ON THE SELECTION OF COORDINATE FUNCTIONS FOR THE SOLUTION OF BOUNDARY PROBLEMS BY GALERKIN'S METHOD

A. I. Kaidanov

UDC 518.6:536.21

We propose a method for determining the coordinate functions to be employed in Galerkin's method for the solution of boundary problems, which increases significantly the precision of the calculations in a first approximation.

Galerkin's method, because of its simplicity and universality, has found wide application in many problems of mathematical physics.

An approximate solution (ϑ) of a boundary problem is represented in the form of a linear combination of coordinate functions

$$\vartheta = \sum_{i=1}^{n} C_i u_i. \tag{1}$$

The constants C_i are determined from the orthogonality condition [1]:

$$I_V - I_S = 0; \quad I_V = \iiint_V L(\vartheta)\, u_i dV; \quad I_S = \iint_S \Gamma(\vartheta)\, u_i dS. \tag{2}$$

Here L is the differential operator of the boundary problem; Γ is the operator of the boundary conditions.

The functions u_i must possess the property of completeness in order for the approximate solution (1) to converge to the exact solution. This property is possessed, in particular, by algebraic and trigonometric polynomials. For homogeneous boundary conditions the convergence may be improved substantially by seeking ϑ in the form [1]

$$\vartheta = \omega (C + Bx + Dy + Ez + \ldots). \tag{3}$$

Here $\omega(x, y, z)$ is an arbitrary function, which, together with its partial derivatives of arbitrary order in a region V, is nonnegative and continuous and satisfies on the surface of the region S the condition

$$\omega(S) = 0. \tag{4}$$

As pointed out in [2], one of the fundamental drawbacks of Galerkin's method is the following: the properties of the boundary problem operator are accounted for by only a finite number of constants (C_i). Kantorovich's method, applied to ordinary differential equations [1], remedies this drawback only partially.

The chief merit of Galerkin's method, namely its simplicity, is maintained only when, with satisfactory precision, one can limit oneself to the first approximation in the sum (1). Further precision (second, and further, approximations) leads to substantial complexity in the computational formulas.

We present below a method for selecting the coordinate functions, which enables us to obtain a high degree of precision with only the first approximation.

We restrict ourselves to a consideration of homogeneous boundary conditions of the first kind. We illustrate the method by using, as an example, the heat conduction equation for an anisotropic homogeneous parallelepiped with internal heat sources:

Institute of Rational Mechanics and Optics, Leningrad. Translated from Inzhenerno-Fizicheskii Zhurnal, Vol. 18, No. 2, pp. 309-315, February, 1970. Original article submitted March 21, 1969.

$$\lambda_x \frac{\partial^2 t}{\partial x^2} + \lambda_y \frac{\partial^2 t}{\partial y^2} + \lambda_z \frac{\partial^2 t}{\partial z^2} + q_V = 0 \tag{5}$$

with the boundary conditions:

$$t(j = \pm l_j) = 0; \quad j = x, y, z. \tag{6}$$

In dimensionless form, assuming for simplicity that q_V = const, we obtain

$$\varepsilon_x \frac{\partial^2 N}{\partial \bar{x}^2} + \varepsilon_y \frac{\partial^2 N}{\partial \bar{y}^2} + \varepsilon_z \frac{\partial^2 N}{\partial \bar{z}^2} + 1 = 0;$$

$$N(\bar{j} = \pm 1) = 0;$$

$$\varepsilon_j = \frac{\lambda_j l_m^2}{\lambda_m l_j^2}; \qquad \bar{j} = \frac{j}{l_j}; \quad N = \frac{t\lambda_m}{q_V l_m^2}. \tag{7}$$

For the differential operator we have the expression

$$L(\vartheta) \equiv \varepsilon_x \frac{\partial^2 \vartheta}{\partial \bar{x}^2} + \varepsilon_y \frac{\partial^2 \vartheta}{\partial \bar{y}^2} + \varepsilon_z \frac{\partial^2 \vartheta}{\partial \bar{z}^2} + 1. \tag{8}$$

For homogeneous boundary conditions of the first kind,

$$\Gamma(\vartheta) \equiv \vartheta(S). \tag{9}$$

We seek an approximate solution (ϑ) in the form (3), taking into account the condition (4). Then $I_S = 0$ and, consequently,

$$I_V = 0. \tag{10}$$

We restrict ourselves to the first approximation $\vartheta = C\omega$. To determine the coordinate functions we initially use Kantorovich's method. We write ϑ in the form

$$\vartheta = f_{2x}\varphi_{1y}\varphi_{1z}; \quad \varphi_{1y} = 1 - \bar{y}^2; \quad \varphi_{1z} = 1 - \bar{z}^2.$$

Then

$$L(\vartheta) = \varepsilon_x f''_{2x}\varphi_{1y}\varphi_{1z} - f_{2x}2(\varepsilon_y\varphi_{1z} + \varepsilon_z\varphi_{1y}) + 1.$$

The orthogonality conditions may be written as follows [1]:

$$\int_0^1 \int_0^1 L(\vartheta)\varphi_{1y}\varphi_{1z}d\bar{y}d\bar{z} = 0.$$

After integrating and simplifying, we obtain:

$$f''_{2x} - p_{2x}^2 f_{2x} = -W_{2x};$$

$$p_{2x}^2 = \frac{1}{\varepsilon_x}(\varepsilon_y\eta_{1y} + \varepsilon_z\eta_{1z}); \qquad W_{2x} = \frac{\eta_{1y}\eta_{1z}}{\varepsilon_z};$$

$$\eta_{1j} = \frac{I_{11j}}{I_{21j}}; \qquad I_{11j} = \int_0^1 \varphi_{1j}d\bar{j}; \qquad I_{21j} = \int_0^1 \varphi_{1j}^2 d\bar{j}. \tag{11}$$

For selected φ_{1y} and φ_{1z}, we have $\eta_{1j} = 5/2$. From condition (4) it follows that

$$f_{2x}(\bar{x} = \pm 1) = 0. \tag{12}$$

We obtain the solution of Eq. (11), subject to the conditions (12), in the form

$$f_{2x} = \frac{W_{2x}}{p_{2x}^2}\left(1 - \frac{\operatorname{ch} p_{2x}\bar{x}}{\operatorname{ch} p_{2x}}\right) = \frac{W_{2x}}{p_{2x}^2}\varphi_{2x}. \tag{13}$$

We remark that to solve the problem by Galerkin's method, the function $\varphi_{1x} = 1 - \bar{x}^2$ would have had to be given beforehand.

The solution obtained by Kantorovich's method has the following properties: a) for $l_{y,z} \to \infty$ expressions for the temperature field in an unbounded plate are not obtainable; b) in bodies with diagonal symmetry (cube, square prism), one should also have diagonal symmetry of the temperature field, for example,

$$N(x = a, \quad y = b, \quad z = c) = N(x = b, \quad y = a, \quad z = c),$$

TABLE 1. Comparison of Approximate (ϑ_m) and Exact (N) Solutions at the Center of a Parallelepiped

$\frac{l_z}{l_x}$	$\frac{l_z}{l_y}$	N	m=1		m=2		m=∞	
			ϑ	δ, %	ϑ	δ, %	ϑ	δ, %
0	0	0,500	0,782	65,6	0,500	0,0	0,500	0,0
	0,5	0,447	0,626	40,0	0,453	+1,5	0,452	+1,3
	0,8	0,351	0,476	35,5	0,356	+1,5	0,355	+1,3
	1,0	0,289	0,391	35,0	0,293	+1,5	0,292	+1,3
0,5	0,5	0,423	0,521	23,0	0,419	−1,0	0,401	−5,1
	0,8	0,342	0,413	21,0	0,339	−1,0	0,325	−5,0
	1,0	0,286	0,348	21,0	0,283	−0,9	0,273	−4,6
0,8	0,8	0,293	0,343	17,0	0,289	−1,3	0,282	−3,8
	1,0	0,253	0,296	17,0	0,250	−1,2	0,244	−3,5
1,0	1,0	0,224	0,260	15,6	0,221	−1,3	0,218	−2,5

however the symmetry is violated because of the dissimilar form of the coordinate functions φ_{2x} and φ_{1y}, φ_{1z}:

$$\varphi_{2x}(x=a) \neq \varphi_{1y}(y=a) = \varphi_{1z}(z=a);$$

c) in contrast to Galerkin's method the method of Kantorovich allows one to partially account for properties of the operator L and to obtain a coordinate function with respect to x, which admits of a limiting approach to the exact solution for $l_x \to 0$ and $l_x \to \infty$.

The problem here is to keep the meritorious property c) of Kantorovich's method and at the same time rid ourselves of its drawbacks a), b).

As was pointed out, the function φ_{2x} is "better" than the function φ_{1x} in the sense that it takes into account the properties of the operator L. But since all the coordinate axes are equivalent, then by applying Kantorovich's method independently along the y and z axes, we could obtain precisely the same expressions for φ_{2y}, φ_{2z}, and thus for all the axes

$$\varphi_{2j} = 1 - \frac{\operatorname{ch} p_{2j}\bar{j}}{\operatorname{ch} p_{2j}}, \quad j = x, y, z.$$

In the sense indicated above, the set of functions φ_{2j} is better than the set φ_{1j}. Therefore if we now apply Galerkin's method, putting $\omega_2 = \varphi_{2x}\varphi_{2y}\varphi_{2z}$, we can eliminate the drawbacks a), b) and increase the accuracy. Moreover an approximate solution is determined by the formula $\vartheta_2 = C_2\omega_2$, and the constant C_2 is obtained from condition (10). Next, putting $\vartheta = f_{3x}\varphi_{2y}\varphi_{2z}$, and again applying Kantorovich's method, we can find an "improved" function φ_{3x}, and also φ_{3y} and φ_{3z}. This, in turn, allows us, using Galerkin's method, to find C_3, and so on. We shall call each successive application of the two methods, that of Kantorovich (for determining the φ_{mj}) and of Galerkin (for finding the C_m), a step.* Beginning with the second step (m = 2), the form of the functions φ_{mj} does not change, and with the third step, the form of the p_{mj}. This enables us to obtain recursion formulas for the ϑ_m.

We determine C_m from condition (10). We find

$$L(\vartheta_m) = C_m[\varepsilon_x\varphi''_{mx}\varphi_{my}\varphi_{mz} + \varepsilon_y\varphi''_{my}\varphi_{mz}\varphi_{mx} + \varepsilon_z\varphi''_{mz}\varphi_{mx}\varphi_{my}] + 1.$$

Taking into account that

$$\varphi''_{mj} = p^2_{mj}(\varphi_{mj} - 1);$$
$$\varphi_{mj} = 1 - \frac{\operatorname{ch} p_{mj}\bar{j}}{\operatorname{ch} p_{mj}}, \tag{14}$$

we find, from Eq. (10), after integrating and simplifying:

$$C_m = \frac{\eta_{mx}\eta_{my}\eta_{mz}}{\sum_{j=x,y,z}\varepsilon_j p^2_{mj}(\eta_{mj}-1)}; \qquad \eta_{mj} = \frac{1 - \dfrac{\operatorname{th} p_{mj}}{p_{mj}}}{1 - \dfrac{3}{2}\dfrac{\operatorname{th} p_{mj}}{p_{mj}} + \dfrac{1}{2\operatorname{ch}^2 p_{mj}}};$$
$$p^2_{mx} = \frac{1}{\varepsilon_x}\left[\varepsilon_y p^2_{m-1,y}(\eta_{m-1,y} - 1) + \varepsilon_z p^2_{m-1,z}(\eta_{m-1,z} - 1)\right]. \tag{15}$$

*Not to be confused with an approximation in the sum (1).

TABLE 2. Errors made in Calculating, by Approximate Methods, the Central (δ_0) and Mean-Volume (δ_V) Temperatures in an Isotropic Cube

δ	Approximating method used						
	1	2	3	4	5	6	7
δ_0, %	+23,9	+15,6	+9,3	−25,8	+31,0	−21,3	−2,5
δ_V, %	−9,6	−4,2	−3,0	−39,0	+7,7	−0,3	−0,3

For m = 2

$$p_{2x}^2 = \frac{5}{2}\,\frac{1}{\varepsilon_x}(\varepsilon_y + \varepsilon_z).$$

We find p_{my} and p_{mz} in a similar way.

A direct application of Galerkin's method, with a priori assignment of φ_{1j}, yields the expression

$$\vartheta_1 = \frac{25}{32}\,\frac{(1-\bar{x}^2)(1-\bar{y}^2)(1-\bar{z}^2)}{\varepsilon_x + \varepsilon_y + \varepsilon_z}$$

for ϑ, which can also be obtained from the relations (14), (15) by putting $p_j = 0$ into them and resolving the indeterminacies. Therefore, a priori assignment of φ_{1j} can be looked upon as the first step (m = 1) in finding an approximate solution, wherein $p_{1j} = 0$.

In practical calculations one should use the solution for m = 2. As the number of steps (m) increases the successive solutions (ϑ_m) converge to a limit ($\vartheta_\infty \neq N$), which is independent of the choice of φ_{1j}, whereas for small m (m = 1, 2, 3, . . .) the solution is sensitive to the choice of φ_{1j}. If the values of ϑ_∞ define the accuracy of the proposed method in this way, the derivations ϑ_1, ϑ_2, ϑ_3, . . . from the exact solution are, to a significant degree, random and depend on the form of φ_{1j}. It may happen, by chance, that ϑ_2 will give better agreement with the exact solution than ϑ_∞. One should expect in this connection, however, that ϑ_∞ and ϑ_2 are closer to N than ϑ_1. Formulas (14), (15) were used in calculating the values of ϑ_1, ϑ_2, and ϑ_∞ at the point $\bar{x} = \bar{y} = \bar{z} = 0$ for parallelepipeds of various configurations. The results were compared with the exact solution [3]

$$N(0,\ 0,\ 0) = \frac{64}{\pi^4}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\frac{(-1)^{p+q}\left(1 - \dfrac{1}{\operatorname{ch}\dfrac{\pi}{2}k}\right)}{(2p+1)(2q+1)k^2};$$

$$k = \sqrt{(2q+1)^2\varepsilon_x + (2p+1)^2\varepsilon_y}.$$

In the computations it was assumed that $\lambda_x = \lambda_y = \lambda_z = \lambda_m$ and $l_m = l_z$. Results of the comparison are shown in Table 1. Relative errors (δ_m) were determined according to the formula

$$\delta_m = \frac{\vartheta_m - N}{N}\,100\%.$$

The comparison was made at the center since there the errors of the Galerkin and Kantorovich methods are the largest. In Table 2 we give the errors made in determining, by various approximate methods, the central and mean-volume temperatures of an isotropic cube: 1) Galerkin's method, $\varphi_{1j} = \cos(\pi/2)\bar{j}$; 2) Galerkin's method, $\varphi_{1j} = 1 - \bar{j}^2$; 3) Kantorovich's method, $\varphi_{1j} = 1 - \bar{j}^2$; 4) method of collocations; $\varphi_j = 1 - \bar{j}^2$, point of collocations $\bar{j} = 1/2$; 5) method of least squares, $\varphi_j = 1 - \bar{j}^2$; 6) the method of this paper with m = 2; 7) the method of this paper for $m \to \infty$. The details of methods 4, 5 are presented in [4].

A general analysis of the accuracy of the proposed scheme, as well as that of others and, in particular, that of Galerkin [1, 2], is difficult. Therefore the data given in the tables are, meanwhile, for a single accuracy criterion.

The method proposed here can be recommended not only for the problem we have considered, but also for the case of a finite cylinder, and also for $q_V = q_V(x, y, z)$. For these problems this way of determining the coordinate functions is completely rigorous since they form a complete system. Using the relations

(1)-(4) an arbitrarily better approximation to the exact solution can be obtained. However, in solving applied problems it proves to be the case that not only the principal convergence of the system of functions is important but also the possibility of obtaining satisfactory accuracy in the very first approximation. If we proceed from this utilitarian approach, there is no need to construct a complete system of functions. It is sufficient to find a prescription enabling us to successfully construct a first approximation.

Then the domain of applicability of the proposed method can be extended even to homogeneous boundary conditions of the third kind:

$$\left[\frac{\partial N}{\partial \bar{j}} \pm B_j N\right]_{\bar{j}=\pm 1} = 0; \quad B_j = \frac{\alpha_j l_j}{\lambda_j}; \quad \Gamma(\vartheta) = \frac{\partial \vartheta}{\partial \bar{j}} \pm B_j \vartheta. \tag{16}$$

For boundary conditions (16) the form of the approximate solution (3) is inadmissible; however, by just restricting ourselves to the approximation, we can obtain the following recursion formulas:

$$\vartheta_m = C_m \varphi_{mx} \varphi_{my} \varphi_{mz}; \quad \varphi_{mj} = 1 - \frac{\operatorname{ch} p_{mj} \bar{j}}{a_{mj}};$$

$$C_m = \frac{\eta_{mx}\eta_{my}\eta_{mz}}{\sum\limits_{j=x,y,z} \varepsilon_j p_{mj}^2 (\eta_{mj} - 1)}; \quad a_{mj} = \frac{p_{mj}}{B_j} \operatorname{sh} p_{mj} + \operatorname{ch} p_{mj};$$

$$\eta_{mj} = \left(1 - \frac{\operatorname{sh} p_{mj}}{p_{mj} a_{mj}}\right) \Bigg/ \left(1 - \frac{2 \operatorname{sh} p_{mj}}{p_{mj} a_{mj}} + \frac{\frac{\operatorname{sh} 2p_{mj}}{4p_{mj}} + \frac{1}{2}}{a_{mj}^2}\right);$$

$$p_{mx}^2 = \frac{1}{\varepsilon_x} \left[\varepsilon_y p_{m-1,y}^2 (\eta_{m-1,y} - 1) + \varepsilon_z p_{m-1,z}^2 (\eta_{m-1,z} - 1)\right];$$

$$p_{2x}^2 = \frac{1}{\varepsilon_x} (\varepsilon_y B_y \eta_{1y} + \varepsilon_z B_z \eta_{1z}); \qquad \eta_{1j} = \frac{1 + \frac{1}{3} B_j}{1 + \frac{2}{3} B_j + \frac{2}{15} B_j^2};$$

$$C_1 = \frac{\eta_{1x}\eta_{1y}\eta_{1z}}{\sum\limits_{j=x,y,z} \varepsilon_j B_j \eta_{1j}}; \quad \varphi_{1j} = 1 + \frac{B_j}{2} (1 - \bar{x}^2).$$

When $B_j \to \infty$, these relations become those of Eqs. (14), (15) for boundary conditions of the first kind. Although it is difficult to construct a convergent system of functions for $B_j \neq \infty$, the first approximation here gives better agreement with the exact solution than that with $B_j \to 0$. When $B_j \to 0$, then $\delta \to 0$ also. This means that for m = 2, for all positive B_j and ε_j, the error made in calculating the temperature in the parallelepiped does not exceed 1.5%. For a finite cylinder the error is even less (two-dimensional problem).

The method given here can also be used for nonhomogeneous conditions. This is, however, a problem in itself and outside the scope of this paper. We merely remark that the error in this case turns out to be larger ($\delta \approx 10\%$).

The method proposed makes it possible to obtain an approximate solution of some boundary problems in a relatively simple form with satisfactory accuracy.

NOTATION

ϑ	is the approximate solution of boundary-value problem;
N	is the exact solution of boundary-value problem;
V, S	are the region investigated and its surface, respectively,
x, y, z	are the coordinates;
λ_x, λ_y, λ_z	are the thermal conductivity coefficients in the x, y, z directions;
t	is the temperature;
q_V	is the energy source density in region V;
l_x, l_y, l_z	are the semiedge lengths of parallelepiped in the x, y, z directions;
l_m, λ_m, q_m	are the scale values of length, thermal conductivity, and flow density;
m	is the step number in approximating process;

δ_0, δ_V are the relative errors made in calculating the maximum and mean-volume temperatures by approximate methods;

B_j is the Biot number on the boundaries $\bar{j} = \pm 1$.

LITERATURE CITED

1. L. V. Kantorovich and V. I. Krylov, Approximate Methods of Higher Analysis, Interscience, New York (1958).
2. A. I. Vaindiner, Vestn. Mosk. Gos. Un-ta, Ser. Matem.-Mekh., No. 2, 101-109 (1967).
3. G. N. Dul'nev and É. M. Semyashkin, Heat Transfer in Electronic Instrumentation [in Russian], GEI (1968).
4. B. P. Demidovich, I. A. Maron, and É. Z. Shuvalova, Numerical Methods of Analysis [in Russian], Fizmatgiz (1963).