Mathematics

I. P. Mysovskikh

1963

SovietRxiv

Abstract

Full Text

Mathematics

I. P. Mysovskikh

ESTIMATION OF THE ERROR OF A NUMERICAL SOLUTION OF A NONLINEAR INTEGRAL EQUATION

(Presented by Academician V. I. Smirnov, 3 VI 1963)

Consider the nonlinear integral equation

$$\varphi(s) = \int_a^b F(s, t, \varphi(t))\, dt + f(s), \tag{1}$$

where $F(s,t,u)$ is continuous in the variables (s,t,u) jointly, together with $F'_u(s,t,u)$ and $F''_{u^2}(s,t,u)$, in some domain $\mathcal{D}$ containing in its interior the set of points D_R:

$$a \le s \le b, \qquad a \le t \le b, \qquad |u - g(t)| \le R, \qquad R > 0, \tag{2}$$

where $g(t)$ is a function continuous on $[a,b]$. We shall assume the function $f(s)$ to be continuous on $[a,b]$. The integral equation is written in the form (1) for simplicity of notation. The results presented are valid for multidimensional integral equations.

Let an approximate numerical solution of equation (1) be given,

$$\begin{array}{c|c} t_1 & \Phi_1 \\ t_2 & \Phi_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ t_n & \Phi_n \end{array} \tag{3}$$

where $a \le t_1 < t_2 < t_3 < \dots < t_n \le b$. Below we consider the question of estimating the norm of the difference $\|\varphi - \Phi\|_I$ of the vectors

$$\varphi = (\varphi_1, \varphi_2, \dots, \varphi_n), \qquad \Phi = (\Phi_1, \Phi_2, \dots, \Phi_n),$$

where $\varphi_i = \varphi(t_i)$ and $\varphi(s)$ is the exact solution of equation (1).

One possible way of obtaining such an estimate consists in constructing an interpolating function $\Phi(s)$ from the conditions $\Phi(t_i) = \Phi_i,\ i = 1, 2, \dots, n$, and estimating in C the norm of the difference $\varphi(s) - \Phi(s)$. For this purpose one may use the theorems of L. V. Kantorovich (^{1}) and A. P. Plekhotin (^{2}). Such an estimate is useful not only for $s = t_i$, but also on the whole interval $[a, b]$, and for this reason, as a rule, it will be an overestimate. The actual execution of the error estimate for an analytical solution constructed by interpolation from the numerical solution, for large n, is associated with considerable difficulties.

The construction (explicit or implicit) of the interpolating function $\Phi(s)$ from the values $\Phi_1, \dots, \Phi_n$ is, generally speaking, unavoidable, but after such interpolation has been performed, one should deal only with the values Φ_i. Such an approach was considered in (^{3}) and is used in the present paper. Namely, equation (1) is replaced, by means of a quadrature formula, by a system of equations (algebraic and transcendental), whose solution

is the vector φ. Following S. M. Lozinskii $^{(4)}$, we shall assume that the solution of equation (1) exists and that the domain of its decomposition is known.

We shall use the notation employed in $^{(3)}$. In particular, if $\psi(t)$ is a function specified on $[a, b]$, then by ψ_i we denote its value at the point t_i: $\psi_i = \psi(t_i)$, and ψ denotes the vector $\psi = (\psi_1, \psi_2, \dots, \psi_n)$.

Assume that equation (1) has a solution $\varphi(s)$ and that this solution belongs to D_{R_1} for some $R_1 > 0$. Let the numerical solution (3) satisfy the inequality $\|g - \Phi\|_I \le R_2$. We assume that $D_{R_1+R_2} \subset D$. Thus we may suppose that a positive constant r is known such that

$$\|\varphi - \Phi\|_I \le r \tag{4}$$

(taking, for example, $r = R_1 + R_2$). Our task is to replace r in inequality (4) by a smaller number.

Take the quadrature formula

$$\int_a^b \psi(t)\,dt \simeq \sum_{j=1}^{n} A_j \psi(t_j) \tag{5}$$

with nodes $t_1, t_2, \dots, t_n$, the arguments of table (3). We write equation (1) in the form

$$\varphi(s) = \sum_{j=1}^{n} A_{jF}(s, t_j, \varphi_j) + f(s) + \alpha(s), \tag{6}$$

where

$$\alpha(s) = \int_a^b F(s,t,\varphi(t))\,dt - \sum_{j=1}^{n} A_{jF}(s,t_j,\varphi_j) \tag{7}$$

is the quadrature error of $F(s,t,\varphi(t))$ as a function of t. Here $\varphi(t)$ is the exact solution of equation (1).

Put $s = t_i$ in (6). We obtain

$$\varphi_i = \sum_{j=1}^{n} A_{jF}(t_i,t_j,\varphi_j) + f_i + \alpha_i, \qquad i = 1,2,\dots,n. \tag{8}$$

Since the numerical solution (3) is known, we may regard the numbers ρ_i as known:

$$\Phi_i - \sum_{j=1}^{n} A_{jF}(t_i,t_j,\Phi_j) - f_i = \rho_i, \qquad i = 1,2,\dots,n. \tag{9}$$

Consider the system of n equations

$$\xi_i - \sum_{j=1}^{n} A_{jF}(t_i,t_j,\xi_j) - f_i - \alpha_i = 0, \qquad i = 1,2,\dots,n, \tag{10}$$

with n unknowns $\xi_1,\xi_2,\dots,\xi_n$. Denote $x = (\xi_1,\xi_2,\dots,\xi_n)$ and introduce the vector-function $G(x)$, whose components are

$$G_i(x) = G_i(\xi_1,\xi_2,\dots,\xi_n) = \sum_{j=1}^{n} A_{jF}(t_i,t_j,\xi_j), \qquad i = 1,2,\dots,n.$$

System (10) can now be written in the form

$$x - G(x) - f - \alpha = 0. \tag{11}$$

By virtue of (8), the vector φ is a solution of equation (11). Equality (9) shows that the residual of system (11) at the numerical solution Φ is the vector $\rho - \alpha$. We shall denote the Jacobi matrix of the vector function $G(x)$ by $J(x)$, so that

$$\{J(x)\}_{ij} = -\frac{\partial G_i(\xi_1,\xi_2,\dots,\xi_n)}{\partial \xi_j} = A_j F'_u(t_i,t_j,\xi_j).$$

Theorem. *Suppose that the following conditions are satisfied:*

1) $\lambda = 1$ is not an eigenvalue of the kernel $\mathcal{K}(s,t)$, continuous for $a \leqslant s, t \leqslant b$, and an estimate is known for the norm in C of its resolvent $R(s,t,1)$:

$$\max_{a \leqslant s \leqslant b} \int_a^b |R(s,t,1)|\,dt \leqslant \Gamma.$$

2) *The inequality is satisfied*

$$\varepsilon_1 B < 1, \qquad B = 1 + \Gamma,$$

where

$$\varepsilon_1 = \max_{a \leqslant s \leqslant b} \sum_{j=1}^{n} |A_j \varepsilon(s, t_j)|,$$

$\varepsilon(s,t)$ is the quadrature error of the kernel $\mathcal{K}(s,t)$:

$$\varepsilon(s,t) = \int_a^b \mathcal{K}(s,\tau)\mathcal{K}(\tau,t)\,d\tau - \sum_{j=1}^{n} A_j \mathcal{K}(s,t_j)\mathcal{K}(t_j,t).$$

3) *Suppose that the matrix* $\mathcal{L} = (A_j \mathcal{K}(t_i, t_j))$ *satisfies the condition*

$$\|\mathcal{L} - J(\Phi)\|_I \leqslant \beta,$$

and, moreover, the number β *satisfies the inequality*

$$\beta \frac{1 + \mathcal{K}_1 B}{1 - \varepsilon_1 B} < 1, \qquad \mathcal{K}_1 = \max_{a \leqslant s \leqslant b} \sum_{j=1}^{n} |A_j \mathcal{K}(s, t_j)|.$$

4) *For any vector* x *belonging to the ball* $\|x - \Phi\|_I \leqslant r$, *the inequality is satisfied*

$$\|J(x) - J(\Phi)\|_I \leqslant \omega \|x - \Phi\|_I, \qquad \omega \geqslant$$

5) $h = \widetilde{B}^2 \omega (\|\rho\|_I + \|\alpha\|_I) \leqslant 1/2$, *where*

$$\widetilde{B} = \frac{1 + \mathcal{K}_1 B}{1 - \varepsilon_1 B - \beta(1 + \mathcal{K}_1 B)}.$$

6)

$$\frac{1-\sqrt{1-2h}}{\widetilde{B}\omega} \leqslant r \leqslant \frac{1+\sqrt{1-2h}}{\widetilde{B}\omega}.$$

Then the inequality holds

$$\|\varphi - \Phi\|_I \leqslant \frac{1-\sqrt{1-2h}}{h} B(\|\rho\|_I + \|\alpha\|_I). \tag{12}$$

The proof can be obtained on the basis of the theorem of L. V. Kantorovich on the convergence of Newton' s method (^{1}), as applied to system (11) and the initial approximation Φ to its solution, and of Theorem 2 from (^{5}). The condition that the norm of the second derivative of the left-hand side of the equation be bounded by the number ω should be replaced by the weakened Lipschitz condition for the first derivative with constant ω (^{6}). However, the theorem can be proved directly, if one uses the fact that system (11) has a solution satisfying inequality (4).

The success of applying the theorem depends to a large extent on how accurately we can estimate the norm of the vector α. Let us consider the particular case when

$$F(s,t,u) = M(s,t)H(t,u),$$

where

$$H(t,u) = \sum_{k=0}^{\infty} \beta_k(t)[u - f(t)]^k,$$

and the series converges in some domain $a \leqslant t \leqslant b,\ \|u - f(t)\| \leqslant R_3,\ R_3 > 0$. Then the equality

$$\alpha(s) = \varepsilon_M(s) + \tag{13}$$

$$+\sum_{k=1}^{\infty} \int_a^b \cdots \int_a^b \varepsilon_M(s,\tau_1,\tau_2,\dots,\tau_k) H(\tau_1,\varphi(\tau_1)) \cdots H(\tau_k,\varphi(\tau_k))\, d\tau_1 \cdots d\tau_k,$$

holds, where

$$\varepsilon_M(s) = \int_a^b M(s,t)\beta_0(t)\,dt - \sum_{j=1}^{n} A_j M(s,t_j)\beta_0(t_j),$$

$$\varepsilon_M(s, \tau_1, \tau_2, \dots, \tau_k) = \int_a^b M(s,t)M(t,\tau_1)\cdots M(t,\tau_k)\beta_k(t)\,dt-$$

$$-\sum_{j=1}^{n} A_{jM}(s,t_j)M(t_j,\tau_1)\cdots M(t_j,\tau_k)\beta_k(t_j)$$

are the quadrature errors of known functions. Since the domain in which $\varphi(s)$ lies is known, equality (13) makes it possible to give an upper bound for $\|\alpha\|_I$.

Estimate (12) is a posteriori. The quadrature formula (5) may be regarded as a parameter whose choice is at our disposal. If table (3) is obtained by the method of mechanical quadratures using formula (5), then $\|\rho\|_I = 0$, and (12) is an a priori error estimate for the numerical solution obtained by the method of mechanical quadratures.

Leningrad State University
named after A. A. Zhdanov

Received
30 V 1963

REFERENCES

1. L. V. Kantorovich, G. P. Akilov, *Functional Analysis in Normed Spaces*, 1959.
2. A. P. Plekhotin, *Investigation of Nonlinear Integral and Nonlinear Ordinary Differential Equations*, Abstract of doctoral dissertation, LSU, 1960.
3. I. P. Mysovskikh, DAN, 140, No. 4, 763 (1961).
4. S. M. Lozinskii, Izv. vyssh. uchebn. zaved., Mathematics, No. 5 (6), 52 (1958).
5. I. P. Mysovskikh, Vestn. Leningrad Univ., No. 7, ser. math., mech. and astr., issue 2, 78 (1962).
6. B. A. Vertgeim, DAN, 110, No. 5, 719 (1956).

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.