified as a linear arithmetic expression. It is defined by its value $u(X_i)$ at the bundles and ends of the element. Taking into account that in a continual problem the function is continuous, its values in each bundle of adjacent elements must agree. To do this we introduce rounding functions: $\lfloor x \rfloor$ – the function of the "floor", which is defined as maximum integer, less or equal to x, that is $\lfloor x \rfloor = n \Leftrightarrow x - 1 < n \le x$; $\lceil x \rceil$ – the function of "ceiling", which is defined as minimum integer, more or equal to x, that is $\lfloor x \rfloor = n \Leftrightarrow x < n \le x + 1$. Using these rounding functions, we introduce the family of piecewise linear continuous functions of the following type

$$u(x)=\{\lfloor\Phi(x-a)\rfloor-\lceil\Phi(x-b)\rceil\}\cdot\left[c+(d-c)\cdot\frac{x-a}{b-a}\right] \tag{1}$$

where $\Phi(x)$ – the function of Heaviside, a unit step function the value of which is zero for negative arguments and one for positive arguments; *a, b* – the function interval *u(x)* where it is does not equal zero ($a \le b$); *c, d* – parameters of line equation at the interval *a, b*. The example of this function is shown at the graph. (Picture 1).

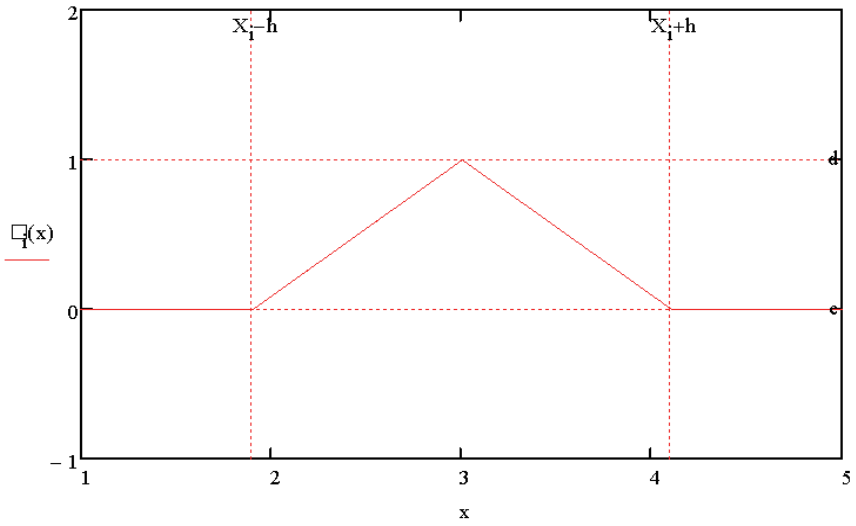


Picture 1
One-dimensional Linear Piecewise Continuous Function

From the graph (Picture 1) it can be seen that the function *u(x)* at the boundary (*a, b*) turns into zero. We use (1) to describe the test function at the arbitrary interval which is specified by half-width $h = (b - a)/2$ of one-dimensional unit final element in relation to the bundle X_i. In this case parameters of piecewise linear continuous function (1) equal correspondingly $a = X_i - h$, $b = X_i + h$, $c = 0$, $d = 1$, and the function looks like this:

$$\begin{aligned}\varphi_i(x)=&\{\lfloor\Phi[x-(X_i-h)]\rfloor-\lceil\Phi(x-X_i)\rceil\}\cdot\frac{x+h-X_i}{h}+\\&+\{\lceil\Phi(x-X_i)\rceil-\lfloor\Phi[x-(X_i+h)]\rfloor\}\cdot\frac{h-x+X_i}{h}\end{aligned} \tag{2}$$

Functions $\varphi_i(x)$ are shown as broken lines and defined by the finite number of parameters. The graph (Picture 2) shows the function of this family.



Picture 2
Piecewise Linear Function (2) to Solve Continuous One-Dimensional Problem Using the Method of Finite Elements

The method of finite elements substitutes the task of finding the function for the task of finding the finite number of its approximate values in particular bundle-points. Here if the initial problem concerning the function consists of differential equation with the corresponding boundary conditions then the aim of the finite elements method concerning its values in bundles is represented by the system of algebraic equations. Reducing the maximum number of elements leads to the increase in the number of bundles and unknown bundle parameters. In addition, the probability to meet the problem equations more precisely also gets higher which brings us closer to the desired solution. For linear problems where unknown functions and operations with them are included in all problem relations only to the first power, the method of finite elements has got a sufficiently full mathematical justification [5]. Further we use a linear problem whose solution is brought by the method of finite elements to solving the systems of linear algebraic equations. Let's consider the application of the method in the definition of one-dimensional temperature profile, specified at the initial time (τ=0 – starting condition). In this case the test function defined by the equation (3) is represented by the linear combination of functions (2) with coefficients $u_i=u_i(0)$:

$$u(x)=\sum_{i=1}^{n} u_i \cdot \varphi_i(x) \qquad (3)$$

To make sure that $u(x_i) = u_i$ in all bundles X_i, functions $\varphi_i(x)$ must meet conditions $\varphi_i(X_i) = 1$ and $\varphi_i(X_j) = 0$ for all bundles X_j when $j \neq i$. Besides, in order to satisfy first-order boundary conditions determine that $u_0 = u_{n+1} = 0$. The method of final elements operates piecewise polynominal functions as $\varphi_i(x)$, different from zero within the range of small number of elements near the bundle X_i. This is what makes the method so effective. As $u(x)$ by its physical meaning must be a continuous function, we choose $\varphi_i(x)$ to be piecewise linear functions different from zero in two elements (Picture 2). Each function of this kind $\varphi_i(x)$, $i = 1, 2, \ldots, n$, equals one in X_i and zero in all other bundles. Here the range of functions $u(x)$ will consist of continuous functions linear within the boundaries of elements having bundles bends and defined by bundle value u_i, $i = 1, 2, \ldots, n$. At the ends of the interval $X_{min} \ldots X_{max}$ they turn into zero. Each of these functions could be shown as a line string. In order to define parameters u_i used in the equation (3) we form the system of linear algebraic equations by the method of Galerkin, integrating the product of test function into the family of piecewise

linear continuous functions at the domain of solution existence:

$$\int_{X_{\min}}^{X_{\max}} \left[\varphi_j(x) \cdot \sum_{i=1}^{n} u_i \cdot \varphi_i(x) \right] dx = \int_{X_{\min}}^{X_{\max}} \left[y(x) \cdot \varphi_j(x) \right] dx \tag{4}$$

where *j=1,2,...,n; y(x)* is the function of the initial temperature profile. In the system of linear algebraic equations that we have got, certain integrals on the left form squarte matrix of the banded type having a three diagonal structure, formed by two types of integrals of matrix leading diagonal:

$$m_{i,i} = \int_{X_{\min}}^{X_{\max}} \varphi_i(x) \cdot \varphi_i(x) dx = \frac{2}{3} \cdot \frac{X_{\max} - X_{\min}}{n+1} \tag{5}$$

where *i=1,2,...,n*. Both above and below the leading diagonal:

$$m_{k,k+1} = m_{l,l-1} = \int_{X_{\min}}^{X_{\max}} \varphi_l(x) \cdot \varphi_{l-1}(x) dx = \frac{1}{6} \cdot \frac{X_{\max} - X_{\min}}{n+1} \tag{6}$$

where k=1,2,…,n-1; l=2,3,…,n.

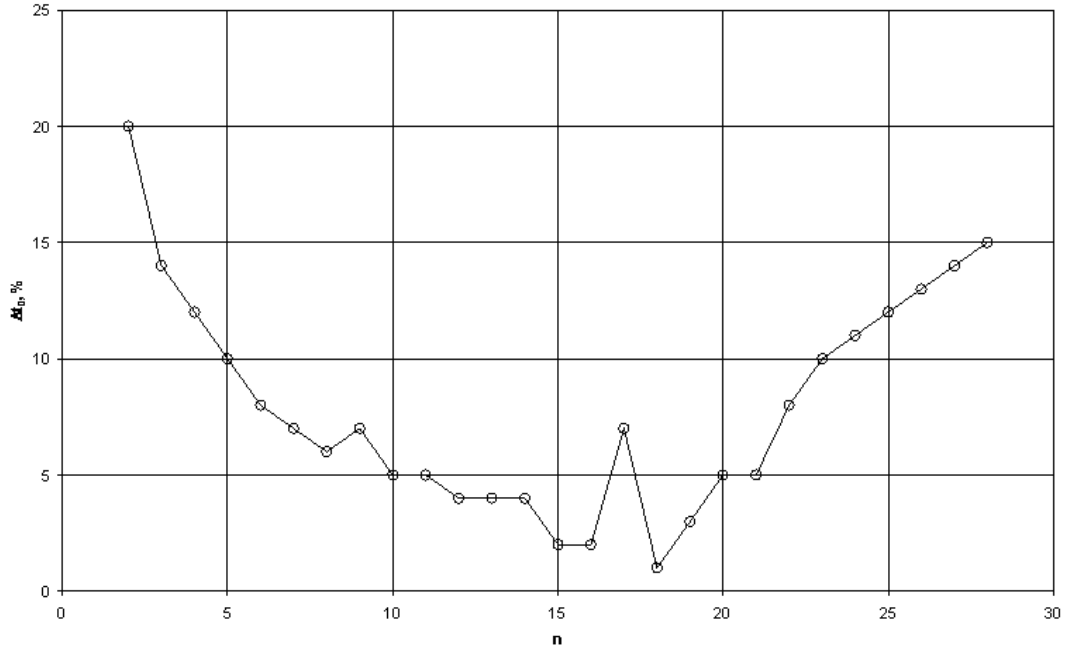
Let's simplify the problem, supposing that *y(x)=1*. In this case the column of absolute terms p_i is the value defined by the integral:

$$p_i = \int_{X_{\min}}^{X_{\max}} \varphi_i(x) dx = \frac{X_{\max} - X_{\min}}{n+1} \tag{7}$$

thereby the equation (4) takes on the matrix form:

$$\begin{vmatrix} 2/3 & 1/6 & 0 & 0 & 0 \\ 1/6 & 2/3 & 1/6 & 0 & 0 \\ 0 & 1/6 & 2/3 & 1/6 & 0 \\ \hdashline 0 & 0 & 1/6 & 2/3 & 1/6 \\ 0 & 0 & 0 & 1/6 & 2/3 \end{vmatrix} \cdot \begin{vmatrix} u_1 \\ u_2 \\ u_3 \\ \hdashline u_{n-1} \\ u_n \end{vmatrix} = \begin{vmatrix} 1 \\ 1 \\ 1 \\ \hdashline 1 \\ 1 \end{vmatrix} \tag{8}$$

Here the matrix is three diagonal and the solution could be found by Gauss method, matrix method (multiplying by the inverse matrix) or by the sweep method. For small dimensions systems solution method is not essential but with the increase of solution accuracy the number of finite elements in domain of solution existence should be increased which leads to the increase in the number of unknown in the equation (8). In order to define the effect of finite elements number within the interval X_{min}… X_{max} on the accuracy of initial approximation, we solved the sytem (8) by the matrix method in the engineering environment MathCAD with different n-numbers, defining the average value *u(x)* in this interval. The relative error for different numbers of finite elements is shown below.



Picture 3
Approximation Accuracy of the Initial Temperature Profile Depending on the Number of Finite Elements Using Matrix Method

It is clear from the data shown that the greatest approximation accuracy of the initial temperature profile is achieved at 18 finite elements on the domain of solution existence. Here the relative error of initial temperature profile approximation makes up 1.5 per cent. If a greater accuracy is required a sweep method should be used which allows to solve the systems of greater dimension without significant rounding errors. This method has relevant limits and is applied only for the systems of linear algebraic equations of the banded type. When the method of finite elements is applied the band width of the banded matrix depends on the bundles numbering. In some cases, the initial problem statement can be so inadequate that even the method of finite elements is useless and thus, the problem statement should be changed. It concerns the system of algebraic equations where slight changes of coefficients or absolute terms can lead to the significant change in solution. These equation systems are called ill-conditioned. Let's consider the sweep method for the system shown above. The sweep method is two-step. At first we calculate subsidiary quantities α_i, β_i:

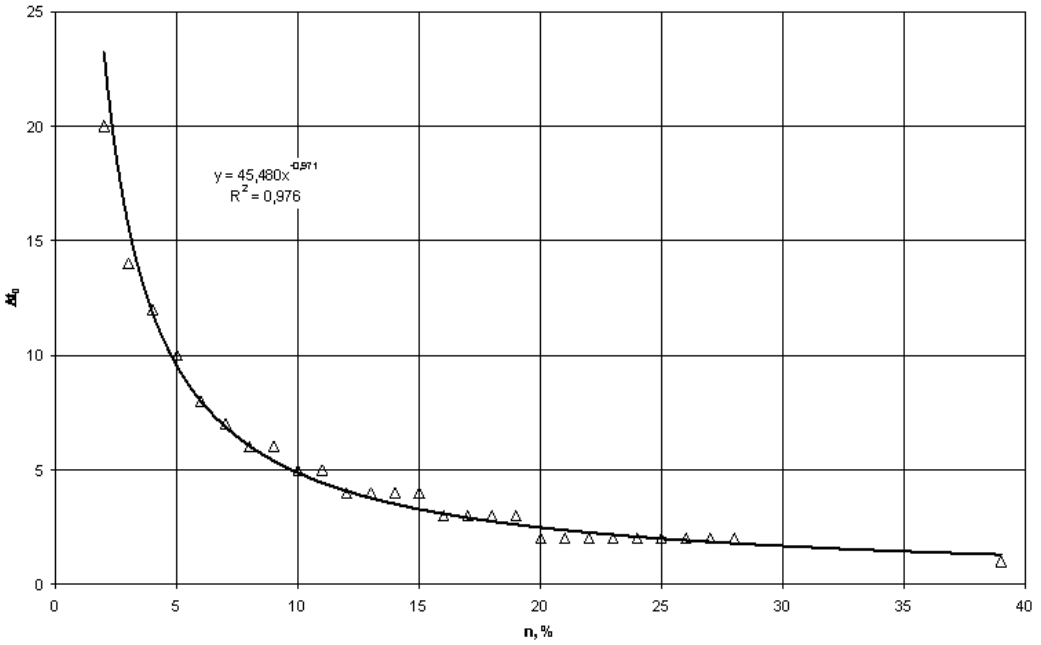
$$\alpha_0 = -\frac{1}{4}, \alpha_1 = -\frac{1}{4+\alpha_0}, \ldots, \alpha_{n-1} = -\frac{1}{4+\alpha_{n-2}} \tag{9}$$

$$\beta_0 = \frac{3}{2}, \beta_1 = \frac{6-\beta_0}{4+\alpha_0}, \ldots, \beta_{n-1} = \frac{6-\beta_{n-2}}{4+\alpha_{n-2}} \tag{10}$$

These quantities are used to calculate weighting coefficients u_i:

$$u_n = \frac{6-\beta_{n-1}}{4+\alpha_{n-1}}, u_{n-1} = \alpha_{n-1} \cdot u_n + \beta_{n-1}, \ldots, u_1 = \alpha_0 \cdot u_2 + \beta_0 \tag{11}$$

To define the effect of the finite elements number in the interval $X_{min} \ldots X_{max}$ on the accuracy of initial approximation we solved the system (9) - (11) in the engineering environment MathCAD with different n-numbers, defining the average value of *u(x)* in this interval.



Picture 4
Accuracy of Initial Temperature Profile Approximation Depending on the Number of Finite Elements Using A Sweep Method

Relative error when the number of finite elements is different is shown above (Picture 4). From the data shown above it is clear that the method of finite elements allows to reduce relative error to even less than 1 per cent when 40 and more finite elements are used. Using the values of weighting coefficients u_i, as the values of unknown temporal functions $u(\tau)$ we can turn to the boundary problem solution. In this case the test function could be presented as the product of coordinate and temporal functions:

$$\Psi(x) = \sum_{i=1}^{n} u_i(\tau) \cdot \varphi_i(x) \tag{12}$$

with boundary conditions $\Psi(0) = \Psi(1) = 0$. However, as the equation has the second derivative on the coordinate and the first derivative of the finite element has discontinuity in the bundles, we apply the following method. Let's mark $R(x) = u(\tau)\cdot\varphi''(x)-u'(\tau)\cdot\varphi(x)$ residual of the initial differential equation. The solution is accurate when $R(x) = 0$. Let's simplify this condition stating that it should be met only for n functions which correspond to test functions $u(\tau)\cdot\varphi(x)$. This method is called Galerkin method. For residual we do integration by parts on condition $\varphi(x) = \varphi_j(x)$ u $\varphi_j(0) = \varphi_j(1) = 0$, then we get the first-order system both on temporal as well as coordinate components:

$$\int_{X_{\min}}^{X_{\max}} \left[u_i(\tau)\cdot\varphi_i''(x) - u_i'(\tau)\cdot\varphi_i(x)\right]\cdot\varphi_j(x)dx = \int_{X_{\min}}^{X_{\max}} \left[-u_i(\tau)\cdot\varphi_i'(x)\cdot\varphi_j'(x) - u_i'(\tau)\cdot\varphi_i(x)\cdot\varphi_j(x)\right]dx = 0 \tag{13}$$

Here the problem includes u', which produces the system of linear algebraic equations in relation to u_i type:

$$-u_i(\tau)\cdot\int_{X_{\min}}^{X_{\max}}\left[\varphi_i'(x)\cdot\varphi_j'(x)\right]dx=u_i'(\tau)\cdot\int_{X_{\min}}^{X_{\max}}\left[\varphi_i(x)\cdot\varphi_j(x)\right]dx \tag{14}$$

Right-hand integrals (14) are represented by expressions (5) and (6) while left-hand integrals in the produced system of linear algebraic equations form the banded type square matrix of three-diagonal structure which is produced by two types of integrals on the leading matrix diagonal:

$$m_{i,i}=\int_{X_{\min}}^{X_{\max}}\varphi_i'(x)\cdot\varphi_i'(x)\,dx=n+1 \tag{15}$$

where $i=1,2,\ldots,n$. Both above and below the leading diagonal:

$$m_{k,k+1}=m_{l,l-1}=\int_{X_{\min}}^{X_{\max}}\varphi_l'(x)\cdot\varphi_{l-1}'(x)\,dx=-\frac{n+1}{2} \tag{16}$$

This matrix is symmetrical which is typical of Galerkin method. Besides, the product is not zero only on condition j = i, j = i±1, when corresponding two elements performing the test functions overlap. As a result, the matrix in this case is three diagonal as integration is performed only on two adjacent elements. Solution of the produced system of algebraic equations allows to produce the expression for derived temporary constituents and get an approximate solution in terms of numerical integration for example by Euler's method. Thereby, for a continual problem the method of finite elements performs an approximate transition to the discrete problem on the basis of corresponding piecewise-polynomial functions different from zero on several adjacent elements containing bundle X_i.

Solution (when Bi =∝) could be shown by the following system of the first-order differential equations:

$$(-1)\cdot\begin{vmatrix} n+1 & -(n+1/2) & 0 & 0 & 0\\ -(n+1/2) & n+1 & -(n+1/2) & 0 & 0\\ 0 & -(n+1/2) & n+1 & -(n+1/2) & 0\\ 0 & 0 & -(n+1/2) & n+1 & -(n+1/2)\\ 0 & 0 & 0 & -(n+1/2) & n+1 \end{vmatrix}\cdot\begin{vmatrix} u_1\\ u_2\\ u_3\\ u_{n-1}\\ u_n \end{vmatrix}=\frac{X_{\max}-X_{\min}}{n+1}\cdot\begin{vmatrix} 2/3 & 1/6 & 0 & 0 & 0\\ 1/6 & 2/3 & 1/6 & 0 & 0\\ 0 & 1/6 & 2/3 & 1/6 & 0\\ 0 & 0 & 1/6 & 2/3 & 1/6\\ 0 & 0 & 0 & 1/6 & 2/3 \end{vmatrix}\cdot\begin{vmatrix} u_1'\\ u_2'\\ u_3'\\ u_{n-1}'\\ u_n' \end{vmatrix} \tag{17}$$

The easiest way of numerical integration for the system of differential equations (17) is Euler's method which can be shown by the following design scheme:

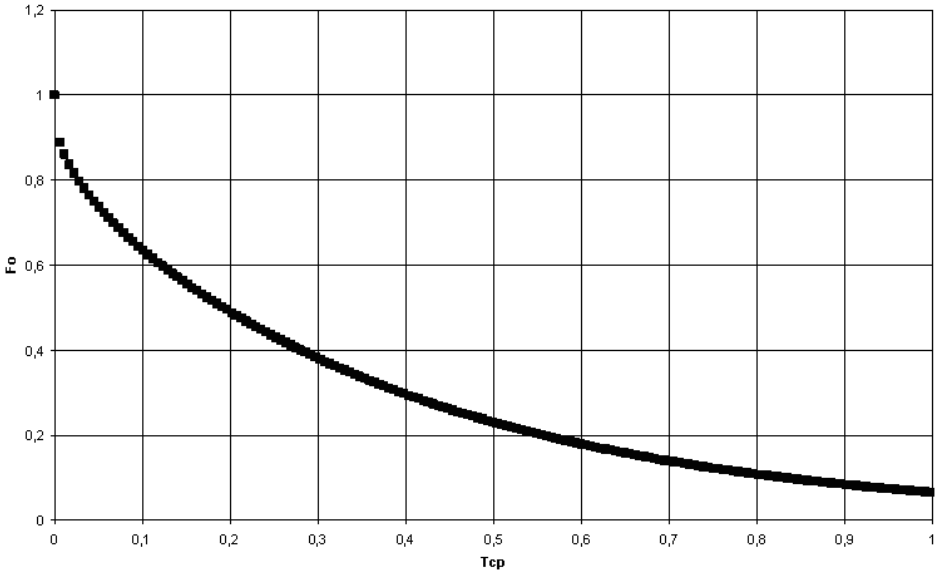
$$\begin{vmatrix} u_1(\tau_k)\\ u_2(\tau_k)\\ u_3(\tau_k)\\ u_{n-1}(\tau_k)\\ u_n(\tau_k) \end{vmatrix}=\begin{vmatrix} u_1(\tau_{k-1})\\ u_2(\tau_{k-1})\\ u_3(\tau_{k-1})\\ u_{n-1}(\tau_{k-1})\\ u_n(\tau_{k-1}) \end{vmatrix}-\left\{\begin{vmatrix} n+1 & -(n+1/2) & 0 & 0 & 0\\ -(n+1/2) & n+1 & -(n+1/2) & 0 & 0\\ 0 & -(n+1/2) & n+1 & -(n+1/2) & 0\\ 0 & 0 & -(n+1/2) & n+1 & -(n+1/2)\\ 0 & 0 & 0 & -(n+1/2) & n+1 \end{vmatrix}\cdot\frac{X_{\max}-X_{\min}}{n+1}\cdot\begin{vmatrix} 2/3 & 1/6 & 0 & 0 & 0\\ 1/6 & 2/3 & 1/6 & 0 & 0\\ 0 & 1/6 & 2/3 & 1/6 & 0\\ 0 & 0 & 1/6 & 2/3 & 1/6\\ 0 & 0 & 0 & 1/6 & 2/3 \end{vmatrix}^{-1}\right\}\cdot\begin{vmatrix} u_1(\tau_{k-1})\\ u_2(\tau_{k-1})\\ u_3(\tau_{k-1})\\ u_{n-1}(\tau_{k-1})\\ u_n(\tau_{k-1}) \end{vmatrix}\cdot\Delta\tau \tag{18}$$

where $\Delta\tau$ – time integration step of the system (17). Thus, the method of finite elements allows to transform a continual problem in partial derivatives to the system of linear piecewise continuous functions with weighting coefficients depending on time. In this case the systems of algebraic equations have a banded structure which allows to apply a sweep method for the problems of high dimensionality achieving sufficient

accuracy of solution with the limited number of finite elements covering domain of solution existence.

Considering what has been said above the solution to the problem of thermal conductivity has been found and the results are shown in picture 5 as changes of average excess temperature from the criteria of homochronity Fo.

The graph shown in picture 5 corresponds to the classical solution and the error of numerical approximation does not exceed 7.5 % of analytical decision.



Picture 5
Changes in Average Excess Temperature from Homochronity Criteria Produced for 8 Finite Elements

CONCLUSION

Suggested rounding functions allow to simplify algorithmization of continual problems using the method of finite elements.

REFERENCES

[1] Akin, J. (2005). *Finite Element Analysis with Error Estimators.* Elsevier.
[2] Chen, Z. (2005). *Finite Element Methods and Their Applications.* Springer.
[3] Solin, P. (2006). *Partial Differential Equations and the Finite Element Method*. Wiley.
[4] Thomee, V. (2006). *Galerkin Finite Element Methods for Parabolic Problems.* Springer.
[5] Strang, G., Fix G. (1977). In Marchuk, G. (Ed.), *Theory of Finite Element Method* (Agoshkova, V., Vasilenko, V. & Shaidurova, V., Trans.). Moscow: Mir.