Jiří Vacek

Dual variational principles for an elliptic partial differential equation

Aplikace matematiky, Vol. 21 (1976), No. 1, 5–27

Persistent URL: http://dml.cz/dmlcz/103619

DUAL VARIATIONAL PRINCIPLES FOR AN ELLIPTIC PARTIAL DIFFERENTIAL EQUATION

JIŘÍ VACEK

(Received August 14, 1974)

Introduction

In this paper we formulate the dual variational principles for the equation

$$-\sum_{i,j=1}^{N} \frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial u}{\partial x_j}\right) + au = f$$

in the domain $\Omega \subset E_N$.

We prove that the variational principles make it possible to obtain a posteriori error estimates and we show how to construct convergent approximations to the exact solutions of variational problems. In the last section some numerical results are presented.

1. Formulation of the problem

Let E_N be the N-dimensional Euclidean space, $\Omega \subset E_N$ an open Lipschitz region with the boundary Γ and let

$$\Gamma = \Gamma_u \cup \Gamma_v \cup \Gamma_h \cup \Gamma_0$$

where Γ_u, Γ_v, Γ_h are mutually disjoint sets open in Γ and $\text{mes}_{n-1}\,\Gamma_0 = 0$; by mes_N we denote the N-dimensional measure.

Let Ω be divided into mutually disjoint open Lipschitz subregions Ω_s, $s = 1, 2, \ldots, m$, i.e., let

$$\Omega = \Omega^0 \cup \Gamma_i$$

where

$$\Omega^0 = \bigcup_{s=1}^{m} \Omega_s$$

and

$$\Gamma_i = \bigcup_{\substack{r,s=1 \\ r \neq s}}^{m} (\bar{\Omega}_r \cap \bar{\Omega}_s)$$

is such that $\text{mes}_N \, \Gamma_i = 0$ and $\text{mes}_{N-1} \, \Gamma \cap \Gamma_i = 0$.

Let us introduce the following real functions:

1) $a_{ij}(x)$, $a(x)$ are bounded piecewise continuous functions in Ω with jumps on Γ_i and such that for $x \in \Omega^0$ they satisfy

(1) $$a_{ij}(x) = a_{ji}(x) \,, \quad i, j = 1, \ldots, N \,,$$

(2) $$a_{ij}(x) \, \xi_i \xi_j \geqq c|\xi|^2 \,, \quad c > 0 \,, \quad \text{for all} \quad \xi \in E_N \,,$$

(3) $$a(x) \geqq a_0 > 0 \,.$$

2) $\alpha(x)$ is a bounded piecewise continuous function on Γ_v with jumps on $\Gamma_v \cap \Gamma_i$ such that for $x \in \Gamma_v - \Gamma_i$

(4) $$\alpha(x) \geqq \alpha_0 > 0 \,.$$

(5) 3) $f \in L^2(\Omega) \,, \quad u_0 \in W^{1,2}(\Omega) \,, \quad g \in L^2(\Gamma) \,.$

We use the following notation: $L^2(\Omega)$, $L^2(\Gamma)$ are the spaces of square integrable functions in Ω, Γ. The usual norms in these spaces are denoted by $\| \, \|_0$, $\| \, \|_{0,\Gamma}$.

$W^{k,2}(\Omega)$ is the Sobolev space of functions with generalized derivatives up to the k-th order belonging to $L^2(\Omega)$. The norm in this space is denoted by $\| \, \|_k$.

We use the convention that a repeated subscript indicates the summation over the range of the space E_N.

The letter c with a possible subscript will denote a positive constant and $u,_i$ will be used instead of $\partial u / \partial x_i$.

If we define the differential operator A by the relation

(6) $$Au = au - (a_{ij} u,_j),_i$$

then (2) says that A is uniformly elliptic in Ω.

We define in $W^{1,2}(\Omega) \times W^{1,2}(\Omega)$ a bilinear form

(7) $$A(v, u) = \int_\Omega (a_{ij} v,_j u,_j + avu) \, \mathrm{d}x$$

The boundedness of a_{ij} and a in Ω implies that $A(v, u)$ is a continuous bilinear form in $W^{1,2}(\Omega) \times W^{1,2}(\Omega)$.

For functions belonging to $W^{1,2}(\Omega)$ the boundary values can be well defined, as is seen from the following theorem:

Theorem 1. *Let Ω be a bounded Lipschitz region with the boundary Γ. Then there exists a uniquely defined linear mapping*

$$T: W^{1,2}(\Omega) \to L^2(\Gamma)$$

which is continuous and such that $Tu = u|_\Gamma$ for $u \in \mathscr{E}(\bar{\Omega})$

Definition. T is called *the operator of traces*, Tu is called *the trace* on Γ of the function $u \in W^{1,2}(\Omega)$.

Here $\mathscr{E}(\bar{\Omega})$ is the set of functions which are infinitely differentiable in $\bar{\Omega}$ and $u|_\Gamma$ denotes the restriction of u to Γ. For the proof of the theorem see [7].

The boundary values of functions belonging to $W^{1,2}(\Omega)$ will be understood in the sense of traces and we shall often write u instead of Tu for the values on the boundary.

We define on $L^2(\Gamma_v) \times L^2(\Gamma_v)$ the bilinear form

$$a(v, u) = \int_{\Gamma_v} \alpha v u \,\mathrm{d}\Gamma . \tag{8}$$

From the boundedness of α on Γ_v and from the continuity of T it follows that $a(v, u)$ is continuous in $W^{1,2}(\Omega) \times W^{1,2}(\Omega)$.

Now we can define in $W^{1,2}(\Omega) \times W^{1,2}(\Omega)$ the bilinear form

$$((v, u)) = A(v, u) + a(v, u) . \tag{9}$$

This form is continuous and thus there exists $c > 0$ such that

$$u, v \in W^{1,2}(\Omega) \Rightarrow ((v, u)) \leqq c\|u\|_1 \|v\|_1 . \tag{10}$$

Definition. Let V be the closure of the set

$$\mathscr{V} = \{v \in \mathscr{E}(\bar{\Omega}) : \operatorname{supp} v \cap \Gamma_u = \emptyset\} \tag{11}$$

in $W^{1,2}(\Omega)$ and let G be a linear functional defined in V by

$$Gv = \int_\Omega vf \,\mathrm{d}x + \int_{\Gamma_v \cup \Gamma_h} vg \,\mathrm{d}\Gamma \tag{12}$$

A function $u \in W^{1,2}(\Omega)$ satisfying

$$u - u_0 \in V \tag{13a}$$

$$v \in V \Rightarrow ((v, u)) = Gv \tag{13b}$$

is called *a weak solution* or simply *a solution of the problem* (13) *with the stable boundary condition u_0*.

If we suppose that the functions a_{ij}, a, α, f, g, u_0, the boundary Γ and the solution u are sufficiently smooth, we can apply Green's theorem obtaining the classical interpretation of our problem: u is the solution of the problem

$$
\begin{aligned}
(14) \qquad -(a_{ij}u,_j),_i + au &= f \quad \text{in} \quad \Omega\,,\\
a_{ij}u,_j n_i + \alpha u &= g \quad \text{on} \quad \Gamma_v\,,\\
a_{ij}u,_j n_i &= g \quad \text{on} \quad \Gamma_h\,,\\
u &= u_0 \quad \text{on} \quad \Gamma_u\,,\\
u \text{ and } a_{ij}u,_j n_i & \text{ are continuous on } \Gamma_i\,,
\end{aligned}
$$

where $n = (n_1, \ldots, n_N)$ is the unit outward normal.

2. Existence of the weak solution

In this section we shall prove that the problem (13) has a unique solution. We recall here some well-known facts about the V-ellipticity from [7].

Definition. The bilinear form $((v, u))$ is called *V-elliptic* if there exists $c > 0$ such that

$$(15) \qquad v \in V \Rightarrow ((v, v)) \geqq c\|v\|_1^2\,.$$

Theorem 2. *The boundary value problem with a V-elliptic associated bilinear form $((v, u))$ has a unique solution u.*
It holds

$$(16) \qquad \|u\|_1 \leqq c(\|f\|_0 + \|u_0\|_1 + \|g\|_{0,\Gamma})\,.$$

To prove the existence and uniqueness of the solution of (13), it is now sufficient to prove the V-ellipticity of the form (9). If we use (2), (3), (4) it is easy to see that

$$(17) \qquad v \in W^{1,2}(\Omega) \Rightarrow ((v, v)) =$$

$$= \int_\Omega (a_{ij}v,_i v,_j + av^2)\,\mathrm{d}x + \int_{\Gamma_v} \alpha v^2\,\mathrm{d}\Gamma \geqq \min(c, a_0)\,\|v\|_1^2 = c_1\|v\|_1^2$$

where $c_1 = \min(c, a_0) > 0$. This means that the form (9) is $W^{1,2}(\Omega)$-elliptic and thus V-elliptic for each $V \subset W^{1,2}(\Omega)$.

By (1), the form $((v, u))$ is symmetric and by (17) $((v, v)) \geqq 0$ and $((v, v)) = 0$ if and only if $v = 0$. Thus $((v, u))$ is a scalar product in $W^{1,2}(\Omega)$ and in accordance with (10), (17) there exist positive constants c_1, c_2 such that

$$(18) \qquad c_1\|v\|_1 \leqq ((v, v))^{1/2} \leqq c_2\|v\|_1$$

which means that the norms $((v, v))^{1/2}$ and $\|v\|_1$ are equivalent.

3. Primary variational principle

In [14] the following theorem is proved:

Theorem 3. *Let the form* $((v, u))$ *be V-elliptic,* $((v, u)) = ((u, v))$ *and* $((v, v)) \geqq 0$. *Then u is a solution of the associated boundary value problem if and only if u minimizes the functional*

$$\mathscr{L}(u) = \tfrac{1}{2}((u, u)) - Gu \tag{19}$$

in the class $u_0 + V$.

In the preceding section we have proved that the assumptions of this theorem are satisfied. Thus the solution of the problem (13) is equivalent to the solution of the corresponding variational problem. The variational principle formulated in Theorem 3 will be called the *primary variational principle.*

4. Dual variational principle

In this section we shall formulate the dual variational principle. We want to find a functional $\mathscr{S}(\lambda)$ and a class of admissible functions $\hat{L}$ such that the equality

$$\min_{u \in u_0 + V} \mathscr{L}(u) = \max_{\lambda \in \hat{L}} \mathscr{S}(\lambda)$$

is satisfied.

Let us define the space

$$H = [L^2(\Omega)]^{N+1} \times L^2(\Gamma_v) . \tag{20}$$

We denote by

$$\|\varphi\|_C = \sum_{i=1}^{N+1} \|\varphi_i\|_0 + \|\varphi_{N+2}\|_{0,\Gamma}$$

the norm of the Cartesian product. In this norm H is complete as the Cartesian product of complete spaces.

Next, we define a mapping $\Phi : W^{1,2}(\Omega) \to H$ by the following relations:

$$\begin{aligned} \Phi_i u &= -a_{ij} u_{,j} , \quad i = 1, \ldots, N , \\ \Phi_{N+1} u &= -au , \\ \Phi_{N+2} u &= -\alpha \, . \, Tu . \end{aligned} \tag{21}$$

Let H_1, H_2, L be the following subsets of the space H:

(22) $$H_1 = \{\varphi \in H : (\exists v \in V)\ \varphi = \Phi v\},$$

(23) $$H_2 = \left\{\varphi \in H : (\forall v \in V) \int_\Omega (\varphi_i v_{,i} + \varphi_{N+1} v)\, \mathrm{d}x + \int_{\Gamma_v} \varphi_{N+2} v\, \mathrm{d}\Gamma = 0\right\},$$

(24) $$L = \left\{\varphi \in H : (\forall v \in V) \int_\Omega (\varphi_i v_{,i} + \varphi_{N+1} v)\, \mathrm{d}x + \int_{\Gamma_v} \varphi_{N+2} v\, \mathrm{d}\Gamma = -Gv\right\}.$$

For $\varphi', \varphi'' \in H$ we can define

(25) $$(\varphi', \varphi'')_H = \int_\Omega \left(a_{ij}^{-1} \varphi'_i \varphi''_j + \frac{1}{a} \varphi'_{N+1} \varphi''_{N+1}\right) \mathrm{d}x + \int_{\Gamma_v} \frac{1}{\alpha} \varphi'_{N+2} \varphi''_{N+2}\, \mathrm{d}\Gamma,$$

where a_{ij}^{-1} are elements of the inverse matrix to the matrix a_{ij}. (1), (2) implies that the inverse matrix exists and that the functions a_{ij}^{-1} are bounded, piecewise continuous and in Ω^0 they satisfy

$$a_{ij}^{-1}(x) = a_{ji}^{-1}(x), \quad i, j = 1, \ldots, N,$$
$$a_{ij}^{-1}(x)\, \xi_i \xi_j \geqq c|\xi|^2, \quad c > 0, \quad \text{for each} \quad \xi \in E_N.$$

Hence we can deduce that the bilinear form $(\varphi', \varphi'')_H$ is symmetric and there exist positive constants c_1, c_2 such that

(26) $$c_1 \|\varphi\|_C \leqq (\varphi, \varphi)_H^{1/2} \leqq c_2 \|\varphi\|_C.$$

Thus $(\varphi', \varphi'')_H$ is a scalar product in H and the norms $\|\varphi\|_C$, $\|\varphi\|_H = (\varphi, \varphi)_H^{1/2}$ are equivalent. H with the scalar product $(\varphi', \varphi'')_H$ is a Hilbert space and in the sequel H will be understood in this sense.

A simple argument shows that

(27) $$u', u'' \in W^{1,2}(\Omega) \Rightarrow (\Phi u', \Phi u'')_H = ((u', u'')).$$

H_1, H_2 are linear manifolds in H. We shall prove that H_1 is a closed subspace of H.

Let $\{\varphi^{(k)}\}_{k=1}^\infty \subset H_1$, $\varphi^{(k)} \to \varphi$. By the definition of H_1, to any k there exists $v_k \in V$ such that $\varphi^{(k)} = \Phi v_k$. $\{\varphi^{(k)}\}$ is a Cauchy sequence and thus, by (18) and (27), $\{v_k\}$ is a Cauchy sequence and $v = \lim_{k \to \infty} v_k$ exists. As V is a closed subspace of $W^{1,2}(\Omega)$, we have $v \in V$ and if we denote $\varphi' = \Phi v$, it holds $\varphi' \in H_1$ and, by (18), (27), $\varphi^{(k)} \to \varphi'$. By the uniqueness of the limit we have $\varphi' = \varphi$, thus $\varphi \in H_1$ and we have proved that H_1 is a closed subspace of H.

Let us state here the well-known lemma about the orthogonal complement. For the proof see e.g. [11], Theorem 4.82.-A.

Lemma 1. *Let H be a Hilbert space and let M be its closed subspace. Then H is the direct summ of M and its orthogonal complement $M^{\perp}$,*

$$H = M \oplus M^{\perp} .$$

As a corollary of this lemma we can prove

Theorem 4.

$$H_1^{\perp} = H_2 , \quad H = H_1 \oplus H_2 .$$

Proof. It is sufficient to prove the first assertion. The other assertion is then an immediate consequence of the first one, closedness of H_1 and of the preceding lemma.

Let $\varphi' \in H_1$. Then there exists $v \in V$ such that $\varphi = \Phi v$ and thus

$$(\varphi', \varphi'')_H = - \int_\Omega (a_{ij}^{-1} a_{ik} v_{,k} \varphi_j'' + v \varphi_{N+1}'') \, \mathrm{d}x - \int_{\Gamma_v} v \varphi_{N+2}'' \, \mathrm{d}\Gamma$$

$$= - \int_\Omega (\varphi_i'' v_{,i} + \varphi_{N+1}'' v) \, \mathrm{d}x - \int_{\Gamma_v} \varphi_{N+2}'' \, v \, \mathrm{d}\Gamma .$$

If $\varphi'' \in H_2$, then $(\varphi', \varphi'')_H = 0$ by the definition of H_2 and thus $H_2 \subset H_1^{\perp}$.
Conversely, if $(\varphi', \varphi'')_H = 0$ for any $\varphi' \in H_1$, then for any $v \in V$ it is

$$\int_\Omega (\varphi_i'' v_{,i} + \varphi_{N+1}'' v) \, \mathrm{d}x + \int_{\Gamma_v} \varphi_{N+2}'' v \, \mathrm{d}\Gamma = 0$$

and $H_1^{\perp} \subset H_2$. Thus $H_1^{\perp} = H_2$ which was to be proved.

Theorem 5. *Let u^0 be a solution of (13). Then*

$$\varphi^0 - \varphi_0 \in H_1 , \quad \varphi \in L \Rightarrow \varphi - \varphi^0 \in H_2 ,$$

where $\varphi^0 = \Phi u^0$, $\varphi_0 = \Phi u_0$.

Proof. By the definition of the solution, it holds

1° $$u^0 = u_0 + v^0 , \quad v^0 \in V ,$$

2° $$v \in V \Rightarrow \int_\Omega (a_{ij} u^0{}_{,j} v_{,i} + a v u^0) \, \mathrm{d}x + \int_{\Gamma_v} \alpha v u^0 \, \mathrm{d}\Gamma = Gv .$$

From 1° it follows that $\varphi^0 - \varphi_0 = \Phi v^0 \in H_1$.
By 2° we have

$$v \in V \Rightarrow - \int_\Omega (\varphi_i^0 v_{,i} + \varphi_{N+1}^0 v) \mathrm{d}x - \int_{\Gamma_v} \varphi_{N+2}^0 v \, \mathrm{d}\Gamma = Gv .$$

Now, if $\varphi \in L$, then

$$v \in V \Rightarrow \int_\Omega [(\varphi_i^0 - \varphi_i)\, v_{,i} + (\varphi_{N+1}^0 - \varphi_{N+1})\, v]\, \mathrm{d}x + \int_{\Gamma_v} (\varphi_{N+2}^0 - \varphi_{N+2})\, v\, \mathrm{d}\Gamma = 0\,.$$

and thus $\varphi - \varphi^0 \in H_2$.

Let us define in H the functional

$$\mathscr{T}(\varphi) = -\tfrac{1}{2}(\varphi, \varphi)_H + (\varphi, \varphi_0)_H \tag{28}$$

where $\varphi_0 = \Phi u_0$, and let us investigate its properties.

Theorem 6. *The problem*

$$\mathscr{T}(\varphi) = \max\,, \quad \varphi \in L \tag{29}$$

has the unique solution $\varphi^0 = \Phi u^0$, *where* u^0 *is the solution of* (13).

P r o o f. An easy calculation yields

$$\mathscr{T}(\varphi) = -\tfrac{1}{2}[(\varphi - \varphi_0, \varphi - \varphi_0)_H - (\varphi_0, \varphi_0)_H] = -\tfrac{1}{2}\|\varphi - \varphi_0\|_H^2 + \tfrac{1}{2}\|\varphi_0\|_H^2\,,$$

$$\mathscr{T}(\varphi^0) = -\tfrac{1}{2}\|\varphi^0 - \varphi_0\|_H^2 + \tfrac{1}{2}\|\varphi_0\|_H^2\,,$$

thus

$$\mathscr{T}(\varphi^0) - \mathscr{T}(\varphi) = \tfrac{1}{2}[\|\varphi - \varphi_0\|_H^2 - \|\varphi^0 - \varphi_0\|_H^2]\,. \tag{30}$$

Let $\varphi \in L$. Then, by Theorems 4 and 5, it holds $\varphi - \varphi^0 \perp \varphi^0 - \varphi_0$ and thus

$$\|\varphi - \varphi_0\|_H^2 = \|\varphi - \varphi^0\|_H^2 + \|\varphi^0 - \varphi_0\|_H^2 \geqq \|\varphi^0 - \varphi_0\|_H^2\,.$$

It follows that $\mathscr{T}(\varphi) \leqq \mathscr{T}(\varphi^0)$ and the equality holds if and only if $\varphi = \varphi^0$.

Let us now investigate the relation between the functionals $\mathscr{L}(u)$ and $\mathscr{T}(\varphi)$. Let $u \in W^{1,2}(\Omega)$, $\varphi = \Phi u$. Then, by (27),

$$(\varphi, \varphi)_H = ((u, u))\,.$$

By subtracting $\mathscr{T}(\varphi)$ from $\mathscr{L}(u)$ we get

$$\mathscr{L}(u) - \mathscr{T}(\varphi) = ((u - u_0, u)) - Gu\,.$$

If $u = u^0$ where u^0 is the solution of (13), then $u^0 - u_0 \in V$ and $((u^0 - u_0, u^0)) = G(u^0 - u_0)$ and thus

$$\mathscr{T}(\varphi^0) = \mathscr{L}(u^0) + Gu_0\,. \tag{31}$$

Theorem 6 together with the relation (31) represents the dual variational principle, but this form is rather unpractical. Firstly, the construction of the trial vectors $\varphi \in L$ would be very complicated and secondly, the calculation of Gu_0 in (31) requires the knowledge of the values of the function u_0 in the whole region Ω, while in most cases only the values on Γ_u are given as a boundary condition. We shall try to find another form of the dual variational principle to avoid these drawbacks. In order to succeed, we have to reduce the class of admissible functions.

For later use, let us state here Green's theorem. For the proof see [7].

Lemma 2. (*Green's theorem.*) *Let Ω be a Lipschitz region with the boundary Γ and let $u, v \in W^{1,2}(\Omega)$. Then*

$$(32) \qquad \int_\Omega u_{,i} v \, dx = \int_\Gamma uv \, n_i \, d\Gamma - \int_\Omega uv_{,i} \, dx \,, \quad i = 1, \ldots, N$$

holds, n being the unit outward normal.

Let us define

$$(33) \qquad \hat{H} = \{\lambda : \lambda = (\lambda_1, \ldots, \lambda_N) \,, \quad \lambda_i \in W^{1,2}(\Omega) \,, \quad i = 1, \ldots, N\} \,,$$

$$(34) \qquad \hat{L} = \{\lambda \in \hat{H} : -\lambda_i n_i = g \quad \text{on} \quad \Gamma_h\} \,.$$

We can define in $\hat{H}$ the symmetric bilinear form

$$(35) \qquad \langle \lambda', \lambda'' \rangle = \int_\Omega \left(a_{ij}^{-1} \lambda'_i \lambda''_j + \frac{1}{a} \lambda_{i,i} \lambda_{j,j} \right) dx + \int_{\Gamma_v} \frac{1}{\alpha} (\lambda'_i n_i)(\lambda''_j n_j) \, d\Gamma \,.$$

It holds

$$\langle \lambda, \lambda \rangle \geqq c \sum_{i=1}^N \|\lambda_i\|_0^2 \,,$$

thus $\langle \lambda, \lambda \rangle \geqq 0$, $\langle \lambda, \lambda \rangle = 0$ if and only if $\lambda = 0$. Thus $\langle \lambda', \lambda'' \rangle$ is a scalar product in $\hat{H}$. We define a norm in $\hat{H}$ by

$$\|\lambda\|_{\hat{H}} = \langle \lambda, \lambda \rangle^{1/2}$$

Using Theorem 1 we can prove that

$$\int_{\Gamma_v} \frac{1}{\alpha} (\lambda_i n_i)^2 \, d\Gamma \leqq c \int_{\Gamma_v} \sum_{i=1}^N (T\lambda_i)^2 \, d\Gamma \leqq c \|T\|^2 \sum_{i=1}^N \|\lambda_i\|_1^2 \,.$$

Further,

$$\int_\Omega \left[a_{ij}^{-1} \lambda_i \lambda_j + \frac{1}{a} (\lambda_{i,i})^2 \right] dx \leqq c \sum_{i=1}^N \|\lambda_i\|_1^2$$

and thus we have proved that there exist positive constants c_1, c_2 such that

(36) $$c_1 \sum_{i=1}^{N} \|\lambda_i\|_0^2 \leqq \langle\lambda, \lambda\rangle \leqq c_2 \sum_{i=1}^{N} \|\lambda_i\|_1^2 .$$

Let us define in $\hat{H}$ the functional

(37) $$\mathscr{S}(\lambda) = -\tfrac{1}{2} \int_\Omega \left[a_{ij}^{-1} \lambda_i \lambda_j + \frac{1}{a}(f - \lambda_{i,i})^2 \right] \mathrm{d}x - -\tfrac{1}{2} \int_{\Gamma_v} \frac{1}{\alpha} (g + \lambda_i n_i)^2 \,\mathrm{d}\Gamma - \int_{\Gamma_u} \lambda_i n_i u_0 \,\mathrm{d}\Gamma$$

and let us define the mapping $\Lambda : W^{1,2}(\Omega) \to [L^2(\Omega)]^N$:

(38) $$\Lambda_i u = -a_{ij} u_{,j} , \quad i = 1, \ldots, N .$$

We can prove

Theorem 7. *Let a solution u^0 of (13) exist such that $\Lambda u^0 \in \hat{H}$. Then the problem*

(39) $$\mathscr{S}(\lambda) = \max , \quad \lambda \in \hat{L}$$

has a unique solution $\lambda^0 = \Lambda u^0$ and it holds

(40) $$\mathscr{S}(\lambda^0) = \mathscr{L}(u^0) .$$

Proof. Let us define in $\hat{H}$ the linear functional $\mathscr{G}$ by the relation

(41) $$\mathscr{G}\lambda = \int_\Omega \frac{f}{a} \lambda_{i,i} \,\mathrm{d}x - \int_{\Gamma_v} \frac{g}{\alpha} \lambda_i n_i \,\mathrm{d}\Gamma - \int_{\Gamma_u} \lambda_i n_i u_0 \,\mathrm{d}\Gamma .$$

Then we can write

(42) $$\mathscr{S}(\lambda) = -\tfrac{1}{2} \langle\lambda, \lambda\rangle + \mathscr{G}\lambda - \tfrac{1}{2} \int_\Omega \frac{f^2}{a} \,\mathrm{d}x - \tfrac{1}{2} \int_{\Gamma_v} \frac{g^2}{\alpha} \,\mathrm{d}\Gamma .$$

Let us prove first the following assertion:

(43) $$\lambda \in \hat{L}, \quad \lambda^0 = \Lambda u^0 \Rightarrow \langle\lambda^0, \lambda^0 - \lambda\rangle = \mathscr{G}(\lambda^0 - \lambda) .$$

By the definition of u^0 and λ^0 it holds

(44) $$\begin{aligned} \lambda_{i,i}^0 &= f - a u^0 \text{ in } \Omega , \\ \lambda_i^0 n_i &= \alpha u^0 - g \quad \text{on } \Gamma_v , \\ \lambda_i^0 n_i &= -g \quad \text{on } \Gamma_h . \end{aligned}$$

We have

$$\int_\Omega a_{ij}^{-1} \lambda_i^0 (\lambda_j^0 - \lambda_j) \, dx = \int_\Omega a^{-1}_{ij}(-a_{ik} u^0_{,k}) (\lambda_j^0 - \lambda_j) \, dx =$$

$$= - \int_\Omega u^0_{,i} (\lambda_i^0 - \lambda_i) \, dx$$

and by the assumptions u^0, λ_i^0, $\lambda_i \in W^{1,2}(\Omega)$. So we can use Green's theorem concluding

$$\int_\Omega a_{ij}^{-1} \lambda_i^0 (\lambda_j^0 - \lambda_j) \, dx = - \int_\Gamma u^0 (\lambda_i^0 - \lambda_i) \, n_i \, d\Gamma + \int_\Omega u^0 (\lambda_i^0 - \lambda_i)_{,i} \, dx \, .$$

Using (44) and the relation $\lambda_i n_i = -g$ on Γ_h, we have

(45a)
$$\int_\Omega a_{ij}^{-1} \lambda_i^0 (\lambda_j^0 - \lambda_j) \, dx = \int_{\Gamma_u} (\lambda_i - \lambda_i^0) \, n_i u_0 \, d\Gamma +$$

$$+ \int_{\Gamma_v} (\lambda_i n_i - \alpha u^0 + g) \, u^0 \, d\Gamma + \int_\Omega (f - a u^0 - \lambda_{i,i}) \, u^0 \, dx \, .$$

Similarly,

(45b)
$$\int_\Omega \frac{1}{a} \lambda_{i,i}^0 (\lambda_{j,j}^0 - \lambda_{j,j}) \, dx = \int_\Omega (\lambda_{i,i} + a u^0 - f) \, u^0 \, dx +$$

$$+ \int_\Omega \frac{f}{a} (f - \lambda_{i,i} - a u^0) \, dx \, ,$$

(45c)
$$\int_{\Gamma_v} \frac{1}{\alpha} (\lambda_i^0 n_i) (\lambda_j^0 - \lambda_j) \, n_j \, d\Gamma = \int_{\Gamma_v} (\alpha u^0 - \lambda_i n_i - g) \, u^0 \, d\Gamma +$$

$$+ \int_{\Gamma_v} \frac{g}{a} (\lambda_i n_i + g - \alpha u^0) \, d\Gamma \, .$$

Summing up (45a), (45b), (45c) we get

$$\langle \lambda^0, \lambda^0 - \lambda \rangle = \int_\Omega \frac{f}{a} \, (f - a u^0 - \lambda_{i,i}) \, dx - \int_{\Gamma_v} \frac{g}{\alpha} (\alpha u^0 - g - \lambda_i n_i) \, d\Gamma -$$

$$- \int_{\Gamma_u} (\lambda_i^0 - \lambda_i) \, n_i u_0 \, d\Gamma \, ,$$

but this means exactly $\mathcal{G}(\lambda^0 - \lambda)$ and (43) is proved.

Now, it holds

$$\mathscr{S}(\lambda^0) - \mathscr{S}(\lambda) = -\tfrac{1}{2}\langle\lambda^0, \lambda^0\rangle + \tfrac{1}{2}\langle\lambda, \lambda\rangle + \mathscr{G}(\lambda^0 - \lambda) =$$
$$= \tfrac{1}{2}\langle\lambda^0 - \lambda, \lambda^0 - \lambda\rangle - \langle\lambda^0, \lambda^0 - \lambda\rangle + \mathscr{G}(\lambda^0 - \lambda)$$

and from (43) we have

(46) $$\mathscr{S}(\lambda^0) - \mathscr{S}(\lambda) = \tfrac{1}{2}\langle\lambda^0 - \lambda, \lambda^0 - \lambda\rangle\,.$$

Thus $\mathscr{S}(\lambda^0) - \mathscr{S}(\lambda) \geqq 0$ for any $\lambda \in \hat{L}$ and the equality holds if and only if $\lambda = \lambda^0$.

By the assumption of the theorem $\lambda^0 = \Lambda u^0 \in \hat{H}$ and by the third equality in (44) $\lambda_i^0 n_i = -g$ on Γ_h, thus $\lambda^0 \in \hat{L}$ and the existence and uniqueness of the solution of (39) is proved.

We have to prove the equality (40).

For $\lambda \in \hat{H}$ let us define a vector

(47a) $$\varphi = (\varphi_1, \ldots, \varphi_{N+2})\,,$$

where

(47b) $$\begin{aligned} \varphi_i &= \lambda_i\,, \quad i = 1, \ldots, N\,, \\ \varphi_{N+1} &= \lambda_{i,i} - f\,, \\ \varphi_{N+2} &= -(g + \lambda_i n_i) \quad \text{on} \quad \Gamma_v\,. \end{aligned}$$

Obviously $\varphi \in H$. We shall prove that $\varphi \in L$ for any $\lambda \in \hat{L}$ and

(48) $$\mathscr{T}(\varphi) = \mathscr{S}(\lambda) + Gu_0\,.$$

Let $v \in V$. Using Green's theorem we obtain

$$\int_\Omega (\varphi_i v_{,i} + \varphi_{N+1} v)\,\mathrm{d}x + \int_{\Gamma_v} \varphi_{N+2} v\,\mathrm{d}\Gamma = \int_\Gamma \varphi_i n_i v\,\mathrm{d}\Gamma -$$
$$- \int_\Omega (\varphi_{i,i} - \varphi_{N+1}) v\,\mathrm{d}x + \int_{\Gamma_v} \varphi_{N+2} v\,\mathrm{d}\Gamma = -Gv$$

and thus $\varphi \in L$.

Rearranging the term $(\varphi, \varphi_0)_H$ we conclude

$$\int_\Omega a_{ij}^{-1} \varphi_i \Phi_j u_0\,\mathrm{d}x = -\int_\Omega \varphi_i u_{0,i}\,\mathrm{d}x = -\int_\Gamma u_0 \varphi_i n_i\,\mathrm{d}\Gamma + \int_\Omega \varphi_{i,i} u_0\,\mathrm{d}x\,,$$

thus

$$(\varphi, \varphi_0)_H = \int_\Omega \left(a_{ij}^{-1} \varphi_i \Phi_j u_0 + \frac{1}{a} \varphi_{N+1} \Phi_{N+1} u_0\right) \mathrm{d}x + \int_{\Gamma_v} \frac{1}{\alpha} \varphi_{N+2} \Phi_{N+2} u_0 \,\mathrm{d}\Gamma =$$

$$= \int_\Omega (\varphi_{i,i} - \varphi_{N+1})\, u_0 \,\mathrm{d}x - \int_{\Gamma_h} \varphi_i n_i u_0 \,\mathrm{d}\Gamma - \int_{\Gamma_u} \varphi_i n_i u_0 \,\mathrm{d}\Gamma -$$

$$- \int_{\Gamma_v} (\varphi_i n_i + \varphi_{N+2})\, u_0 \,\mathrm{d}\Gamma = \int_\Omega f u_0 \,\mathrm{d}x + \int_{\Gamma_h \cup \Gamma_v} g u_0 \,\mathrm{d}\Gamma - \int_{\Gamma_u} \varphi_i n_i u_0 \,\mathrm{d}\Gamma =$$

$$= G u_0 - \int_{\Gamma_u} \varphi_i n_i u_0 \,\mathrm{d}\Gamma$$

and substituting into (28) we get (48).

Thus for $\lambda^0 \in \hat{L}$ it holds

$$\mathcal{T}(\varphi^0) = \mathcal{S}(\lambda^0) + G u_0 \,, \tag{48'}$$

where φ^0 is defined by the relations (47) with $\lambda = \lambda^0$.

If $\Lambda u^0 \in \hat{H}$, we can apply Green's theorem to (13) and by (14), (21), (47) we have $\varphi^0 = \Phi u^0$. In such a case (31) is valid and (31), (48') yields immediately (40) which completes the proof.

The dual variational principle formulated in Theorem 7 can be used in practical calculations. If the boundary condition g is a piecewise polynomial function, then the construction of the functions belonging to $\hat{L}$ is not difficult (we can use piecewice polynomial functions satisfying the boundary conditions). Also all the values appearing in the definition of $\mathcal{S}(\lambda)$ are known. However, if the assumption of Theorem 7 is not satisfied, then the assertion need not be valid. We shall show in the next section that even in this case we can use the dual variational principle for a posteriori error estimates of the approximate solution.

5. A posteriori error estimates

We have already mentioned in a remark after Theorem 6 that the construction of vectors belonging to the set L would be difficult. Nonetheless, if we are able to find approximate solutions to both the problems $\mathcal{L}(u) = \min$ and $\mathcal{T}(\varphi) = \max$, we can have an element to which the exact value of the error is known. The method used in the next theorem is known as the *method of hypercircle.*

Theorem 8. *Let $u \in u_0 + V$ be an approximate solution of the primary variational problem, let $\varphi \in L$ be an approximate solution of the dual variational problem* (29) *and let $\varphi^0 = \Phi u^0$, where u^0 is a solution of* (13). *Then*

$$\left\|\tfrac{1}{2}(\Phi u + \varphi) - \varphi^0\right\|_H = \tfrac{1}{2}\left\|\Phi u - \varphi\right\|_H \,. \tag{49}$$

Proof. As $u \in u_0 + V$, there exists $v \in V$ such that $u = u_0 + v$. Further, there exists $v^0 \in V$ such that $u^0 = u_0 + v^0$. It means that $u - u^0 = v - v^0 \in V$ and thus

$$\Phi u - \varphi^0 = \Phi(u - u^0) = \Phi v \in H_1 .$$

It is $\varphi \in L$ and thus, by Theorem 5, $\varphi - \varphi^0 \in H_2$. By Theorem 4 $H_1 \perp H_2$ and thus we have

$$\|\tfrac{1}{2}(\Phi u + \varphi) - \varphi^0\|_H^2 = \|\tfrac{1}{2}(\Phi u - \varphi^0) + \tfrac{1}{2}(\varphi - \varphi^0)\|_H^2 =$$
$$= \tfrac{1}{4}\|\Phi u - \varphi^0\|_H^2 + \tfrac{1}{4}\|\varphi - \varphi^0\|_H^2 = \tfrac{1}{4}\|(\Phi u - \varphi^0) - (\varphi - \varphi^0)\|_H^2 =$$
$$= \tfrac{1}{4}\|\Phi u - \varphi\|_H^2 .$$

In the following we shall show how to use the variational principle from Theorem 7 to a posteriori error estimates. Generally, we cannot assume that the assumption of Theorem 7 is satisfied. In such a case $\lambda^0 \in \hat{L}$ need not hold and the problem $\mathscr{S}(\lambda) = \max$ has no solution in $\hat{L}$.

Let us have $\lambda \in \hat{L}$. If we define φ by the relations (47), we have $\varphi \in L$. By Theorem 6, the problem $\mathscr{T}(\varphi) = \max$, $\varphi \in L$ has a solution φ^0 and (48), (31) yields

$$\mathscr{T}(\varphi) = \mathscr{S}(\lambda) + Gu_0 \leqq \mathscr{T}(\varphi^0) = \mathscr{L}(u^0) + Gu_0 \leqq \mathscr{L}(u) + Gu_0 .$$

So we have proved the following theorem:

Theorem 9. *Let $\lambda \in \hat{L}$, $u \in u_0 + V$ and let u^0 be a solution of* (13). *Then*

(50) $$\mathscr{S}(\lambda) \leqq \mathscr{L}(u^0) \leqq \mathscr{L}(u) .$$

Corollary. *Under the assumptions of Theorem* 9,

(51) $$0 \leqq \tfrac{1}{2}((u - u^0, u - u^0)) \leqq \mathscr{L}(u) - \mathscr{S}(\lambda) .$$

Proof. It is $0 \leqq \tfrac{1}{2}((u - u^0, u - u^0)) = \mathscr{L}(u) - \mathscr{L}(u^0)$ and (50) implies

$$0 \leqq \mathscr{L}(u) - \mathscr{L}(u^0) \leqq \mathscr{L}(u) - \mathscr{S}(\lambda) .$$

The inequality (51) represents an *a posteriori error estimate for the approximate solution of the primary variational problem.*

If a solution u^0 of (11) exists such that $\Lambda u^0 \in \hat{H}$, we have by (39, (40)

$$\mathscr{S}(\lambda) \leqq \max_{\lambda \in \hat{L}} \mathscr{S}(\lambda) = \mathscr{S}(\lambda^0) = \mathscr{L}(u^0) = \min_{u \in u_0 + V} \mathscr{L}(u)$$

which implies

Theorem 10. *Let $u \in u_0 + V$, $\lambda \in L$. Let a solution u^0 of* (13) *exist such that $\Lambda u^0 \in \hat{H}$. Then*

(52) $$\mathscr{S}(\lambda) \leqq \mathscr{S}(\lambda^0) \leqq \mathscr{L}(u) .$$

Corollary. *Under the assumption of Theorem* 10,

$$0 \leqq \tfrac{1}{2}\langle \lambda - \lambda^0, \lambda - \lambda^0 \rangle \leqq \mathscr{L}(u) - \mathscr{S}(\lambda)\,. \tag{53}$$

Proof. By (46)

$$\tfrac{1}{2}\langle \lambda - \lambda^0, \lambda - \lambda^0 \rangle = \mathscr{S}(\lambda^0) - \mathscr{S}(\lambda)$$

and by (52)

$$0 \leqq \mathscr{S}(\lambda^0) - \mathscr{S}(\lambda) \leqq \mathscr{L}(u) - \mathscr{S}(\lambda)\,.$$

The inequality (53) represents an *a posteriori error estimate for the approximate solution of the dual variational proble.* Moreover, if we are able to calculate $\varphi \in L$ for $\lambda \in \hat{L}$ by (47), we can use the estimate (49).

6. Approximate solutions of variational problems

In this section we shall show how to construct convergent approximations to the solutions of variational problems formulated in Sections 3 and 4. We have to solve the following problems:

$$\mathscr{L}(u) = \min\,, \quad u \in u_0 + V\,, \tag{54}$$

$$\mathscr{S}(\lambda) = \max\,, \quad \lambda \in \hat{L}. \tag{55}$$

We shall look for the approximate solutions in the finite dimensional subspaces. Our basic assumption will be the following one:

Let h, $0 < h < 1$, be a parameter. For an integer $r > 1$ let S_r^h be any finite dimensional subspace of $W^{1,2}(\Omega)$ which satisfies the condition

(*) *For each $u \in W^{r,2}(\Omega)$ there exists $\bar{u} \in S_r^h$ and a constant c independent of h and u such that*

$$\|u - \bar{u}\|_1 \leqq c\, h^{r-1} \|u\|_r\,. \tag{56}$$

In the case of a polygonal domain Ω possible examples of such subspaces are e.g. the spaces of Lagrange or Hermite interpolation polynomials on a given triangulation. For these and other examples see e.g. [3], [10], [14].

Definition. Let

$$V_r^h = V \cap S_r^h\,, \quad L_r^h = \hat{L} \cap [S_r^h]\, N\,.$$

The solutions of the problems

$$\mathscr{L}(u) = \min\,, \quad u \in u_0 + V_r^h\,, \tag{54a}$$

$$\mathscr{S}(\lambda) = \max\,, \quad \lambda \in L_r^h \tag{55a}$$

are called *the Ritz-Galerkin approximate solutions of the original problems* (54), (55).

Following [3], [10] we can prove

Theorem 11. *Let u^0 be an exact solution, let u_h^0 be a Ritz-Galerkin solution of the problem* (54). *Let $u_0 \in S_r^h$. Then*

$$\lim_{h\to 0} \|u^0 - u_h^0\|_1 = 0\,. \tag{56}$$

If $u^0 \in W^{r,2}(\Omega)$, $r > 1$, then

$$\|u^0 - u_h^0\|_1 \leqq ch^{r-1}\|u^0\|_r\,. \tag{57}$$

The convergence of the Ritz-Galerkin solution of the dual problem to its exact solution will be proved here under the assumption $\Gamma_h = \emptyset$.

Theorem 12. *Let $\Gamma_h = \emptyset$, Let $\lambda^0 \in \hat{L}$ be an exact solution of* (55), *let $\lambda_h^0 \in L_r^h$ be a Ritz-Galerkin solution of* (55). *Then*

$$\lim_{h\to 0} \|\lambda^0 - \lambda_h^0\|_{\hat{H}} = 0\,. \tag{58}$$

If $\lambda^0 \in [W^{r,2}(\Omega)]^N$, $r > 1$, then

$$\|\lambda^0 - \lambda_h^0\|_{\hat{H}} \leqq ch^{r-1}\sum_{i=1}^{N}\|\lambda_i^0\|_r\,. \tag{59}$$

Proof. Under the assumption $\Gamma_h = \emptyset$ we have $\hat{L} = \hat{H}$, $L_r^h = [S_r^h]^N$ and thus $\lambda^0 \in \hat{H}$, $\lambda_h^0 \in [S_r^h]^N$. Because $\mathscr{E}(\bar{\Omega})$ is dense in $W^{1,2}(\Omega)$, then for any $i = 1, \ldots, N$ and for any $\varepsilon_1 > 0$ there exists $\bar{\lambda}_i \in \mathscr{E}(\bar{\Omega})$ such that

$$\|\lambda_i^0 - \bar{\lambda}_i\|_1 \leqq \varepsilon_1\,.$$

By (36) we have

$$\|\lambda^0 - \bar{\lambda}\|_{\hat{H}} \leqq c\sum_{i=1}^{N}\|\lambda_i^0 - \bar{\lambda}_i\|_1 \leqq cN\varepsilon_1\,. \tag{60}$$

For any $r > 1$ it is $\bar{\lambda} \in [W^{r,2}(\Omega)]^N$ and thus, by (*), there exists $\bar{\lambda}_h \in [S_r^h]^N = L_r^h$ such that

$$\|\bar{\lambda} - \bar{\lambda}_h\|_{\hat{H}} \leqq c\sum_{i=1}^{N}\|\bar{\lambda}_i - \bar{\lambda}_{h,i}\|_1 \leqq c_1 h^{r-1}\sum_{i=1}^{N}\|\bar{\lambda}_i\|_r\,. \tag{61}$$

Inequalities (60), (61) imply the following assertion:

For any $\varepsilon > 0$ there exist

1° $\bar{\lambda} \in [\mathscr{E}(\bar{\Omega})]^N$ such that

$$\|\lambda^0 - \bar{\lambda}\|_{\hat{H}} \leqq \tfrac{1}{2}\varepsilon\,.$$

$2°$ h_0 and $\bar{\lambda}_h \in L_r^h$ such that

$$h < h_0 \Rightarrow \|\bar{\lambda} - \bar{\lambda}_h\|_{\bar{H}} \leqq \tfrac{1}{2}\varepsilon\,.$$

Summarizing 1°, 2° we get: For any $\varepsilon > 0$ there exist h_0 and $\bar{\lambda}_h \in L_r^h$ such that

$$h < h_0 \Rightarrow \|\lambda^0 - \bar{\lambda}_h\| \leqq \varepsilon\,. \tag{62}$$

By (46), (55a) we have

$$\tfrac{1}{2}\|\lambda^0 - \lambda_h^0\|_{\bar{H}} = \mathscr{S}(\lambda^0) - \mathscr{S}(\lambda_h^0) = \mathscr{S}(\lambda^0) - \max_{\lambda \in L_r^h} \mathscr{S}(\lambda) \leqq$$

$$\leqq \mathscr{S}(\lambda^0) - \mathscr{S}(\bar{\lambda}_h) = \tfrac{1}{2}\|\lambda^0 - \bar{\lambda}_h\|_{\bar{H}}\,,$$

thus

$$\|\lambda^0 - \lambda_h^0\|_{\bar{H}} \leqq \|\lambda^0 - \bar{\lambda}_h\|_{\bar{H}} \tag{63}$$

and combining (62), (63) we immediately obtain (58).

If $\lambda^0 \in [W^{r,2}(\Omega)]^N$, then, by (*), there exists $\bar{\lambda}_h \in L_r^h$ such that

$$\|\lambda^0 - \bar{\lambda}_h\|_H \leqq ch^{r-1} \sum_{i=1}^{N} \|\lambda_i^0\|_r\,.$$

Using (63) we get (59) which completes the proof.

7. Numerical results

As a numerical example, we solve the following problem:

$$-\Delta u + u = -2x(x-1) + x(x-1)\,y(y-1) - 2y(y-1)$$

in $\Omega = \langle 0, 1\rangle \times \langle 0, 1\rangle$

$$u = 0 \quad \text{on} \quad \Gamma\,.$$

It is easy to find that the function

$$u(x, y) = x(x-1)\,y(y-1)$$

with the derivatives

$$u_{,1}(x, y) = (2x - 1)\,y(y-1)\,,$$

$$u_{,2}(x, y) = (2y - 1)\,x(x-1)$$

is a solution of this problem.

In order to find the approximate solution we used the finite element method. We triangulated the domain Ω. Let us denote by D_1 the division formed in the following way: joining the centres of the opposite sides, we split the square $\langle 0, 1\rangle \times \langle 0, 1\rangle$

into four equal squares and we split each of them into four equal triangles by its diagonals. The division D_k, $k > 1$, is obtained from D_{k-1} by splitting each element of D_{k-1} into two equal right-angled triangles. If we denote by N_k the number of elements and by h_k the length of the longest side of the division D_k, then we have

$$N_1 = 16\,, \quad N_k = 2N_{k-1} \quad \text{for} \quad k > 1\,,$$
$$h_1 = 0.5\,, \quad h_k = 2h_{k-1} \quad \text{for} \quad k > 1\,.$$

Now, $\overset{\bullet}{S}{}_r^h$ will be the space of functions which are linear on each triangle of the division with the maximal side length h. Relation (*) holds for this space with $r = 2$.

Using the algorithms described in [12] we solved the problem numerically with the use of a computer. The solution of both primary and dual variational problems leads to the solution of a system of n linear algebraic equations with a symmetric band matrix with the band width m. The number of operations needed for a solution of such a problem by Gaussian elimination is roughly $\frac{1}{2}nm^2$. To reduce our problem, we can use its symmetry and solve it only in one eighth of Ω. The values n, m, $\frac{1}{2}nm^2$ are listed in Table 1.

Table. 1. Dimensions of matrices associated with the problem

			prim. problem			dual problem		
k	h_k	N_k	n	m	$p = \frac{1}{2}nm^2$	n	m	$p = \frac{1}{2}nm^2$
1	·5	2	2	1	1	3	2	6
2	·25 $\sqrt{2}$	4	3	2	6	6	4	48
3	·25	8	6	2	12	10	6	180
4	·125 $\sqrt{2}$	16	10	4	80	20	9	810
5	·125	32	20	6	360	36	14	3 528
6	·0625 $\sqrt{2}$	64	36	8	1 152	72	17	10 404

The approximate solutions corresponding to the division D_k will be denoted $u^{(k)}$, $\lambda^{(k)}$.

The assumptions of Theorems 9, 10 are satisfied. Using the a posteriori error estimates given in the corollaries to these theorems we can estimate the reduction of the error when refining the division. We define

$$\Delta_k = \sqrt{\left(\frac{\mathscr{L}(u)^{(k)}) - \mathscr{S}(\lambda^{(k)})}{\mathscr{L}(u^{(k-1)}) - \mathscr{S}(\lambda^{(k-1)})}\right)}\,.$$

The reduction of the error given by the a priori estimates of Theorems 11, 12 is

$$\delta_k = \frac{h_k}{h_{k-1}} = \sqrt{2}\,.$$

Table 2. A posteriori error estimates and the reduction of the error

k	$\mathscr{L}(u^{(k)}) - \mathscr{L}(u)$	$\mathscr{S}(\lambda) - \mathscr{S}(\lambda^{(k)})$	$\mathscr{L}(u^{(k)}) - \mathscr{S}(\lambda^{(k)})$	Δ_k^2
1	·002201	·008548	·010749	—
2	·001519	·004961	·006580	1·6336
3	·000562	·002276	·002838	2·3185
4	·000406	·001179	·001585	1·7903
5	·000140	·000576	·000716	2·2126
6	·000099	·000291	·000390	1·8219

From Table 2 we can see that the reduction of the error achieved in our calculation is in accordance with the a priori estimates.

In our problem

$$\lambda_i = -u_{,i}\,, \quad i = 1, 2$$

and the dual variational principle is thus a method for obtaining the approximate value of the derivatives of the solution. These values can be calculated also from the solution of the primary problem. If we use linear polynomials in the finite element method, the derivatives are constant in each element and we calculate the value of the derivative at a given point as the average of the values of the derivatives in all triangles meeting at this point.

As an example, let us calculate the values of the derivatives at the vertices of the division D_3. We only need to find the values of $u_{,1}$ in the set

$$M_1 = \{(\cdot 125, \cdot 125), (\cdot 375, \cdot 125), (\cdot 25, \cdot 25), (\cdot 375, \cdot 375)\}$$

and the values of $u_{,2}$ in the set

$$M_2 = \{(\cdot 25, \cdot 0), (\cdot 5, \cdot 0), (\cdot 375, \cdot 125), (\cdot 5, \cdot 25)\}\,.$$

All the other values are then known from the boundary condition and from the symmetry of the problem.

As a measure of inaccuracy we use the following quantities:

$$\varrho_k = \max_{i=1,2} \max_{(x,y)\in M_i} \left|u_{,i}^{(k)}(x, y) - u_{,i}(x, y)\right|\,,$$

$$\sigma_k = \sqrt{\Big(\sum_{i=1}^{2} \sum_{(x,y)\in M_i} [u_{,i}^{(k)}(x, y) - u_{,i}(x, y)]^2\Big)}\,.$$

In technical and physical applications we are often interested mainly in the values of derivatives on the boundary. In that case we can measure the error by the quantity

$$\tau_k = \max\big(\left|u_{,2}^{(k)}(\cdot 25, \cdot 0) - u_{,2}(\cdot 25, \cdot 0)\right|\,, \\ \left|u_{,2}^{(k)}(\cdot 5, \cdot 0) - u_{,2}(\cdot 5, \cdot 0)\right|\big)\,.$$

The values of ϱ_k, σ_k, τ_k found from the results obtained by both primary and dual methods are listed in Tables 3,3a.

If we want to compare both methods, we need to estimate the number of operations needed for the computation of derivatives from the solution of the primary problem.

Table 3. Errors in the calculation of derivatives

k	$\sigma_k^{(p)}$	$\sigma_k^{(d)}$	$\sigma_k^{(p)}/\sigma_k^{(d)}$	$\varrho_k^{(p)}$	$\varrho_k^{(d)}$	$\varrho_k^{(p)}/\varrho_k^{(d)}$
2	·10881	·04731	2·2999	·0934	·0297	3·1448
3	·07136	·02160	3·3037	·0552	·0117	4·7179
4	·05474	·00512	10·6914	·0407	·0033	12·3333
5	·03385	·00511	6·6243	·0268	·0026	10·3077

Table 3a. Errors in the calculation of derivatives on the boundary

k	$\tau_k^{(p)}$	$\tau_k^{(d)}$	$\tau_k^{(p)}/\tau_k^{(d)}$
2	·0934	·0187	4·9947
3	·0552	·0099	5·5758
4	·0407	·0033	12·3333
5	·0268	·0026	10·3077

We must perform at least two operations to find the derivative in each triangle and about 48 operations for the averaging at the points of $M_1 \cup M_2$. If we are interested only in the values on the boundary, we need about 24 operations. The numbers of operations $P^{(p)}$, $R^{(p)}$ and $P^{(d)}$ (here and in all tables the upper index p stays for primary and d for dual method) needed for the calculation of derivatives are thus given by the following relations:

$$P_k^{(p)} = p_k^{(p)} + 4N_k + 48\,,$$

$$R_k^{(p)} = p_k^{(p)} + 24\,,$$

$$P_k^{(d)} = p_k^{(d)}\,.$$

These values are listed in Table 4.

If we compare the ratios of necessary operations from Table 4 with the ratios of errors from Tables 3, 3a, we can see that in our case the dual method of calculating the derivatives is more effective than the primary one.

Table 4. Numbers of operations needed for the calculation of derivatives

k	$P^{(d)}$	$P^{(p)}$	$R^{(p)}$	$P^{(d)}/P^{(p)}$	$P^{(d)}/R^{(p)}$
2	48	70	30	·6857	1·6
3	180	92	36	1·9565	5·0
4	810	192	104	4·2188	7·7885
5	3528	532	384	6·6316	9·1875

In Table 5, the values of derivatives calculated by the primary method with the division D_4 $(P^{(p)} = 192)$ and by the dual method with the division D_3 $(P^{(d)} = 180)$ are given. We can see that the dual method is much better for the calculation of derivatives on the boundary. Thus all our results suggest that the dual method ought to be preferred when we are interested mainly in the values of derivatives on the boundary.

Table 5. Values of derivatives $u,_i$ calculated by primary method (division D_4) and by dual method (division D_3)

x	y	i	prim. $u,_i^{(p)}$	dual $u,_i^{(d)}$	exact $u,_i^{(ex)}$	$\|u,_i^{(ex)} - u,_i^{(p)}\|$	$\|u,_i^{(ex)} - u,_i^{(d)}\|$
·25	·0	2	·1552	·1958	·1875	·0323	·0083
·5	·0	2	·2093	·2699	·25	·0407	·0099
·125	·125	1	·0749	·0731	·0820	·0071	·0089
·375	·125	1	·0250	·0236	·0273	·0023	·0037
		2	·1680	·1640	·1757	·0077	·0117
·25	·25	1	·1061	·0994	·0938	·0067	·0056
·5	·25	2	·1200	·1300	·125	·0050	·0050
·375	·375	1	·0561	·0549	·0586	·0025	·0037

8. Conclusion

The problem (14) is a typical problem of a nuclear reactor theory, where it is known as a one-group neutron diffusion equation. This paper gives a theoretical background for its solution by the finite element method. The primary variational principle allows us to obtain the neutron fluxes. It is well-known and used for calculations, see e.g. [5], [6], [8]. In [2], [10] the difficulties arising from the singularities of the solution in the presence of corners and interfaces are investigated and some ways of avoiding them are suggested. By the solution of the dual problem we can obtain the neutron currents. A modification of this method is suggested in [9].

The use of the dual variational principle to a posteriori estimates of errors is suggested in [13]. Neither [9] nor [13] contain a general formulation of the dual variational principle with combined, nonhomogeneous boundary conditions. The convergence of the approximate solutions of the dual problem is not proved.

From the theoretical point of view, the works [1], [4] deal with a similar problem, but under the assumption of an infinitely differentiable boundary. This is not the case in many practical applications. Nor the combinations of all three types of boundary conditions are allowed, though they can appear simultaneously in some problems of the nuclear reactor theory.

Numerical results presented in Section 7 suggest that the dual method can serve a practical tool for approximate calculation of derivatives.

References

[1] *Aubin J. P., Burchard H. G.,* Some aspects of the method of the hypercircle applied to elliptic variational problems, 1—68, Numerical solution of partial differential equations — II, SYNSPADE 1970, ed. B. Hubbard, Academic Press, New York 1971.

[2] *Babuška I., Kellog R. D.,* Numerical solution of the neutron diffusion equation in the presence of corners and interfaces, Numerical reactor calculations, Panel IAEA-SM-154/59, Vienna 1973.

[3] *Bramble J. H., Zlámal M.,* Triangular elements in the finite element method, Math. of Comp., *24,* (1970), 809—821.

[4] *Grenacher F.,* A posteriori error estimates for elliptic partial differential equations, Technical Note BN-743, Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, 1972.

[5] *Kang C. M., Hansen K. F.,* Finite element method for the neutron diffusion equation, Trans. Am. Nucl. Soc. *14* (1971), 199.

[6] *Kaper H. G., Leaf G. K., Lindeman A. J.,* Applications of finite element method in reactor mathematics, ANL-7925, Argonne National Laboratory, Argonne, Illinois, 1972.

[7] *Nečas J.,* Les méthodes directes en théorie des équations elliptiques, Academia, Praha 1967.

[8] *Semenza L. A., Lewis E. E., Rossow E. C.,* A finite element treatment of neutron diffusion, Trans. Am. Nucl. Soc. *14,* (1971), 200.

[9] *Semenza L. A., Lewis E. E., Rossow E. C.,* Dual finite element methods for neutron diffusion, Trans. Am. Nucl. Soc., *14* (1971), 662.

[10] *Strang G., Fix G. J.,* An analysis of the finite element method, Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

[11] *Taylor A. E.,* Introduction to functional analysis, John Willey & Sons, New York, 1967.

[12] *Vacek J.,* Dual variational principles for neutron diffusion equation, thesis, MFF UK, Praha, 1974 (in Czech).

[13] *Yasinsky J. B., Kaplan S.,* On the use of dual variational principles for the estimation of error in approximate solutions of diffusion problems, Nucl. Sci. Eng., *31* (1968), 80.

[14] *Zlámal M., Ženíšek A.,* Mathematical aspects of the finite element method, Trans. of ČSAV, *81* (1971), Praha.

Souhrn

DUÁLNÍ VARIAČNÍ PRINCIPY PRO ELIPTICKOU PARCIÁLNÍ DIFERENCIÁLNÍ ROVNICI

JIŘÍ VACEK

V článku jsou zformulovány duální variační principy pro eliptickou parciální diferenciální rovnici druhého řádu s kombinovanými okrajovými podmínkami. Jsou odvozeny aposteriorní odhady chyb přibližného řešení a pro jistou třídu úloh je dokázána konvergence přibližných řešení duální úlohy k jejímu přesnému řešení. V závěru je uveden numerický příklad. Rozbor přibližných řešení naznačuje, že zvláště v úlohách, v kterých nás zajímají především hodnoty derivace podle konormály na hranici, může být duální metoda efektivním prostředkem přibližného řešení.

Author's address: Ing. *Jiří Vacek*, odd. Vývoj a výpočty ZVJE Škoda Plzeň, 316 00 Plzeň.