Numerical methods using defects

BORIS S. DOBRONETS

The paper deals with numerical methods using defects. The defects are used to smooth numerical solutions, to construct a posteriori error estimates and difference schemes, to correct solutions.

Методы, использующие невязки

B. C. ДОБРОНЕЦ

В работе рассматриваются методы, использующие невязки. Невязки применяются для слаживания численных решений, построения апosterиорных оценок погрешности, уточнения решений и исправления разностных схем.

Introduction

Let $\mathbb{R}^n$ be the space of $n$-dimensional vectors. In what follows, we denote interval numbers $a = [a, \bar{a}]$ with bold font: $a$, $b$, $c$, $f$, $w$, $\text{wid}(a) = \bar{a} - a$, $\mathbb{R}^n$ is the space of $n$-dimensional interval vectors, $H^l(\Omega)$ denotes the usual $L^2$-Sobolev space of order $l$. Finally, we denote by $H^l_0(\Omega)$ the space all $u \in H^l(\Omega)$ with $u = 0$ on $\partial \Omega$ (in the sense of trace).

1. Approximation of numerical solutions by finite elements of high degrees

Let us consider the Dirichlet problem:

\begin{align*}
Lu &= f, \quad x \in \Omega, \\
\phi(x) &= 0, \quad x \in \partial \Omega
\end{align*}

(1)  (2)

where $\Omega$ is a bounded open convex domain in $\mathbb{R}^2$, with piecewise smooth boundary $\partial \Omega$.

\[ Lu = -\sum_{i=1}^{2} \frac{\partial}{\partial x_i} \left( a_i \frac{\partial}{\partial x_i} u \right) + qu. \]

We assume that the coefficients $a_i \in C^1(\Omega)$, $q, f \in C(\Omega)$ and that

\[ a_i \geq c > 0, \quad q \geq 0, \quad x \in \Omega. \]

Let $T_h$ be a partition of $\Omega$ comprised of elements $T$ and

\[ \Omega = \cup T_i, \quad T_i \cap T_j = \emptyset, \quad \text{or common edge, or common corner}, \quad i \neq j. \]
384
B. S. DOBRONETS

\[ \Omega_h = \{ x_i \}_{i=1}^N \] are the nodes of the partition and \( u_h^i = u^h(x_i) \), \( i = 1, \ldots, N \) are the values of some numerical solutions of the problem (1), (2).

If \( T_h \) is a triangulation, then the finite element space \( S_h^p \) is defined by introducing a piecewise-polynomial basis on a \( T_h \):

\[ S_h^p = \{ s(x) | s \in H^1(\Omega) \cap H^0_0(\Omega), s|_T \in P^n, T \in T_h \} \]  

(3)

where \( P^n \) is the set of polynomials of degree \( n \). In the case that \( T_h \) is a rectangular partition the subspace \( S_h^p \) is defined by piecewise Hermite polynomials of degree \( n \) [4, 14].

Consider the problem of approximation of \( u_h \) by \( s \in S_h^p \), \( k \geq 3 \), such that

\[ \| s \|_{L^2(\Omega)} \leq K h^{\sigma_1}, \]

\[ \| s - f \|_{L^2(\Omega)} \leq K h^{\sigma_2}, \] for some appropriate \( \sigma_1, \sigma_2 \geq 0 \).

In order to construct finite elements of high degree [4, 14] we need some subset of the values of \( s(x) \), \( \partial_{i+j} s(x) \), \( i, j = 0, 1, 2, \ldots \) for some points \( x \in \Omega \). Let \( x_0 = (x_{0,1}, x_{0,2}) \) be one of such points.

For this purpose consider a local grid \( Z_{r,\delta,d} = \{ z_{i,j} \} \),

\[ z_{i,j,1} = \text{sign}(i) \text{abs}(i\delta)/d + x_{0,1}, \]

\[ z_{i,j,2} = \text{sign}(j) \text{abs}(j\delta)/d + x_{0,2}, \] \( i, j = 0, \pm 1, \pm 2, \ldots \)

where \( r, \delta, d \) are parameters and \( r \geq 1, \delta > 0, d > 0 \).

Define \( p \) as a polynomial \( p = \sum_{i=0}^{n_p} a_i \psi_i(x - x_0) \) such that

\[ \sum_{i=1}^{n_1} \alpha_i \| p(v_i) - u_h(v_i) \|^2 + \sum_{i=1}^{n_2} \beta_i \| Lp(v_i) - f(v_i) \|^2 \rightarrow \min \]  

(4)

where \( \psi_i \) are given linearly independent functions. Let, for definiteness,

\[ \psi_0(x) = 1, \psi_1(x) = x, \psi_2(x) = x^2, \ldots, \psi_{n_p}(x) = x_n, \]

\( n_p = (n + 2)(n + 1)/2 \)

\( \{ v_i \}_{i=1}^{n_1} \) and \( \{ v_i \}_{i=1}^{n_2} \) be the nodes of the auxiliary grid, disposed in the neighborhood of the point \( x_0 \). Here \( v_i \in \Omega_h, v_i \in Z_{r,\delta,d}, n_1 + n_2 = n_p + 1, n_1 \geq 2n + 1, \)

\[ \alpha_i = \alpha_{-1}/(\rho(v_i, x_0)^{n+1} + \alpha_{-2}), \]

\[ \beta_i = \beta_{-1}/(\rho(v_i, x_0)^{n+1} + \beta_{-2}) \]

where \( \rho(x, y) \) is a distance between the points \( x, y \); \( \alpha_{-1}, \beta_{-1} \) is expressible in terms of the accuracy of the numerical solution \( u_h \):

\[ \alpha_{-1}, \alpha_{-2}, \beta_{-1} \approx ||u - u_h||_{L^2(\Omega_h)}, \]

\[ \beta_{-2} \approx h||u - u_h||_{L^2(\Omega_h)}. \]

The problem (4) is reduced to the solution of the system of linear algebraic equations

\[ Ba = d \]
NUMERICAL METHODS USING DEFECTS

where
\[ B = \{ b_{ij} \}_{i,j=0}^{n} \]
\[ a = (a_0, a_1, a_2, \ldots, a_n) \]
\[ b_{ij} = \sum_{i=1}^{n} \alpha_i \psi_i(v_i^2 - x_0) \psi_j(v_i^2 - x_0) + \sum_{i=1}^{n} \beta_i L \psi_i(v_i^2 - x_0) L \psi_j(v_i^2 - x_0) \]
\[ d_i = \sum_{i=1}^{n} \alpha_i u^h(v_i^2) \psi_i(v_i^2 - x_0) + \sum_{i=1}^{n} \beta_i f(v_i^2) L \psi_i(v_i^2 - x_0) \]

We can put \( s(x_0) = p(x_0), \partial_{i,j} s(x_0) = \partial_{i,j}^h p(x_0) \).

**Theorem 1** [7]. Let \( u \in C^4(\Omega) \) be the exact solution, \( \Omega_h \) is a uniform rectangular grid and \( u^h \) is the numerical solution of problem (1), (2) by a certain difference scheme [12], \( s \in S^h_2 \) constructed as above and \( S^h_2 \) is the piecewise Hermite polynomials subspace. Then
\[
\| u - u^h \|_{L^2(\Omega_h)} \leq K h^2, \\
\| s \|_{L^2(\Omega_h)} \leq K h^4, \\
\| L s - f \|_{L^2(\Omega_h)} \leq K h^2
\]

where \( h \) is the meshsize, \( K \) is a constant independent of \( h \).

2. A posteriori error estimate

Let us consider the use of an approximation of numerical solutions by finite elements of high degrees for an a posteriori error estimate of boundary value problems for elliptic equations. The base of this method is the principle of monotony [2, 3, 15].

Let \( u^h \) be a numerical solution of the problem (1), (2) obtained by a certain difference scheme [12] and \( s \in S^h_2 \) constructed as above. We can use the defect
\[ \varphi(x, s) = L s - f(x), \quad x \in \Omega. \] (5)

We solve numerically the additional problem
\[ Lu_1 = 1, \quad x \text{ in } \Omega, \] (6)
\[ u_1(x) = 0, \quad x \text{ on } \partial \Omega. \] (7)

and also build \( s_1 \) in an analogous way. As a consequence of Theorem 1, for sufficiently small \( h > 0 \) the inequalities
\[ L s_1 > 1 - K h^2 > 0 \quad \text{on } \partial \Omega \]
hold. Then the interval solution has the form
\[ u = s + [\alpha, \bar{\alpha}] s_1 + [\beta, \bar{\beta}] \]

where
\[ \bar{\alpha} = \max_{\partial \Omega} (-\varphi/L s_1), \quad \alpha = \min_{\partial \Omega} (-\varphi/L s_1), \]
\[ \bar{\beta} = \max_{\partial \Omega} (-\bar{\alpha} s_1 - s, 0), \quad \beta = \min_{\partial \Omega} (-\bar{\alpha} s_1 - s, 0). \]

**Theorem 2** [7]. Assume that the assumptions of Theorem 1 hold. Then
\[ \| \text{wid}(u) \|_{L^2(\Omega)} \leq K h^2 \]

where \( h \) is the meshsize, \( K \) is a constant independent of \( h \).
3. Defect correction method

Defect correction method is the common name of special discrete Newton’s method [1, 9, 13]. In this section we consider the defect correction method for problem (1). We will use the finite element method only with linear elements, but the final solution will have accuracy corresponding to cubic elements.

The solution of (1) is understood in the following weak sense: find a function \( u \in H_0^1(\Omega) \) such that

\[
\mathcal{L}(u, v) = (f, v), \quad \forall v \in H_0^1(\Omega)
\]

where \((\cdot, \cdot)\) is the inner product in \(L_2\).

\[
\mathcal{L}(u, v) = \int_\Omega \sum_{|i| \leq 2} a_i \partial_i u \partial_i v + quvd\Omega.
\]

We define the finite element solution \( u^h \) of the problem (1) as a function from \( S_1^h \) that satisfies the equation

\[
\mathcal{L}(u^h, v) = (f, v), \quad \forall v \in S_1^h,
\]

\[
u^h = 0, \quad \text{on } \partial\Omega.
\]

Further, using the finite element solution \( u^h \) we construct \( s \in S_1(\Omega) \) according to Section 1.

Consider the identity

\[
\mathcal{L}(s, v) = (f, v) - \varphi, \quad \forall v \in H_0^1(\Omega)
\]

where \( \varphi \equiv (f, v) - \mathcal{L}(s, v) \). Subtracting (10) from (8) we have:

\[
\mathcal{L}(u - s, v) = \varphi, \quad \forall v \in H_0^1(\Omega).
\]

Denoting \( \varepsilon = u - s \) we write the equation for \( \varepsilon \) in the weak form

\[
\mathcal{L}(\varepsilon, v) = \varphi, \quad \forall v \in H_0^1(\Omega).
\]

Further, let \( \varepsilon^h \) be an approximate solution to the problem (12) by the finite element method in \( S_1^h \). Then the corrected solution is

\[
s_{\text{cor}} = s + \varepsilon^h.
\]

Theorem 3 [5]. Let \( u \in H^4(\Omega), \Omega_h \) be a uniform grid.

Then

\[
||u - s_{\text{cor}}||_{L_2(\Omega)} \leq Kh^4||u||_{H^4(\Omega)}
\]

where \( h \) is the meshsize, \( K \) is a constant independent of \( h \).

If we take the initial solution \( \varepsilon^h = 0 \) and \( s_0 = 0 \), then the correction procedure should be done several times:

\[
s_{i+1} = s_i + s_i^h, \quad i = 0, 1, \ldots
\]

where \( s_i^h \in S_1^h \) is the approximation of \( \varepsilon_i^h \).
4. The cell approximate solution method

In this section, we present a method for the construction of difference schemes for the problem (1) on nonregular grids.

Choose the node $x_0 \in \Omega_h$ and let $x_1, \ldots, x_n \in \Omega_h$ be nodes that lie in the neighborhood of $x_0$. We define the cell $Q_0$ as a minimal convex polygon, such that $Q_0 \ni x_i$, $i = 0, 1, \ldots, n$. We will seek an approximate solution $s \equiv p$ in $Q_0$ by the process (4), with $n_1 = n$, $v_i = x_i$ ($i = 1, \ldots, n$), and with $u_i^h = u^h(x_i)$ ($i = 1, \ldots, n$) as parameters. Put $u_0^h = s(x_0)$ and

$$s = s(x, u_1^h, \ldots, u_n^h).$$

Then we obtain

$$u_0^h = \sum_{i=1}^{n} \gamma_i u_i^h + F_0 h^k$$

where $\gamma_i$, $F_0$ are constants, depending on the coordinates of the points $x_i$, $i = 0, 1, 2, \ldots, n$ and coefficients of the problem (1), $h$ is the diameter of $Q_0$. If we take $x_0$ equal to each point of the mesh $\Omega_h$, then we obtain a difference scheme for $u^h$.

The precision of the difference scheme (13) depends on the precision of the approximate solution $s$. Assume, that $u_i = u(x_i)$ and the inequality

$$|u(x_0) - s(x_0, u_1, \ldots, u_n)| \leq C h^l$$

is fulfilled. Then, following [12], we have

$$|u(x) - u^h(x)| \leq C h^{l-k}, \quad x \in \Omega_h.$$ 

Observe that in the general case $l = n + 1$ and $k = 2$ for $s \in P^n$. Following this method we can construct practically all known difference schemes for elliptic and parabolic equations. Taking $s$ in the form of the generalized polynomial, we can construct some difference schemes with given properties.

4.1. Two-point boundary-value problems

Consider

$$Lu \equiv -\frac{d}{dx} \left( p \frac{d}{dx} u \right) + qu = f, \quad x \in (0, 1)$$

$$u(0) = u(1) = 0.$$ 

Suppose

$$p(x) > c_1 > 0, \quad q(x) > 0, \quad x \in (0, 1),$$

$$q, f \in C^r[0, 1], \quad p \in C^{r+1}[0, 1]$$

for some integer $r \geq 0$.

For the interval $[0, 1]$ we choose the mesh points

$$\Omega_h = \{0 = x_0 < x_1 < \cdots < x_N = 1\}, \quad N \geq 2.$$
Consider the cell $\Omega_i = [x_{i-1}, x_{i+1}]$. Let $P^n[x_{i-1}, x_{i+1}]$ be the set of polynomials of degree $n$ on the interval $[x_{i-1}, x_{i+1}]$. We will seek the approximate solution $s \in P^n$ in the cell $\Omega_i$ of the form

$$s = \sum_{i=0}^{n} a_i (x - x_i)^i$$

and require that for $s$ the conditions

$$s(x_{i-1}) = u_{i-1}^h, \quad s(x_{i+1}) = u_{i+1}^h$$

be valid. In the cell $\Omega_i$, we choose the knot sequence

$$\omega_i = \{x_{i-1} < z_1 < \cdots < x_{i-1} < z_i < \cdots < x_{i+1}\}.$$ 

In order to construct $s$ we use the methods of Section 1 with $n_1 = 2$, $v_1 = x_{i-1}$, $v_2 = x_{i+1}$, $n_2 = n - 1$, $v_3 = z_i$:

$$Ls(z_i) = f(z_i), \quad l = 1, \ldots, n - 1.$$  

Thus to determine $s$ we have to solve the system of linear algebraic equations

$$Ba = d$$

where the vector $\hat{d}$ has the form:

$$\hat{d} = (u_{i-1}^h, u_{i+1}^h, f(z_1), f(z_1), \ldots, f(z_{n-1})).$$

Note that $B$ is invertible. Let $B^{-1}$ be the inverse matrix to $B$. Then we have

$$u_i^h = s(x_i) = a_0 = \sum_{l=1}^{n+1} B_{1i}^{-1} \hat{d}_l$$

and the difference scheme

$$u_i^h = \gamma_1 u_{i-1}^h + \gamma_2 u_{i+1}^h + F_i h^2$$

where $F_i$ are linear combinations of $f(z_1), \ldots, f(z_{n-1})$.

Consider the accuracy of the scheme (18). Let $p \in P^n[x_{i-1}, x_{i+1}], u \in C^n[x_{i-1}, x_{i+1}]$. If $p$ interpolates $u$, then

$$|p(x) - u(x)| \leq K h^{n+1}, \quad x \in [x_{i-1}, x_{i+1}].$$

Hence, it follows from [12] that

$$|u(x) - u^h(x)| \leq C h^{n-1}, \quad x \in \Omega_h$$

where $K$ is a constant independent of $h$.

This scheme is similar to the "exact" difference scheme [12]. But this approach does not require the exact solution of problem (14) and is applied here, in particular, to boundary-value problems with small parameter $\varepsilon > 0, \varepsilon \ll 1$.

Consider

$$Lu = -\frac{d}{dx} \left( \frac{d}{dx} u \right) + qu = f, \quad x \in (0, 1),$$

$$u(0) = u(1) = 0.$$
We will seek the approximate solution in the cell \( \Omega_i \) of the form
\[
s = \sum_{i=0}^{n} a_i(x - x_i)^i + a_{n+1} \exp \left( \lambda_0(x)/\varepsilon \right) + a_{n+2} \exp \left( \lambda_1(x)/\varepsilon \right).
\]

The functions \( \lambda_0, \lambda_1 \) are solutions of the initial value problems
\[
\begin{align*}
\varepsilon \lambda''_0 + (\lambda'_0)^2 - q &= 0, \quad x \in (x_{i-1}, x_{i+1}) \\
\lambda_0(x_{i-1}) &= 0, \quad \lambda'_0(x_{i-1}) = \sqrt{q(x_{i-1})} \\
\varepsilon \lambda''_1 + (\lambda'_1)^2 - q &= 0, \quad x \in (x_{i-1}, x_{i+1}) \\
\lambda_1(x_{i+1}) &= 0, \quad \lambda'_1(x_{i+1}) = \sqrt{q(x_{i+1})}.
\end{align*}
\]

As a numerical example, we consider the problem [8]
\[
- \varepsilon^2 u'' + u = -\cos^2(\pi x) - 2\varepsilon^2 \pi^2 \cos(2\pi x),
\]
\[
u(0) = \nu(1) = 0.
\]

The exact solution is
\[
u(x) = \left( \exp \left( -(1-x)/\varepsilon \right) + \exp(-x/\varepsilon) \right) / \left( 1 + \exp(-1/\varepsilon) \right) - \cos^2(\pi x).
\]

The problem (21) can be solved on uniform grids \( \Omega_h \) with some meshsize \( h \), \( n = 2 \). Numerical results are presented in Table 1.

<table>
<thead>
<tr>
<th>( \varepsilon )</th>
<th>( h = 0.2 )</th>
<th>( h = 0.1 )</th>
<th>( h = 0.05 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>2.65E-3</td>
<td>1.71E-3</td>
<td>1.40E-3</td>
</tr>
<tr>
<td>0.1</td>
<td>7.90E-4</td>
<td>7.64E-5</td>
<td>5.10E-6</td>
</tr>
<tr>
<td>0.01</td>
<td>4.58E-5</td>
<td>8.70E-6</td>
<td>3.21E-7</td>
</tr>
<tr>
<td>0.001</td>
<td>5.36E-7</td>
<td>1.78E-7</td>
<td>5.96E-8</td>
</tr>
</tbody>
</table>

Table 1. Errors \( \max_{x \in \Omega_h} |u(x) - u^h(x)| \)

### 4.2. Elliptic partial differential equations

The subject matter of this section is the application of CASM to elliptic partial differential equations.

To construct a difference scheme we use polynomials \( p \in \mathbb{P}^4 \) on a nonuniform grid. The convergence order for this finite difference scheme is \( k = 3 \) [6].

Consider the model problems
\[
\begin{align*}
\Delta u_i &= f_i, \quad \text{in } \Omega_i, \\
u_i &= 0, \quad \text{on } \partial \Omega_i,
\end{align*}
\]

where
\[
\begin{align*}
f_1 &= \left( - \frac{1}{r} - (1 - r)^2 \right) \sin(x_1x_2) - 4x_1x_2 \cos(x_1x_2) / r, \quad r = \sqrt{x_1^2 + x_2^2}, \\
f_2 &= -\left( 1 - \exp(1 - r^2) \right)^2 / \cos^2(x_1x_2) - 8 \exp(1 - r^2) \tan(x_1x_2) x_1x_2 \\
&\quad + 4 \exp(1 - r^2)(1 - r^2) \ln \left( \cos(x_1x_2) \right), \\
f_3 &= \left( -10/r + 100 \right) \exp(-10r), \\
\Omega_1 = \Omega_2 := \{ x_1 > 0, x_2 > 0, x_1^2 + x_2^2 < 1 \}.
\end{align*}
\]
The exact solutions to these problems for $i = 1, 2$ are:

$$u_1 = (1 - r) \sin(x_1 x_2),$$
$$u_2 = \left(1 - \exp(1 - r^2)\right) \ln \left(\cos(x_1 x_2)\right)$$

and for $i = 3$ is:

$$\Omega_3 := \{x_1^2 + x_2^2 < 1\},$$
$$u_3 = \exp(-10r) - \exp(-10), \quad r = \sqrt{x_1^2 + x_2^2}$$

Numerical solutions are obtained by using grids with the number of nodes from 16 to 256. The grid were optimized with respect to minimizing the defects.

<table>
<thead>
<tr>
<th>Number of problem</th>
<th>$| \cdot |_{L^2}$</th>
<th>$| \cdot |_{L^\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.40</td>
<td>3.41</td>
</tr>
<tr>
<td>2</td>
<td>4.95</td>
<td>4.21</td>
</tr>
<tr>
<td>3</td>
<td>5.67</td>
<td>4.28</td>
</tr>
</tbody>
</table>

Table 2. Convergence order

Since polynomials $p$ smooth numerical solution, we have the effect of superconvergence, and the convergence order of finite difference schemes for problems 1–3 is greater than the theoretical estimate.

References


Received: June 26, 1994
Revised version: August 15, 1995

Computer Center of Siberian Department of Russian Academy of Sciences
Akademgorodok
660036 Krasnoyarsk
Russia
E-mail: dobr@intra.krasnoyarsk.su