LUTZ ANGERMANN

Numerical solution of second-order elliptic equations on plane domains


<http://www.numdam.org/item?id=M2AN_1991__25_2_169_0>
NUMERICAL SOLUTION OF SECOND-ORDER ELLIPTIC EQUATIONS ON PLANE DOMAINS (*)

Lutz ANGERMANN (1)

Communicated by P. G. CIARLET

Abstract. — The paper presents a general discretization method for convective diffusion equations. The schemes are based on an integral formula and have the following advantages:
1. They are effective particularly in the case when convection is dominated;
2. Solutions obtained by them satisfy a discrete conservation law;
3. A discrete maximum principle is valid.

We show that the finite element solution converges to the exact one with the rate $O(h)$ in $H^1(D)$.

Résumé. — On propose une méthode générale de discrétisation pour la solution numérique des équations de diffusion avec convection. Les règles pour la discrétisation sont basées sur les formules d'intégration pour les champs et possèdent les avantages suivants:
1. Elles restent efficaces aussi dans le cas de la dominance du terme de convection;
2. pour les solutions obtenues on peut vérifier une loi de conservation discrète;
3. on obtient également un principe du maximum discret.

On démontre la convergence de la solution des éléments finis vers la solution exacte avec la vitesse $O(h)$ concernant la norme en $H^1$.

INTRODUCTION

Let us consider the following convective diffusion equation
\[ - \Delta u + b \cdot \nabla u + cu = f \quad (1) \]
in some polygonal domain $D$, where $b$ is a $\mathbb{R}^2$-valued function and $c, f$ are real-valued functions. Throughout this paper we assume the validity of the condition
\[ c(x) - \frac{1}{2} \nabla \cdot b(x) \geq a_0 > 0 \quad \text{on } D. \]

(*) Received December 1988, revised August 1989.
(1) Technische Universität Dresden, Abt. Mathematik, Mommsenstraße 13, Dresden, O-8027, Germany.
For the sake of simplicity we suppose Dirichlet boundary conditions to be fulfilled. We note, however, that this is not a principal restriction. The discretization method proposed below can be applied e.g. to mixed inhomogeneous boundary conditions, too.

The purpose of this paper is to present a conservative discretization scheme preserving coercivity and monotonicity properties of (1). Especially, the scheme is effective even in the case when the convection is dominated and it gives a nonnegative solution when the source term is nonnegative.

CONSTRUCTION OF THE DISCRETIZATION

Our scheme is based on the so-called control region method which has been theoretically investigated in the case of elliptic equations in the divergence form

\[-\nabla \cdot (k \nabla u) + cu = f,\]

see e.g. [3, 5, 6].

This type of methods is playing an important role in discretizing semiconductor device problems [10, 11, 12].

In the sequel, by \( n \) we shall denote the unit outer normal to the boundary of \( D \) (or a subset of \( D \)) with positive measure.

We first triangulate \( D \) in such a way that for some \( h_0 > 0 \) the usual regularity condition is satisfied.

The triangles \( T \) meet only in entire common sides or in vertices. Each triangle contains a circle of radius \( c_1 h \) and is contained in a circle of radius \( c_2 h \), where \( 0 < h = h_0 \) is the mesh parameter and the constants \( c_1, c_2 > 0 \) do not depend on \( T \) and \( h \).

It is wellknown that the condition last mentioned is equivalent to Zlámal's «minimum angle condition», i.e. it is assumed the existence of some number \( \alpha_0 > 0 \) such that all interior angles of the triangles are not less than \( \alpha_0 \).

Moreover we suppose this triangulation to satisfy the so-called inverse assumption (see [14]) and the following condition (A2) There exists a number \( \alpha_1 \) with \( 0 < \alpha_1 < \pi/2 - \alpha_0 \) such that each angle is either not greater than \( \pi/2 - \alpha_1 \) or equal to \( \pi/2 \).

The meaning of this condition will become clear later. Now we define the finite-dimensional subspace \( X^0_h \) of

\[Y^0 = \{ v \in H^1(D) \mid v = 0 \text{ on } \partial D \}\]

\[X^0_h = \{ v_h \in Y^0 \mid v_h \text{ is linear on each } T \}\]
By $v_h$ we shall denote the value of $v_h$ at the nodal point $x_i$. For the set of indices of the nodes lying in $D$ we shall use the notation $V$. Finally, let $V_t$ contain the indices of all nodal points adjacent to $x_t \in \overline{D}$.

Before formulating completely the discrete scheme we give a motivation of the proposed construction. We consider for all $i \in V$ the corresponding control regions (Dirichlet domains) $D_t = \{x \in D : \text{dist}(x, x_i) < \text{dist}(x, x_j)\}$ for all nodes $x_j \in \overline{D}, j \neq i$. In view of the relation

$$b \cdot \nabla u = \nabla \cdot (bu) - (\nabla \cdot b) u,$$

after integrating (1) over $D_t$ we obtain

$$- \int_{\partial D_t} n \cdot [\nabla u - bu] \, ds - \int_{D_t} [\nabla \cdot b - c] u \, dx = \int_{D_t} f \, dx.$$

With $B_{ij} = \partial D_t \cap \partial D_j$ it follows

$$- \sum_{j \in V_t} \int_{B_{ij}} n \cdot [\nabla u - bu] \, ds - \int_{D_t} [\nabla \cdot b - c] u \, dx = \int_{D_t} f \, dx. \quad (2)$$

At this place, the control region method applied to elliptic equations in the divergent form is continued with replacing the term $k(n \cdot \nabla u)$ to be integrated over $B_{ij}$ by a constant. In practice, this means that instead of $k(n \cdot \nabla u)$ the corresponding difference quotient $k(x_{ij}) (u_h - u_{ht})/d_{ij}$ is taken, where $x_{ij}$ is the midpoint of the straight-line segment between $x_i$ and $x_j$ and $d_{ij}$ denotes the length of this segment.

In our case we shall use the idea, too.

Therefore we replace the first integrand in (2) by constants, i.e.

$$- n \cdot [\nabla u - bu] \approx S_{ij}.$$

Then we can attempt to seek for a function $w^h$ being the solution of the corresponding differential equation on the straight-line segment between $x_i$ and $x_j$. With $x = yn, y \in \mathbb{R}^1$, we have

$$- \frac{d w^h}{dy} + (n \cdot b) w^h = S_{ij}.$$

As a rule, this equation cannot be solved analytically for arbitrary $b$. Since the approximation of the second term in (2) containing $\nabla \cdot b$ is usually performed by means of

$$\int_{D_t} (\nabla \cdot b) u \, dx \approx u(x_i) \int_{D_t} (\nabla \cdot b) \, dx = u(x_i) \int_{\partial D_t} (n \cdot b) \, ds$$

$$= u(x_i) \sum_{j \in V_t} \int_{B_{ij}} (n \cdot b) \, ds \approx u(x_i) \sum_{j \in V_t} N_{ij} m_{ij},$$

vol 25, n° 2, 1991
where $N_{ij}$ is some constant and $m_{ij}$ is the length of $B_{ij}$ (see [13]), it is reasonable to approximate the coefficient $n \cdot b$ in the same way, i.e. we set $n \cdot b \approx N_{ij}$. Then we look for a function $w^h$ satisfying the differential equation

$$-rac{d w^h}{d y} + N_{ij} w^h = S_{ij}.$$  

Obviously, the general solution of this equation is

$$w^h(y) = C_{ij} e^{N_{ij} y} + S_{ij}/N_{ij}.$$  

In this representation we have two unknowns: $C_{ij}$ and $S_{ij}$. Demanding $w^h(y(x_i)) = u_{hi}$ and $w^h(y(x_j)) = u_{hj}$, where $u_{hi}$ and $u_{hj}$ are the values of the final approximate solution $u_h$ at $x_i$ and $x_j$ respectively, we obtain a linear system of algebraic equations to determine $C_{ij}$ and $S_{ij}$:

$$\begin{cases}
  u_{hi} = C_{ij} e^{N_{ij} x_i} + S_{ij}/N_{ij} \\
  u_{hj} = C_{ij} e^{N_{ij} x_j} + S_{ij}/N_{ij}.
\end{cases}$$  

It is clear that for different $y_i$ and $y_j$ this system always admits a unique solution. Thus we get

$$C_{ij} = \frac{u_{hi} - u_{hj}}{e^{N_{ij} y_i} - e^{N_{ij} y_j}}, \quad S_{ij} = \frac{u_{hj} e^{N_{ij} y_i} - u_{hi} e^{N_{ij} y_j}}{e^{N_{ij} y_i} - e^{N_{ij} y_j}}.$$  

Now we can replace the integral over $B_{ij}$ by $S_{ij} m_{ij}$. In view of

$$\|x_j - x_i\| = \|n(y_j - y_i)\| = |y_j - y_i| = y_j - y_i$$

we get

$$S_{ij} m_{ij} = \frac{u_{hj} - u_{hi} e^{N_{ij} d_{ij}}}{1 - e^{N_{ij} d_{ij}}} N_{ij} m_{ij}.$$  

We note here that the procedure described above for obtaining a suitable approximation of the term

$$- \int_{\partial D_i} n \cdot (\nabla u - bu) \ ds$$

can be interpreted in various ways. One of these approaches is the application of the idea of exact schemes in connection with appropriate quadrature rules for the occurring integrals, see e.g. [10, 14].
The discretization of the remaining terms
\[ \int_{D_i} cu\,dx \quad \text{and} \quad \int_{D_i} f\,dx \]
will be performed via lumping-technique, i.e. we take
\[ \int_{D_i} cu\,dx = c_i u_i m_i \quad \text{and} \quad \int_{D_i} f\,dx = f_i m_i , \]
where \( m_i \) denotes the area of \( D_i \). Thus we obtain a system of discrete equations
\[ \sum_{j \in V_i} \left\{ 1 - \left[ 1 - r(N_{ij}, d_{ij}) \right] N_{ij} d_{ij} \right\} (u_{hi} - u_{hj}) \frac{m_{ij}}{d_{ij}} + \]
\[ + c_i u_{hi} m_i = f_i m_i , \quad i \in V , \]
with \( r(z) = 1 - \frac{1 - B(z)}{z} \) and \( B(z) = \frac{z}{e^z - 1} \).

Multiplying the \( i \)-th equation by an arbitrary number \( v_{hi} \) and adding all these expressions, this system can be rewritten as follows:

The unknown function \( u_h \in X_h^0 \) satisfies the variational equality
\[ a_i(u_h, v_h) = (f, v_h)_l \quad \text{for all} \quad v_h \in X_h^0 \]
where
\[ a_i(u_h, v_h) = b_i(u_h, v_h) + c_i(u_h, v_h) \]
with
\[ b_i(u_h, v_h) = \sum_{j \in V} \sum_{j \in V} \left\{ 1 - \left[ 1 - r(N_{ij}, d_{ij}) \right] N_{ij} d_{ij} \right\} (u_{hi} - u_{hj}) v_{hi} \frac{m_{ij}}{d_{ij}} \]
and
\[ c_i(u_h, v_h) = \sum_{i \in V} c_i u_{hi} v_{hi} m_i , \quad (f, v_h)_l = \sum_{i \in V} f_i v_{hi} m_i . \]

In the following we shall use the notation
\[ \| v_h \|_l = \sqrt{(v_h, v_h)_l} . \]

We observe that the function \( r \) has the following properties:

(P1) \( r(z) \) is monotone for all real \( z \),

(P2) \( \lim_{z \to -\infty} r(z) = 0 , \quad \lim_{z \to \infty} r(z) = 1 , \)
Moreover, the function $r$ is Lipschitz-continuous on the whole real axis. In view of some practical aspects which will be explained later we shall formulate this property in a weaker variant:

(P6) $r(z)z$ is Lipschitz-continuous on $\mathbb{R}$.

As a consequence of (P3) and (P4) we obtain the relation

(P7) $1 - [1 - r(z)]z \geq 0$ for all real $z$.

Indeed, from (P4) we get

$$1 - [1 - r(z)]z = 1 + r(-z)(-z) > 0.$$ 

Now we examine the properties of scheme (3).

BASIC PROPERTIES OF THE DISCRETIZATION

LEMMA 1: Let the condition (A1) be satisfied. If $b \in [W^2_\infty(D)]^2$, $c \in W^1_\infty(D)$ and if the approximations $N_{ij}$ are such that

$$|N_{ij}| \leq \|b; [L_\infty(B_{ij})]^2\| \quad \text{and} \quad N_{ij}m_{ij} - \int_{B_{ij}} (n \cdot b) \, ds = 0$$

for $b$ linear on $B_{ij}$, then for sufficiently small $h_0 > 0$ there exists a constant $K > 0$ such that for all $h \in (0, h_0]$ and $u_h \in X_h^0$ the relation

$$a_i(u_h, u_h) \leq \|u_h; H^1(D)\|^2 + K\|u_h\|^2_l$$

holds, where the constant $K$ is independent of $h$. ■

For future reference, we introduce the function

$$R(z) = 1 - \left[\frac{1}{2} - r(z)\right]z$$

and record the following important property of it.

COROLLARY 1: The function $R$ satisfies the relation

$$\max \left\{1, \frac{|z|}{2}\right\} \leq R(z) \leq 1 + \frac{|z|}{2}$$

for all real $z$. ■
Proof: The estimate $R(z) \geq 1$ follows immediately from (P5). The remaining relations will be proved for nonnegative $z$ only. This is sufficient in view of $R(z) = R(-z)$.

The left-hand side inequality results from (P7):

$$R(z) = 1 - [1 - r(z)] z + \frac{z}{2} \geq \frac{z}{2}.$$  

Further we have

$$R(z) = 1 - \frac{z}{2} + z r(z) \leq 1 - \frac{z}{2} + z = 1 + \frac{z}{2},$$

because (P1) and (P2) imply $r(z) \equiv 1$. $\blacksquare$

Proof of Lemma 1: At first we decompose $b_1(u_h, u_h)$ into two parts:

$$b_1(u_h, u_h) = b_1^{(1)}(u_h, u_h) + b_1^{(2)}(u_h, u_h)$$

with

$$b_1^{(1)}(u_h, u_h) = \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} [R(z_{ij}) u_{hi}$$

$$- \{1 - [1 - r(z_{ij})] z_{ij}\} u_{hj}] u_{hi} \frac{m_{ij}}{d_{ij}}$$

and

$$b_1^{(2)}(u_h, u_h) = -\frac{1}{2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} u_{hi}^2 N_{ij} m_{ij}, \text{ where } z_{ij} = N_{ij} d_{ij}.$$

Now, in order to treat the first component $b_1^{(1)}$ we use a symmetry argument.

Namely, changing the succession of summation and taking into consideration the boundary values of $u_h$, $b_1^{(1)}$ can be written in the following manner:

$$b_1^{(1)}(u_h, u_h) = \sum_{j \in \mathcal{V}} \sum_{i \in \mathcal{V}_j} [R(z_{ij}) u_{hi}$$

$$- \{1 - [1 - r(z_{ij})] z_{ij}\} u_{hj}] u_{hi} \frac{m_{ij}}{d_{ij}}.$$

Furthermore, replacing $i$ by $j$ and, vice versa, $j$ by $i$, we obtain

$$b_1^{(1)}(u_h, u_h) = \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} [R(z_{ji}) u_{hj}$$

$$- \{1 - [1 - r(z_{ji})] z_{ji}\} u_{hi}] u_{hj} \frac{m_{ji}}{d_{ji}}.$$
In view of the relations
\[ N_{ij} = -N_{ji}, \quad m_{ij} = m_{ji}, \quad d_{ij} = d_{ji}, \]
we have
\[
b_l^{(1)}(u_h, u_h) = \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} \left[ R(-z_{ij}) u_{hj} \right. \]
\[ - \left\{ 1 + \left[ 1 - r(-z_{ij}) \right] z_{ij} \right\} u_{hi} \frac{m_{ij}}{d_{ij}}. \]
Together with the original expression of \( b_l^{(1)} \) this leads to the representation
\[
b_l^{(1)}(u_h, u_h) = \frac{1}{2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} \left[ R(z_{ij}) u_{hi}^2 + u_{hj}^2 \right. \]
\[ - \left\{ 1 - \left[ 1 - r(z_{ij}) \right] z_{ij} \right\} u_{hi} \frac{m_{ij}}{d_{ij}}. \]
Since \( R \) is an even function, it follows
\[
b_l^{(1)}(u_h, u_h) = \frac{1}{2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} \left[ R(z_{ij}) u_{hi}^2 + u_{hj}^2 \right. \]
\[ - \left\{ 2 + \left[ r(z_{ij}) - r(-z_{ij}) \right] z_{ij} \right\} u_{hi} \frac{m_{ij}}{d_{ij}}. \]
Using (P4), we have
\[
b_l^{(1)}(u_h, u_h) = \frac{1}{2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} \left[ R(z_{ij}) (u_{hi}^2 + u_{hj}^2) \right. \]
\[ - \left\{ 2 + \left[ r(z_{ij}) - r(-z_{ij}) \right] z_{ij} \right\} u_{hi} \frac{m_{ij}}{d_{ij}} = \frac{1}{2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} R(z_{ij}) (u_{hi} - u_{hj})^2 \frac{m_{ij}}{d_{ij}}. \]
Hence, from Corollary 1 we get
\[
b_l^{(1)}(u_h, u_h) \equiv \frac{1}{2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}_i} (u_{hi} - u_{hj})^2 \frac{m_{ij}}{d_{ij}} = \left| u_h ; H^1(D) \right|^2, \]
where the last relation is a consequence of [8 : Lemma 2.3].
Finally we consider the remaining terms
\[
b_l^{(2)}(u_h, u_h) + c_l(u_h, u_h) = \sum_{k=1}^{3} \delta_{kh} \]
\[ M^2 \ AN \ Modélisation \ mathématique \ et \ Analyse \ numérique \]
\[ Mathematical \ Modelling \ and \ Numerical \ Analysis \]
with
\[ \delta_{1h} = - \frac{1}{8} \sum_{i \in \mathcal{V}} u_{hi}^2 \sum_{j \in \mathcal{V}_i} \left[ N_{ij} m_{ij} - \int_{B_{ij}} (n \cdot b) \, ds \right], \]
\[ \delta_{2h} = \sum_{i \in \mathcal{V}} \int_{D_i} (c_i - c) u_{hi}^2 \, dx, \]
\[ \delta_{3h} = \sum_{i \in \mathcal{V}} \int_{D_i} \left[ c - \frac{1}{2} \nabla \cdot b \right] u_{hi}^2 \, dx. \]

The first difference is treated as follows:
\[ 2 |\delta_{1h}| \leq \sum_{i \in \mathcal{V}} u_{hi}^2 \sum_{j \in \mathcal{V}_i} \left| \int_{B_{ij}} [\tilde{N}_{ij} - n \cdot b] \, ds \right|. \]

We denote by \( D_{ij}^+ \) the interior of the triangle having \( B_{ij} \) as one side and \( x_i \) as the opposite vertex. Then we can write
\[ \int_{B_{ij}} [\tilde{N}_{ij} - n \cdot b] \, ds = \text{meas} (B_{ij}) \int_{T_0^+} [\tilde{N}_{ij} - n \cdot \tilde{b}] \, d\tilde{s}, \]
where \( T_0^+ \) is the corresponding face of the reference triangle \( T^+ \). Now we can estimate
\[ \left| \int_{T_0^+} [\tilde{N}_{ij} - n \cdot \tilde{b}] \, d\tilde{s} \right| \leq 2 \left\| \tilde{b} \right\| \left[ L_\infty (T_0^+) \right]^2 \]
\[ \leq 2 \left\| \tilde{b} \right\| \left[ L_\infty (T^+) \right]^2 \leq 2 \left\| \tilde{b} \right\| \left[ W_\infty^2 (T^+) \right]^2. \]

Thus the integral is a linear continuous functional on \( [W_\infty^2 (T^+)]^2 \) vanishing for linear functions \( \tilde{b} \). By Bramble-Hilbert’s lemma, we obtain
\[ \left| \int_{T_0^+} [\tilde{N}_{ij} - n \cdot \tilde{b}] \, d\tilde{s} \right| \leq C \left\| \tilde{b} \right\| \left[ W_\infty^2 (T^+) \right]^2. \]

The back-transformation yields
\[ \left| \int_{T_0^+} [\tilde{N}_{ij} - n \cdot \tilde{b}] \, d\tilde{s} \right| \leq C \left\| \left( \frac{\partial F_{ij}^+}{\partial x} \right)^{-1} \right\|^{1/2} \left\| \tilde{b} \right\| \left[ W_\infty^2 (D_{ij}^+) \right]^2, \]
where \( F_{ij}^+ \) is the affin-linear map realizing the transformation \( D_{ij}^+ \rightarrow T^+ \).

From [4] we conclude that
\[ \left| \int_{B_{ij}} [\tilde{N}_{ij} - n \cdot \tilde{b}] \, ds \right| \leq C h^3 \left\| \tilde{b} \right\| \left[ W_\infty^2 (D_{ij}^+) \right]^2. \]
holds. Therefore we have
\[ |\delta_{1h}| \leq C h \| b \| \left[ W^2_\infty(D) \right]^2 \sum_{i \in V} u_{h_i}^2 h^2 \]
\[ \leq C h \| b \| \left[ W^2_\infty(D) \right]^2 \| u_h \|^2 \| I_i \| . \]

In a similar way one can derive the estimate
\[ |\delta_{2h}| \leq C h \| c \| W^1_\infty(D) \| u_h \|^2 \| I_i \| . \]

Summarizing the results and using assumption (A1) we obtain
\[ a_l(u_h, u_h) \geq |u_h ; H^1(D)|^2 + \{a_0 - C h \| b \| \left[ W^2_\infty(D) \right]^2 \} + |c ; W^1_\infty(D)| \| u_h \|^2 \| I_i \|. \]

Now it remains to choose \( h_0 \) such that for all \( h \in (0, h_0] \) the term in the braces is positive. \( \blacksquare \)

Remark 1: It is wellknown that the relation
\[ C \| u_h ; L^2(D) \| \leq \| u_h \| , \quad u_h \in X^0_h , \]
holds [8]. Therefore the functional \( a_l \) is \( X^0_h \)-coercive. \( \blacksquare \)

From assumption (A1) it follows that the «continuous» problem is inverse-isotone. Especially, if there is given a nonnegative right-hand side \( f \), then we obtain a nonnegative solution \( u \), too. The same fact is true for the discrete problem (3).

Before formulating and proving the corresponding result, we need some further notations. Let \( I_h : C(\bar{D}) \rightarrow X^0_h \) denote the usual interpolation operator. The symbol \( u_h^+ \) we shall understand as the positive part of the function \( u_h \), i.e. \( u_h^+ = \max \{0, u_h\} \). The negative part of \( u_h \) is defined by \( u_h^- = u_h^+ - u_h \).

**Lemma 2:** For all \( u_h \in X^0_h \) it holds
\[ a_l(I_h u_h^+, I_h u_h^-) \leq 0 . \]

**Proof:** It is easy to see that \( c_l(I_h u_h^+; I_h u_h^-) = 0 \). Now we consider the first term \( b_l \). We have
\[ b_l(I_h u_h^+, I_h u_h^-) = \sum_{i \in V} \sum_{j \in V} \{1 - [1 - r(N_{ij} d_{ij})] N_{ij} d_{ij}\} (u_{h_i}^+ - u_{h_i}^-) u_{h_i}^- \frac{m_{ij}}{d_{ij}} \]
\[ = - \sum_{i \in V} \sum_{j \in V} \{1 - [1 - r(N_{ij} d_{ij})] N_{ij} d_{ij}\} u_{h_i}^- u_{h_i}^- \frac{m_{ij}}{d_{ij}} . \]
Therefore, in view of the property (P7) of \( r \), we obtain
\[
b_i(I_h u^+, I_h u^-) \leq 0 .
\]

**COROLLARY 2:** Let the assumptions of Lemma 1 be fulfilled. Then the discrete problem (3) is inverse-isotone.

**Proof:** Since the problem is a linear one, it is sufficient to show that a nonnegative right-hand side \( f \in C(\overline{D}) \) leads to a nonnegative solution \( u_h \in X_h^0 \).

Let \( u_h \in X_h^0 \) be the solution of the problem (3) with a nonnegative right-hand side \( f \in C(\overline{D}) \). Taking the special trial function \( v_h = I_h u^- \), we obtain:
\[
0 \leq (f, I_h u^-)_I = a_i(u_h, I_h u^-) = a_i(I_h u^+, I_h u^-) - a_i(I_h u^-, I_h u^-) \leq -a_i(I_h u^-, I_h u^-).
\]

From Lemma 1 and Remark 1 respectively we conclude that
\[
\|I_h u^-; L_2(D)\|^2 \leq 0 ,
\]
i.e. \( u^-_h = 0 \).

**Remark 2:** It is wellknown that the inverse isotonocity is an important part of the sufficient conditions which imply \( L_\infty \)-stability.

**LEMMA 3:** The scheme (3) is conservative.

**Proof:** We have to derive an analogous discrete relation to the identity
\[
\int_{\partial D} [-n \cdot \nabla u + (n \cdot b) u] \, ds - \int_D [\nabla \cdot (b - c)] u \, dx = \int_D f \, dx . \tag{5}
\]

For this aim we consider (3) with the test function \( v_h = i_h \), where \( i_h \) is a function from \( X_h^0 \) with \( i_h(x_j) = 1 \) for all \( j \in V \). Then we can write
\[
b_i(u_h, i_h) = b_i^{(3)}(u_h, i_h) + b_i^{(4)}(u_h, i_h) + b_i^{(5)}(u_h, i_h)
\]
with
\[
b_i^{(3)}(u_h, i_h) = - \sum_{i \in V} \sum_{j \in V} (u_{hj} - u_{hi}) m_{ij}/d_{ij},
\]
\[
b_i^{(5)}(u_h, i_h) = - \sum_{i \in V} u_{hi} \sum_{j \in V} N_{ij} m_{ij}
\]
and
\[
b_i^{(4)}(u_h, i_h) = b_i(u_h, i_h) - b_i^{(3)}(u_h, i_h) - b_i^{(5)}(u_h, i_h) .
\]
At first we consider $b_l^{(3)}$. In contrast to the proof of Lemma 1, here we must be more attentive to the situation near the boundary. Denoting by $Z$ the index set of all nodal points lying on $\partial D$, we have

$$b_l^{(3)}(u_h, i_h) = - \sum_{j \in V, j \in V_j} \sum_{i \in V_j \cap V} (u_{b_j} - u_{h_i}) m_{ij} d_{ij} - \sum_{j \in V, j \in V_j \cap Z} (u_{b_j} - u_{h_i}) m_{ij} d_{ij}.$$  

It is not difficult to see that the first addend in this representation vanishes. The second sum can be rewritten as follows:

$$b_l^{(3)}(u_h, i_h) = - \sum_{j \in Z, j \in V_j} (u_{b_j} - u_{h_i}) m_{ij} d_{ij}.$$  

For $b_l^{(4)}$ we get

$$b_l^{(4)}(u_h, i_h) = \sum_{i \in V, j \in V_j \cap V} \{r(N_{ij} d_{ij}) u_{h_i} + [1 - r(N_{ij} d_{ij})] u_{b_j}\} N_{ij} m_{ij} + \sum_{i \in V, j \in V_j \cap Z} \{r(N_{ij} d_{ij}) u_{h_i} + [1 - r(N_{ij} d_{ij})] u_{b_j}\} N_{ij} m_{ij} = b_l^{(41)}(u_h, i_h) + b_l^{(42)}(u_h, i_h).$$  

Applying again a symmetry argument to $b_l^{(41)}$, as in the proof of Lemma 1, we observe that in view of (P4)

$$b_l^{(41)}(u_h, i_h) = \sum_{i \in V, j \in V_j \cap V} \{r(N_{ij} d_{ij}) u_{b_j} + [1 - r(N_{ij} d_{ij})] u_{h_i}\} N_{ij} m_{ij} = - \sum_{i \in V, j \in V_j \cap V} \{[1 - r(N_{ij} d_{ij})] u_{b_j} + r(N_{ij} d_{ij}) u_{h_i}\} N_{ij} m_{ij} = - b_l^{(41)}(u_h, i_h)$$  

holds, from which $b_l^{(41)} = 0$ follows.

Hence we get the relation

$$b_l^{(41)}(u_h, i_h) = - \sum_{j \in Z, i \in V_j} \{[1 - r(N_{ij} d_{ij})] u_{b_j} + r(N_{ij} d_{ij}) u_{h_i}\} N_{ij} m_{ij}.$$  

Taking into consideration the remaining terms in (3), we have

$$\sum_{j \in V, i \in V_j} \left\{-\frac{u_{b_j} - u_{h_i}}{d_{ij}} + N_{ij} u_{h_{b_j}}\right\} m_{ij} - \sum_{i \in V} \left\{\sum_{j \in V_j} N_{ij} m_{ij} - c_i m_i\right\} u_{h_i} = \sum_{i \in V} f_i m_i.$$  

M² AN Modélisation mathématique et Analyse numérique
Mathematical Modelling and Numerical Analysis
where \( u_{ij} = r(N_{ij} d_{ij}) u_{hi} + [1 - r(N_{ij} d_{ij})] u_{hj} \) can be regarded as an approximation of \( u_h \) on \( B_{ij} \), and this is the desired identity. ■

**CONVERGENCE IN THE \( H^1 \)-NORM**

**Theorem:** Suppose that the assumptions of Lemma 1 are fulfilled and that \( f \in W^1_q(D) \) with some \( q > 2 \). Furthermore, let the assumption \((A2)\) be satisfied and let the « continuous » solution \( u \) of (1) belong to \( Y^0 \cap H^2(D) \).

Then, for sufficiently small \( h_0 > 0 \) the estimate

\[
\| u - u_h ; H^1(D) \| \leq C h
\]

holds, where \( h \in (0, h_0] \) and \( C \) is a constant independent of \( h \). ■

**Proof:** From assumption \((A1)\) it follows that the bilinear form

\[
a(u, v) = \int_D \nabla u \cdot \nabla v \, dx + \int_D [b \cdot \nabla u + cu] v \, dx
\]

is \( H^1 \)-coercive. In particular, it holds

\[
C \| u_h - w_h ; H^1(D) \|^2 \leq a(u_h - w_h, u_h - w_h),
\]

(6)

where \( w_h \) is an arbitrary function from \( X^0_h \). The right-hand side of this inequality can be treated as follows:

\[
a(u_h - w_h, u_h - w_h)
\]

\[
= a(u - w_h, u_h - w_h) + a(u_h, u_h - w_h) - a(u_h, u_h - w_h)
\]

\[
= a(u - w_h, u_h - w_h) + a(u_h, u_h - w_h) - a_l(u_h, u_h - w_h)
\]

\[
+ (f, u_h - w_h)_l - (f, u_h - w_h).
\]

Now we have to consider the expressions

\[
\delta a = a(u_h, v_h) - a_l(u_h, v_h)
\]

and

\[
\delta f = (f, v_h) - (f, v_h)_l,
\]

where we set \( v_h = u_h - w_h \). Making use of the relation

\[
\int_D \nabla u_h \cdot \nabla v_h \, dx = \sum_{i \in \bar{V}} \sum_{j \in \bar{V}_i} (u_{hi} - u_{hj}) v_{hi} m_{ij}/d_{ij}
\]
(see [8 : Lemma 2.3]), the term $\delta a$ can be represented as follows:

$$\delta a = \int_D \nabla \cdot (b u_h) v_h \, dx - \sum_{i \in \mathbb{V}} \sum_{j \in \mathbb{V}_i} \left[ r_{ij} u_{hi} + (1 - r_{ij}) u_{hj} \right] N_{ij} v_{hi} m_{ij}$$

$$- \int_D [\nabla \cdot b - c] u_h v_h \, dx + \sum_{i \in \mathbb{V}} \left\{ \sum_{j \in \mathbb{V}_i} N_{ij} m_{ij} - c, m_i \right\} u_{hi} v_{hi},$$

where $r_{ij} = r(N_{ij} d_{ij}).$

At first we consider the term

$$\delta b_1 = \int_D \nabla \cdot (b u_h) v_h \, dx - \sum_{i \in \mathbb{V}} \sum_{j \in \mathbb{V}_i} \left[ r_{ij} u_{hi} + (1 - r_{ij}) u_{hj} \right] N_{ij} v_{hi} m_{ij}$$

and decompose it additively into the following parts:

$$\delta b_{11} = \sum_{i \in \mathbb{V}} \sum_{j \in \mathbb{V}_i} \int_{B_{ij}} \left\{ u_h - r_{ij} u_{hi} - [1 - r_{ij}] u_{hj} \right\} N_{ij} v_{hi} \, ds,$$

$$\delta b_{12} = \sum_{i \in \mathbb{V}} \sum_{j \in \mathbb{V}_i} \int_{B_{ij}} (n \cdot b - N_{ij}) u_h v_{hi} \, ds,$$

$$\delta b_{13} = \sum_{i \in \mathbb{V}} \int_{D_i} \nabla \cdot (b u_h) (v_h - v_{hi}) \, dx.$$

Now we start the detailed estimations. For $\delta b_{11}$ we have, by a symmetry argument,

$$\delta b_{11} = \frac{1}{2} \sum_{i \in \mathbb{V}} \sum_{j \in \mathbb{V}_i} \int_{B_{ij}} \left\{ [u_h - r_{ij} u_{hi} - (1 - r_{ij}) u_{hj}] N_{ij} v_{hi} \right.$$  

$$+ [u_h - r_{ji} u_{hj} - (1 - r_{ji}) u_{hi}] N_{ji} v_{hj} \} \, ds$$

$$= \frac{1}{2} \sum_{i \in \mathbb{V}} \sum_{j \in \mathbb{V}_i} \int_{B_{ij}} \left\{ [u_h - r_{ij} u_{hi} - (1 - r_{ij}) u_{hj}] N_{ij} (v_h - v_{hj}) \right.$$  

$$+ (1 - r_{ij} - r_{ji}) N_{ij} (u_{hi} - u_{hj}) v_{hj} \} \, ds.$$

Since $(1 - r_{ij} - r_{ji}) N_{ij} d_{ij} = 0$, in view of (P4), we obtain

$$\delta b_{11} = \frac{1}{2} \sum_{i \in \mathbb{V}} \sum_{j \in \mathbb{V}_i} \int_{B_{ij}} [u_h - r_{ij} u_{hi} - (1 - r_{ij}) u_{hj}] \times$$

$$\times N_{ij} (v_h - v_{hj}) \, ds. \quad (7)$$
For the sake of simplicity in the notation we shall investigate the integral

\[ \int_{B_{ij}} u_{hij} v_{hij} \, ds, \]

where

\[ u_{hij} = u_h - r_{ij} u_{hi} - (1 - r_{ij}) u_{hj} \]

and

\[ v_{hij} = N_{ij} (v_{hi} - v_{hj}). \]

Note that \( v_{hij} \) can also be written as

\[ v_{hij} = N_{ij} d_{ij} (n \cdot \nabla v_h)|_{D_{ij}}. \]

Obviously, the point \( x_i \) and the segment \( B_{ij} \) define some triangle \( \tilde{D}_{ij}^+ \) in \( \tilde{D}_i \) in such a way that \( B_{ij} \) is one side and \( x_i \) is the corresponding vertex (see Lemma 1). A second triangle lying in \( \tilde{D}_i \) is defined by \( x_j \) and \( B_{ij} \). We denote the interior of the union of these two triangles by \( D_{ij} \).

It is clear that \( \tilde{D}_{ij} \) can be represented in another manner, namely as the union of two triangles \( \tilde{D}_{ij}^{(1)} \) and \( \tilde{D}_{ij}^{(2)} \) having the common vertices \( x_i \) and \( x_j \).

Further we set \( B_{ij}^{(k)} = B_{ij} \cap \tilde{D}_{ij}^{(k)}, \quad k = 1; 2. \) Let us mention that from assumption (A2) the regularity of the resulting triangulation follows. For each \( \tilde{D}_{ij}^{(k)} \) we are able to find an affin-linear mapping \( \tilde{F}_{ij}^{(k)} \) which transforms the triangle \( \tilde{D}_{ij}^{(k)} \) onto a reference element \( T \):

\[ \tilde{F}_{ij}^{(k)} : \quad \tilde{D}_{ij}^{(k)} \to T. \]

By \( \tilde{G}_{ij}^{(k)} \) we denote the restriction of \( \tilde{F}_{ij}^{(k)} \) to \( B_{ij}^{(k)} \), i.e. \( \tilde{G}_{ij}^{(k)} : \tilde{B}_{ij}^{(k)} \to T_0 \).

Now we return to the estimation of the integral over \( B_{ij} \). We have

\[ \int_{B_{ij}} u_{hij} v_{hij} \, ds = \sum_{k=1,2} \int_{B_{ij}^{(k)}} u_{hij} v_{hij} \, ds = \sum_{k=1,2} \text{meas}(B_{ij}^{(k)}) \int_{T_0} \tilde{u}_{hij}^{(k)} \tilde{v}_{hij}^{(k)} \, d\tilde{s}. \]

The desired estimate follows from the study of the functional

\[ J(\tilde{u}_h^{(k)}) = \int_{T_0} \tilde{u}_{hij}^{(k)} \tilde{v}_{hij}^{(k)} \, d\tilde{s}, \]

where we assume temporarily that \( v_h \) is fixed. Obviously, \( \tilde{u}_{hij}^{(k)} \) can be

vol 25, n° 2, 1991
supposed to belong to $H^2(T)$ since it is affine-linear on $T$. Moreover, $\tilde{v}_{hij}^{(k)}$ is constant on $T$. Therefore it holds

$$|J(\tilde{u})| \leq 2 \| \tilde{u}; L_\infty(T) \| \int_{T_0} \tilde{v}_{hij}^{(k)} \ d\mathcal{S} = 2 \frac{\text{meas}(T_0)}{\text{meas}(T)} \| \tilde{u}; L_\infty(T) \| \| \tilde{v}_{hij}^{(k)}; L_1(T) \| .$$

From Sobolev’s imbedding theorem it follows

$$|J(\tilde{u})| \leq C \| \tilde{u}; H^2(T) \| \| \tilde{v}_{hij}^{(k)}; L_1(T) \| .$$

Hence $J$ is a linear continuous functional on $H^2(T)$ with the norm

$$\|J; (H^2(T))^*\| \leq C \| \tilde{v}_{hij}^{(k)}; L_1(T) \| .$$

In addition, we observe that $J(\tilde{u})$ vanishes for all functions $\tilde{u}$ being constant on $T$. By a modification of the well-known lemma by Bramble & Hilbert [6] we obtain

$$|J(\tilde{u})| \leq C \| \tilde{v}_{hij}^{(k)}; L_1(T) \| \left\{ \| \tilde{u}; H^1(T) \|^2 + |\tilde{u}; H^2(T)|^2 \right\}^{1/2} .$$

In view of $|\tilde{u}; H^2(T)| = 0$ it follows

$$|J(\tilde{u})| \leq C \| \tilde{v}_{hij}^{(k)}; L_1(T) \| |\tilde{u}|_{hij}^{(k)}; H^1(T) | .$$

Therefore we have

$$\left| \int_{D_{ij}^{(k)}} u_{hij} v_{hij} \ d\mathcal{S} \right| \leq C \| (\mathcal{D} F_{ij}^{(k)})^{-1} \| \| \det \mathcal{D} F_{ij}^{(k)} \| \left\{ \| u_{hij}; H^1(D_{ij}^{(k)}) \| + \| v_{hij}; L_1(D_{ij}^{(k)}) \| \right\}$$

with $\mathcal{D} F_{ij}^{(k)} = \left( \frac{\partial F_{ij}^{(k)}}{\partial x} \right)$, where we used the transformation theorem from [4]. The same book gives the estimate

$$\| (\mathcal{D} F_{ij}^{(k)})^{-1} \| \leq Ch_{ij}^{(k)} ,$$

where $h_{ij}^{(k)}$ denotes the diameter of $D_{ij}^{(k)}$ and $C > 0$ is a constant independent of $D_{ij}^{(k)}$. Furthermore it is known that under assumption (A2) the relation

$$|\det \mathcal{D} F_{ij}^{(k)}| \leq C (h_{ij}^{(k)})^{-2}$$
with a independing of \( D^{(k)}_ij \) constant \( C > 0 \) holds. Hence we obtain
\[
\left| \int_{B_{ij}} u_{hij} v_{hij} \, ds \right| \leq C (h^{(k)}_{ij})^{-2} |u_h; H^1(D^{(k)}_ij)| \| v_{hij}; L^1(D^{(k)}_ij) \|.
\]

Now we derive an estimate for the \( L^1 \)-norm of \( v_{hij} \) on \( D^{(k)}_ij \):
\[
\| v_{hij}; L^1(D^{(k)}_ij) \| \leq d_{ij} \| N_{ij} \| \| n \cdot \nabla v_h; L^2(D^{(k)}_ij) \| \| 1; L^2(D^{(k)}_ij) \|
\leq C (h^{(k)}_{ij})^2 \| N_{ij} \| \| v_h; H^1(D^{(k)}_ij) \|.
\]

Thus we have
\[
\left| \int_{B_{ij}} u_{hij} v_{hij} \, ds \right| \leq C \sum_{k = 1, 2} \text{meas} (B^{(k)}_{ij}) \| u_h; H^1(D^{(k)}_ij) \| \| v_h; H^1(D^{(k)}_ij) \|
\leq Ch \| b; [L^\infty(D)]^2 \| \sum_{k = 1, 2} \| u_h; H^1(D^{(k)}_ij) \| \| v_h; H^1(D^{(k)}_ij) \|.
\]

Applying Cauchy’s inequality, we get finally
\[
\left| \int_{B_{ij}} u_{hij} v_{hij} \, ds \right| \leq Ch \| b; [L^\infty(D)]^2 \| \| u_h; H^1(D_{ij}) \| \| v_h; H^1(D_{ij}) \|.
\]

The same inequality applied to \( |\delta b_{11}| \) with \( \delta b_{11} \) from (7) leads to
\[
|\delta b_{11}| \leq Ch \| b; [L^\infty(D)]^2 \| \| u_h; H^1(D) \| \| v_h; H^1(D) \|.
\]

The next step consists in estimating \( \delta b_{12} \). We have
\[
|\delta b_{12}| \leq \frac{1}{2} \sum_{i \in \nu \cup \zeta, j \in \nu, i} \int_{B_{ij}} \left| (n \cdot b - N_{ij}) u_h (v_{hi} - v_{hij}) \right| \, ds
\]
\[
\leq \frac{1}{2} \sum_{i \in \nu \cup \zeta, j \in \nu} \| u_h; L^\infty(B_{ij}) \| \| v_{hi} - v_{hij} \| \int_{B_{ij}} |n \cdot b - N_{ij}| \, ds.
\]

In the proof of Lemma 1 (see the estimate of \( \delta^{(1)}_h \)) we obtained the relation
\[
\int_{B_{ij}} |n \cdot b - N_{ij}| \, ds \leq Ch^3 \| b; [W^2_\infty(D_{ij})]^2 \|,
\]
and from this we conclude that
\[
|\delta b_{12}| \leq Ch^2 \| b; [W^2_\infty(D)]^2 \| \| u_h; L^\infty(D) \| \sum_{i \in \nu \cup \zeta, j \in \nu} \| v_{hi} - v_{hij} \| h
\]
holds. In view of
\[ |v_{hi} - v_{hj}| = d_{ij} |n \cdot \nabla v_h| \bigg|_{D_i}, \]
and
\[ d_{ij} h \leq C \text{ meas } (D_{ij}) \]
it follows
\[ |\delta b_{12}| \leq Ch^2 |b| \cdot |W_2^2(D)|^2 \cdot \|u_h; L_\infty(D)\| \cdot \|v_h; H^1(D)\|. \]

The norm-equivalence theorem (see [4]) implies
\[ \|u_h; L_\infty(D)\| \leq Ch^{-1} \|u_h; L_2(D)\|, \]
i.e. we get
\[ |\delta b_{12}| \leq Ch \cdot |W_2^2(D)|^2 \cdot \|u_h; L_2(D)\| \cdot \|v_h; H^1(D)\|. \]
The last term \( \delta b_{13} \) can be estimated as follows:
\[ |\delta b_{13}| \leq \sum_{i \in V} \|\nabla \cdot (bu_h); L_2(D_i)\| \cdot \|v_h - v_{hi}; L_2(D_i)\| \]
\[ \leq \|\nabla \cdot (bu_h); L_2(D)\| \cdot \left\{ \sum_{i \in V} \|v_h - v_{hi}; L_2(D_i)\|^2 \right\}^{1/2}, \]
where we applied Cauchy’s inequality. The term in the braces satisfies the relation
\[ \left\{ \sum_{i \in V} \|v_h - v_{hi}; L_2(D_i)\|^2 \right\}^{1/2} \leq Ch \cdot \|v_h; H^1(D)\|, \]
see [8]. Thus we obtain
\[ |\delta b_{13}| \leq Ch \cdot \|\nabla \cdot (bu_h); L_2(D)\| \cdot \|v_h; H^1(D)\|, \]
and therefore it holds
\[ |\delta b_1| \leq Ch \cdot \|v_h; H^1(D)\|. \quad (8) \]
The next step consists in estimating the expression
\[ \delta b_2 = \int_D (\nabla \cdot b - c) u_h v_h \, dx - \sum_{i \in V} \left\{ \sum_{j \in V_i} N_{ij} m_{ij} - c_i m_i \right\} u_{hi} v_{hi} \]
\[ = \int_D (\nabla \cdot b) u_h v_h \, dx - \sum_{i \in V} \sum_{j \in V_i} N_{ij} m_{ij} u_{hi} v_{hi} \]
\[ - \left\{ \int_D c u_h v_h \, dx - \sum_{i \in V} c_i u_{hi} v_{hi} m_i \right\} = \delta b_{21} - \delta b_{22}. \]
The term $\delta b_{21}$ can be handled as follows:

$$\delta b_{21} = \sum_{i \in V} \int_{D_i} (\nabla \cdot \mathbf{b}) u_h(v_h - v_{h}) \, dx$$
$$+ \sum_{i \in V} \int_{D_i} (\nabla \cdot \mathbf{b}) (u_h - u_{h}) v_{h} \, dx$$
$$+ \sum_{i \in V} u_{h} v_{h} \sum_{j \in V_i} \left[ \int_{\partial D_i} (n \cdot \mathbf{b}) \, ds - N_{ij} m_{ij} \right]$$
$$= \delta b_{211} + \delta b_{212} + \delta b_{213}.$$  

Now we have (see, e.g. [8])

$$|\delta b_{211}| \leq C h |b| |W^1_\infty(D)|^2 |u_h|_{L^2(D)} |v_h|_{H^1(D)}$$

and

$$|\delta b_{212}| \leq C h |b| |W^2_\infty(D)|^2 |u_h|_{H^1(D)} |v_h|_{L^2(D)} \leq C h |b| |W^1_\infty(D)|^2 |u_h|_{H^1(D)} |v_h|_{L^2(D)}.$$  

Using the same argument as in the proof of Lemma 1, where we estimated $\delta_{1h}$, we obtain

$$|\delta b_{213}| \leq C h |b| |W^2_\infty(D)|^2 |u_h|_{L^2(D)} |v_h|_{L^2(D)}.$$  

Finally we need an estimate for $\delta b_{22}$. This is not difficult to achieve:

$$|\delta b_{22}| = \delta b_{221} + \delta b_{222} + \delta b_{223},$$

where

$$|\delta b_{221}| = \left| \sum_{i \in V} \int_{D_i} cu_h(v_h - v_{h}) \, dx \right|$$
$$\leq C h \|c\|_{L^\infty(D)} |u_h|_{L^2(D)} |v_h|_{H^1(D)}$$

and

$$|\delta b_{222}| = \left| \sum_{i \in V} \int_{D_i} c(u_h - u_{h}) v_{h} \, dx \right|$$
$$\leq C h \|c\|_{L^\infty(D)} |u_h|_{H^1(D)} |v_h|_{L^2(D)}.$$  

For

$$\delta b_{223} = \sum_{i \in V} \int_{D_i} (c - c_i) u_{h} v_{h} \, dx$$
we take the corresponding result for $\delta_{2}h$ from the proof of Lemma 1:

$$|\delta b_{223}| \leq Ch |c; W_{\infty}^{1}(D)| \|u_{h}; L_{2}(D)\| \|v_{h}; L_{2}(D)\|.$$ 

Thus we get

$$|\delta b_{2}| \leq Ch \|v_{h}; H^{1}(D)\|. \quad (9)$$

Now it remains to consider the right-hand side difference $\delta f$:

$$\delta f = \int_{D} f v_{h} \, dx - \sum_{i \in V} f_{i} v_{hi} m_{i}$$

$$= \sum_{i \in V} \left\{ \int_{D_{i}} (f - f_{i}) v_{h} \, dx + \int_{D_{i}} f_{i} (v_{h} - v_{hi}) \, dx \right\}.$$ 

The first sum admits the estimation (see [8] again)

$$\left| \sum_{i \in V} \int_{D_{i}} (f - f_{i}) v_{h} \, dx \right| \leq Ch \|f; W_{q}^{1}(D)\| \|v_{h}; L_{2}(D)\|;$$

for the second one we have

$$\left| \sum_{i \in V} \int_{D_{i}} f_{i} (v_{h} - v_{hi}) \, dx \right| \leq Ch \|f; W_{q}^{1}(D)\| \|v_{h}; H^{1}(D)\|.$$ 

Therefore it holds

$$|\delta f| \leq Ch \|f; W_{q}^{1}(D)\| \|v_{h}; H^{1}(D)\|. \quad (10)$$

Summarizing the estimates (8)-(10) and using the boundedness of $a$, we obtain from (6)

$$\|u - u_{h}; H^{1}(D)\| \leq \|u - w_{h}; H^{1}(D)\| + \|u_{h} - w_{h}; H^{1}(D)\|.$$ 

Since we may assume $u_{h} - w_{h} \neq 0$, it follows

$$\|u_{h} - w_{h}; H^{1}(D)\| \leq C [\|u - w_{h}; H^{1}(D)\| + h].$$

The triangle inequality

$$\|u - u_{h}; H^{1}(D)\| \leq \|u - w_{h}; H^{1}(D)\| + \|u_{h} - w_{h}; H^{1}(D)\|$$

yields

$$\|u_{h} - u_{h}; H^{1}(D)\| \leq C [\|u - w_{h}; H^{1}(D)\| + h].$$
If we take \( w_h = I_h u \), then it is wellknown that the relation
\[
\| u - w_h ; H^1(D) \| \leq C h \| u ; H^2(D) \|
\]
holds. Thus we get
\[
\| u - u_h ; H^1(D) \| \leq C h .
\]

**CONCLUSIVE REMARKS**

In the formulation of our theorem we did not attend to the qualitative
dependence of the constant \( C \) on the norms \( |b ; W^2_\infty(D)| \), \( |c ; W^1_\infty(D)| \) and \( |f ; W^1_\infty(D)| \) respectively.

However, the interested reader can understand it without serious
difficulties from the detailed proof.

In the theory of singular perturbations of elliptic differential equations
often there are considered problems of the type
\[
- \epsilon \Delta u + b \cdot \nabla u + cu = f
\]
with \( 0 < \epsilon \ll 1 \) and \( \| b \|^2 + c^2 + f^2 = o(1) \) for \( \epsilon \to 0 \). The corresponding
discrete bilinear form \( a^\epsilon_I \) can be constructed in the described way, i.e. we set
\[
a^\epsilon_I(u_h, v_h) = b^\epsilon_I(u_h, v_h) + c_I(u_h, v_h),
\]
where
\[
b^\epsilon_I(u_h, v_h) = \sum_{i \in V} \sum_{j \in V_i} \{ \epsilon - [1 - r(N_{ij} d_{ij}/\epsilon)] N_{ij} d_{ij} \} (u_{hi} - u_{hj}) v_{hi} \frac{m_{ij}}{d_{ij}}.
\]

Then it is convenient to investigate the coercivness of \( a^\epsilon_I \) with respect to
the so-called « \( \epsilon \)-weighted » norm defined by
\[
\| u \|^2 = \epsilon \| u ; H^1(D) \|^2 + \| u ; L^2_2(D) \|^2 , \quad u \in H^1(D).
\]

Requiring the assumption (A.1) to be fulfilled and analyzing the proof of
Lemma 1 we conclude that \( a^\epsilon_I \) is coercive and, what is significant, that the
bound \( h_0 \) does not depend on the parameter \( \epsilon \).

All assertions concerning the scheme (3) are proved under the assumption
that the control function \( r(x) \) satisfies the properties (P1)-(P5). The
property (P6) was added for the sake of completeness only.

Now one can ask, of course, whether or not there exist further and, may
be, more simple functions \( r \) satisfying the properties mentioned above. Here
we give some examples of such functions :

vol. 25, n° 2, 1991
\[ r(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0, \end{cases} \quad [2, 13, 14] \]

\[ r(z) = \frac{1}{2} (1 + \tanh z), \quad [9] \]

\[ r(z) = \begin{cases} 0, & z < -m \\ \frac{z + m}{2m}, & |z| \leq m, \quad (0 < m \leq 8) \quad [1] \\ 1, & z > m \\ 0, & z < -m \\ 0.5, & |z| \leq m, \quad (0 \leq m \leq 2) \quad [1] \\ 1, & z > m \end{cases} \]

Heinrich [7] takes the function

\[ r(z) = \begin{cases} (1 + \sigma)/2, & z \geq 0 \\ (1 - \sigma)/2, & z < 0, \end{cases} \]

where \( \sigma = \sigma(z) \) is chosen such that the relation

\[ \sigma(z) \geq \sigma_\delta(z) = \max \left\{ 0, 1 - \frac{2}{(1 + \delta) |z|} \right\} \]

holds for some small \( \delta > 0 \).

This function satisfies the properties (P1)-(P7), too.

Let us add that in our situation it is legitimate to permit \( \delta = 0 \). Then the resulting scheme corresponds to the scheme \( E \) (« partial upwind ») proposed in [8].

At this place we notice that the first and the last of these examples, properly speaking, are the reason for the necessity to formulate the property (P6) of \( r \) exactly in the variant given above.

Finally we want to accentuate that the discretization method and the results can be extended easily to the three-dimensional case. An application of the method for two- and three-dimensional domains \( D \) to the spatial discretization of the fundamental equations for the carrier transport in semiconductors was given in [1].

REFERENCES


M² AN Modélisation mathématique et Analyse numérique
Mathematical Modelling and Numerical Analysis


