Riv. Mat. Univ. Parma (4) 8 (1982), 387-405

M. CAPRILI and R. LAZZERETTI (*)

A new curved element for solving transient heat conduction problems (**)

1 - Introduction

It is well known that curved isoparametric elements (Mitchell, [7]; McLeod, [6]), always have linear precision in the original space. This method involves a reduction in the order of the finite element method, and an approximation to the curved boundary. As the reduction in order cannot be avoided this method is restricted in its practibility to elements which are slightly distorted by comparison with their straight-sided counterparts. Numerical results in [2] underline the fact that isoparametric elements are extremely sensitive to distortion from the standard straight-sided elements. Furthermore, it is still a disadvantage that the function describing the boundary shape is not continuously differentiable at the common nodes of two elements, since this might disturb the thermal-stress solution in the near vicinity. It may be noted that there are several engineering problems in which this region is of high interest.

These difficulties can be overcome by the new curved quadrilateral element proposed by the authors who *match boundaries exactly*. In this way the method can be of any specified order and the function describing the boundary shape is continuously differentiable. Any finite region in two-dimensional space can be divided up into quadrilateral elements which are either straight-sided within the region or have one curved side round its perimeter.

(*) Indirizzo degli AA.: M. CAPRILI, Istituto di Matematiche Applicate, Università, 56100 Pisa, Italy; R. LAZZERETTI, Istituto di Macchine, Università, 56100 Pisa, Italy.

(**) Ricevuto: 15-I-1981.

This paper applies the curved element to the solution of parabolic initial boundary value problems. Computational procedure is outlined. Optimal error estimates in the L^∞ norm is derived for the Crank-Nicolson finite element method and stability is considered. A numerical example carried out on cooled turbine blades with listings of the FORTRAN program can be seen in Caprili and Lazzeretti [4]$_1$.

2 - Statement of the problem

The problem consists of cooling homogeneous and isotropic solid material, in which heat is being generated at a constant rate by passing fluid through a system of ducts running through the solid. The fluid is assumed to enter ducts within the solid at a given constant temperature and with a given constant velocity. It is assumed that the temperature is two dimensional, i.e. the temperature is the same for all the normal sections of the solid. To calculate the temperature, which is assumed to be varing with time, it is necessary to solve the parabolic initial boundary value problem

$$\beta(x, y)\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad (x, y, t) \in \Omega \times (0, T], \tag{1}$$

$$\frac{\partial u}{\partial n} + \alpha(x, y) u = g(x, y) \quad (x, y, t) \in \Gamma \times (0, T], \tag{2}$$

$$u(x, y, 0) = u^0(x, y) \quad (x, y) \in \bar{\Omega}, \tag{3}$$

where Ω is a multiply-connected bounded domain in the x, y plane with boundary Γ, $\beta(x, y)$ is positive and continuously differentiable away from the zero function on $\bar{\Omega}$, $\alpha(x, y)$ and $g(x, y)$ are positive continuous functions on Γ; $u^0(x, y)$ is a continuously differentiable function on $\bar{\Omega}$ and $\partial/\partial n$ is outward normal differentiation. Furthermore, we suppose that Γ consists of a finite number of simple closed curves Γ_c ($c = 0, 1, \ldots, r$) belonging to the class $C^{l+\lambda}(0 < \lambda < 1)$; the curves Γ_c ($c = 1, 2, \ldots, r$) lie inside Γ_0 and do not intersect.

The problem (1)-(3) has one and only one solution $u(x, y, t)$ which is continuous in $\bar{\Omega} \times [0, T]$, and has second order spatial derivatives and a first order time derivative which are continuous in $\Omega \times (0, T]$, [**5**].

Before formulating the given problem in weak variational form, let us introduce some notations. By $H^m(\Omega)$ we denote the Sobolev space of real functions which, together with their generalized derivatives up to the m-th order included, are square integrable in Ω. The norm is defined by

$$\|u\|_{m,\Omega} = \{\sum_{|\alpha|\leqslant m} \int_{\Omega} (D^{\alpha}u)^2 \,dx\,dy\}^{\frac{1}{2}},$$

and the seminorm by

$$|u|_{m,\Omega} = \{\sum_{|\alpha|=m} \int_{\Omega} (D^{\alpha}u)^2 \,dx\,dy\}^{\frac{1}{2}}.$$

Remark. For simplicity, we will sometimes write $\|\cdot\|_m$ and $|\cdot|_m$ instead of $\|\cdot\|_{m,\Omega}$ and $|\cdot|_{m,\Omega}$, respectively.

If X is a normed space with norm $\|\cdot\|_X$ and $\varphi\colon [0, T] \to X$, then

$$\|\varphi\|_{L^{\infty}(X)} = \sup_{0\leqslant t\leqslant T} \|\varphi(t)\|_X .$$

Multiplying (1) by $v \in H^1(\Omega)$ and using Green's theorem and boundary condition (2) we obtain the identity

$$(4)\qquad \begin{aligned} &(\beta(x, y)\frac{\partial u}{\partial t}, v) + a(u, v) = \langle g(x, y), v\rangle && (v \in H^1(\Omega),\ t \in (0, T]),\\ &u(x, y, 0) = u^0(x, y) && (x, y) \in \bar{\Omega}, \end{aligned}$$

where

$$(5)\qquad (\beta(x, y)\frac{\partial u}{\partial t}, v) = \int_{\Omega} \beta(x, y)\frac{\partial u}{\partial t} v \,dx\,dy,$$

$$(6)\qquad a(u, v) = \int_{\Omega} (\frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y}) \,dx\,dy + \int_{\Gamma} \alpha(x, y) uv \,d\Gamma$$

$$(7)\qquad \langle g(x, y), v\rangle = \int_{\Gamma} g(x, y) v \,d\Gamma .$$

Hence the variational principle corresponding to the initial boundary value problem (1)-(3) is to find for each time $t \in (0, T]$ the function $u \in H^1(\Omega)$ that satisfies (4).

It is worth noting that, because of the symmetry of the bilinear form (6), the variational formulation (4) is equivalent to the problem of finding, for each time $t \in (0, T]$, the function $u \in H^1(\Omega)$ that, besides the initial condition (3), minimizes the functional

$$(8)\qquad I(v)=\frac{1}{2}\int_\Omega\{(\frac{\partial v}{\partial x})^2+(\frac{\partial v}{\partial y})^2+2\beta(x,y)\,v\,\frac{\partial v}{\partial t}\}\,\mathrm{d}x\,\mathrm{d}y$$

$$+\int_\Gamma\{\frac{1}{2}\,\alpha(x,y)\,v^2-g(x,y)\,v\}\;\mathrm{d}\Gamma.$$

Both these formulations of the problem will be useful in the discussion which follows.

To formulate the problem in terms of the finite element method, we must perform the spacewise discretization of functional (8). We shall use finite element spaces which are subspace of $H^1(\Omega)$. This means that, since Γ is curved, we have to consider curved elements which exactly match the curved boundary. We denote the finite element space to be used, S_h^1.

The finite element discretization of (8) in space gives for each time $t\in(0,T]$ the $w(x,y,t)$ function from S_h^1 such that

$$(9)\qquad \begin{aligned} &I(w)\leqslant I(v) && v\in S_h^1,\\ &w(x,y,0)=\hat{u}^0(x,y) && \hat{u}^0(x,y)\in S_h^1; \end{aligned}$$

$\hat{u}^0(x,y)$ is an approximation of $u^0(x,y)$ and the simplest way is to choose the interpolation of $u^0(x,y)$ for it. It must be noted that the time variable is left continuous and the determination of $w(x,y,t)$ requires the solution of a initial-value problem for a system of ordinary differential equations.

To define a computable approximation we must discretize with respect to time in some fashion. Since the differential system, as will be proved exactly later, is *stiff* we shall use the second order, A-stable, Crank-Nicolson algorithm.

3 - Approximate solution

3.1 – *Spatial approximation.* We begin by covering the domain $\bar{\Omega}$ by a finite union of finite elements which are of two possible types: elements which have no points in common with the boundary Γ are *quadrilaterals* whose sides are all straight, whereas those which are constructed by using the parametric equation of the boundary Γ are generally *curved quadrilaterals*, i.e., a quadrilateral one of whose sides can be distorted in a prescribed way. Two adjacent elements have a common side and the open elements are disjoined. Let e be the generic quadrilateral of the finite element partition of $\bar{\Omega}$. Let us suppose that the quadrilateral whose sides are all straight and which has the same vertices as e is convex, and let us denote by h_e and $\hat{h}_e$ the greatest and the smallest sides of this quadrilateral.

If $h = \max_{e\subset\bar{\Omega}} h_e$ and $\hat{h} = \min_{e\subset\bar{\Omega}} \hat{h}_e$, we suppose that $\hat{h}/h \geqslant c$ where c is a positive constant (notice that in all cases $\hat{h}/h \leqslant 1$).

We denote by Π_h the given partitioning of $\bar{\Omega}$ and by $P_r \equiv (x_r, y_r)$ $(r = 1, 2, \ldots, l)$ its nodes. We will only consider the partitioning such that h is less than one and θ, the smallest angle of all those defined by a diagonal and a side of any quadrilateral (in the case of a curved quadrilateral we mean the quadrilateral whose sides are all straight sides and which has the same vertices), is bounded away from zero by θ_0 when h tends to zero.

Let $P_{r_i} \equiv (x_{r_i}, y_{r_i})$ (r_i positive integer; $i = 1, 2, 3, 4$) be nodes belonging to an element, and we denote by E the square of vertices $R_1 = (0, 0)$, $R_2 = (1, 0)$, $R_3 = (1, 1)$, $R_4 = (0, 1)$ in an auxiliary space ξ, η.

Let e_s be a quadrilateral whose sides are all straight. In Strang and Fix [**9**] appear the

Theorem 1. *The equations*

$$(10)\qquad \begin{aligned} x = x_0(\xi, \eta) &= x_{r_1} + (x_{r_2} - x_{r_1})\xi + (x_{r_4} - x_{r_1})\eta + (x_{r_1} - x_{r_2} + x_{r_3} - x_{r_4})\xi\eta \\ y = y_0(\xi, \eta) &= y_{r_1} + (y_{r_2} - y_{r_1})\xi + (y_{r_4} - y_{r_1})\eta + (y_{r_1} - y_{r_2} + y_{r_3} - y_{r_4})\xi\eta \end{aligned}$$

map the closed square E one-to-one on the closed quadrilateral e_s.

Now, let e_c be the curved quadrilateral with three straight sides and one curved side. Here the curved side is a part of the boundary Γ. These authors in [**4**]$_2$ have proved the

Theorem 2. *If h_{e_c} is sufficiently small the non-linear mapping*

$$(11)\qquad x = x(\xi, \eta) = x_0(\xi, \eta) + \eta\Phi(\xi), \qquad y = y(\xi, \eta) = y_0(\xi, \eta) + \eta\Psi(\xi),$$

where

$$\Phi(\xi) = \varphi[(s_{r_3} - s_{r_4})\xi + s_{r_4}] - x_{r_4} - (x_{r_3} - x_{r_4})\xi,$$

$$\Psi(\xi) = \psi[(s_{r_3} - s_{r_4})\xi + s_{r_4}] - y_{r_4} - (y_{r_3} - y_{r_4})\xi,$$

maps E one-to-one on e_c, the Jacobian $J(\xi, \eta) = \partial(x, y)/\partial(\xi, \eta)$ do not vanish in E, the sides $\overline{R_1R_2}$, $\overline{R_2R_3}$, $\overline{R_4R_1}$ are mapped linearly on the sides $\overline{P_{r_1}P_{r_2}}$, $\overline{P_{r_2}P_{r_3}}$, $\overline{P_{r_4}P_{r_1}}$, respectively and the side $\overline{R_3R_4}$ is mapped on the arc $\widehat{P_{r_3}P_{r_4}}$.

The $x = \varphi(s)$, $y = \psi(s)$, $s_{r_3} \leqslant s \leqslant s_{r_4}$, are the *parametric equations* of the regular arc $\widehat{P_{r_3} P_{r_4}}$ and $x_{r_3} = \varphi(s_{r_3})$, $x_{r_4} = \varphi(s_{r_4})$, $y_{r_3} = \psi(s_{r_3})$ and $y_{r_4} = \psi(s_{r_4})$. Moreover, the functions $\varphi(s)$ and $\psi(s)$ are supposed to be continuous along with their first and second order derivatives.

Furthermore, in actual calculation it is not necessary to obtain the inverse mapping of eqs. (10) and (11).

The first-degree polynomial in any variable ξ, η

$$w^E(\xi, \eta) = a_1 + a_2\xi + a_3\eta + a_4\xi\eta \tag{12}$$

is defined by the values W_1^E, W_2^E, W_3^E and W_4^E which holds at the points R_1, R_2, R_3 and R_4, respectively. In fact, the conditions $w^E(R_i) = W_i^E$, $(i = 1, 2, 3, 4)$, given to the polynomial (12), yield, *in formula*, a linear system of four equations in the unknouns a_1, a_2, a_3, a_4 whose matrix is non-singular.

Equation (12) can now be written

$$w^E(\xi, \eta) = \sum_{i=1}^{4} N_i(\xi, \eta)\, W_i^E\,, \tag{13}$$

where

$$\begin{aligned} &N_1(\xi, \eta) = (1-\xi)(1-\eta)\,, \qquad &&N_2(\xi, \eta) = \xi(1-\eta)\,, \\ &N_3(\xi, \eta) = \xi\eta\,, &&N_4(\xi, \eta) = (1-\xi)\eta\,. \end{aligned} \tag{14}$$

Hence the function

$$w^e(x, y, t) = \sum_{i=1}^{4} N_i[\xi(x, y), \eta(x, y)]\, W_{r_i}(t) \tag{15}$$

defined in e_s or e_c by the inverse mapping of (10) or (11), respectively, assumes the values $W_{r_1}(t)$, $W_{r_2}(t)$, $W_{r_3}(t)$ and $W_{r_4}(t)$ at the vertices P_{r_1}, P_{r_2}, P_{r_3} and P_{r_4}, respectively.

It may easily be proved that the function $w(x, y, t)$ defined in $\bar{\Omega}$, for each $t \in (0, T]$, which is equal to the function $w^e(x, y, t)$ in every element $e \subset \bar{\Omega}$, is continuous on $\bar{\Omega}$. This condition, with the properties of the equations (10), (11) and (13) is sufficient in order that the *trial function* $w(x, y, t)$ belongs, for each $t \in (0, T]$, to the Sobolev space $H^1(\Omega)$.

From Caprili and Lazzeretti [4]$_1$ the functional (8) can be written in the form

$$I_h(W_1, W_2, \ldots, W_l) = \tfrac{1}{2}\sum_{r,s=1}^{l} A_{r,s}\, W_r\, W_s + \tfrac{1}{2}\sum_{r,s=1}^{l} S_{r,s}\, \dot{W}_r\, W_s - \sum_{r=1}^{l} F_r\, W_r\,, \tag{16}$$

where the dot denotes the derivative with respect to time.

Since $A = [A_{r,s}]$ and $S = [S_{r,s}]$, $(r, s = 1, 2, \dots, l)$, are constant symmetric and positive definite matrices $[4]_1$, then the vector $W(t) = [W_1(t), W_2(t), \dots, W_l(t)]^T$ which minimized (16) is the solution of the initial-value problem of l linear differential equations

$$\dot{W} = -S^{-1}AW + S^{-1}F, \qquad W(0) = W^0, \tag{17}$$

where $W^0 = [u^0(x_1, y_1), u^0(x_2, y_2), \dots, u^0(x_l, y_l)]^T$.

To obtain a computable solution we have to discretize (17) in time.

3.2 - *Time approximation.* An important property of the matrix $S^{-1}A$ is that its eigenvalues μ are positive. Furthermore, as will be shown below, the system (17) is *stiff*, so that a method with a region of absolute stability containing the interval $(-\infty, 0)$ must be used to solve it. Such a method is unconditionally stable.

If we let the approximate value of the vector $W(n\Delta t)$, where Δt is the time step, be denoted by W^n, then the value W^{n+1} obtained by the Crank-Nicolson algorithm is determined by solving the linear system

$$\left(S + \frac{\Delta t}{2} A\right) W^{n+1} = \left(S - \frac{\Delta t}{2} A\right) W^n + \Delta t F. \tag{18}$$

The solution of system (18) required the solution of a system of linear algebraic equations at every time step.

Thus, since the matrices A and S are symmetric and definite positive, the matrix $Q = S + (\Delta t/2)A$ has the same properties for any $\Delta t > 0$, and its spectral condition number, as is shown below, is $\mathcal{O}(\Delta t h^{-2})$. The system (18) is therefore normal and no exchanges of rows by the Gauss method are necessary for its solution. In this case, elimination without row exchanges is not only possible but also numerically stable.

4 - Error estimates

In the following we will use C, C_1, C_2, ... to represent positive constants whose values may change when their position changes. Let us now assume

Lemma 1. *Let $u(x, y, t)$ be a function which belongs to $H^2(\Omega)$ at any fixed $\bar{t} \in [0, T]$* and *$\hat{u}(x, y)$ the function in S_h^1 that assumes the same nodal values as $u(x, y, \bar{t})$ in $\bar{\Omega}$. Then*

$$\|u - \hat{u}\|_{1,\Omega} \leq Ch\|u\|_{2,\Omega}. \tag{19}$$

Proof. As $\hat{u} \in H^1(\Omega)$, we need only prove (19) for a single element. We will prove (19) in the case of a curved quadrilateral element e_c as follows

$$\|u - \hat{u}\|_{1,e_c} \leqslant C_1 h \|u\|_{2,e_c} . \tag{20}$$

The inequality (19) for a quadrilateral element e_s is a special case of (20).

For equation (11), let us suppose that $r(\xi, \eta, \bar{t}) \equiv u(x(\xi, \eta), y(\xi, \eta), \bar{t})$ and $\hat{r}(\xi, \eta) \equiv \hat{u}(x(\xi, \eta), y(\xi, \eta))$; then considering that according to [4]$_2$

$$(21)_1 \qquad \left|\frac{\partial(x, y)}{\partial(\xi, \eta)}\right| = \mathcal{O}(h^2), \qquad \left|\frac{\partial x}{\partial \xi}\right| = \mathcal{O}(h), \qquad \left|\frac{\partial y}{\partial \xi}\right| = \mathcal{O}(h),$$

$$(21)_2 \qquad \left|\frac{\partial^2 x}{\partial \xi^2}\right| = \mathcal{O}(h^2), \qquad \left|\frac{\partial^2 y}{\partial \xi^2}\right| = \mathcal{O}(h^2), \qquad \left|\frac{\partial x}{\partial \eta}\right| = \mathcal{O}(h),$$

$$(21)_3 \qquad \left|\frac{\partial y}{\partial \eta}\right| = \mathcal{O}(h), \qquad \frac{\partial^2 x}{\partial \eta^2} = 0, \qquad \frac{\partial^2 y}{\partial \eta^2} = 0,$$

by using simple calculations we obtain

$$\|u - \hat{u}\|^2_{1,e_c} \leqslant C_2 \|r - \hat{r}\|^2_{1,E} . \tag{22}$$

To estimate the quantity $\|r - \hat{r}\|_{1,E}$, we may now define the linear functional on $H^2(E)$

$$F(r) = (r - \hat{r}, s)_{1,E} , \tag{23}$$

where $s \in H^1(E)$. As $r \in H^2(E)$, it is true that

$$\max_E |r| \leqslant C_3 \|r\|_{2,E} \tag{24}$$

for Sobolev's lemma, [9].

At the vertices of E, the polynomial $\hat{r}(\xi, \eta)$ takes on the values $r(\xi, \eta, \bar{t})$, so that, by means of (24), we easily obtain

$$\|\hat{r}\|_{1,E} \leqslant C_4 \|r\|_{2,E} . \tag{25}$$

So, applying (23), we obtain

$$|F(r)| \leqslant \|r - \hat{r}\|_{1,E} \|s\|_{1,E} \leqslant \|s\|_{1,E} \big(\|r\|_{1,E} + \|\hat{r}\|_{1,E}\big)$$

and, using (25),

$$|F(r)| \leq C_5 \|s\|_{1,E} \|r\|_{2,E} .$$

The functional (23) is therefore linear and bounded and, given that the interpolating function $\hat{r} \in S_h^1$ is unique, it is equal to zero for each polynomial of degree ≤ 1 in every variable, so it follows for Bramble and Hilbert, [**3**]$_{1,2}$, that

$$(26) \qquad |F(r)| \leq C_6 \|s\|_{1,E} \left\{ \left\|\frac{\partial^2 r}{\partial \xi^2}\right\|_{0,E}^2 + \left\|\frac{\partial^2 r}{\partial \eta^2}\right\|_{0,E}^2 \right\}^{\frac{1}{2}}.$$

If we choose $s = r - \hat{r}$ (23) and (26) yield

$$\|r - \hat{r}\|_{1,E} \leq C_6 \left\{ \left\|\frac{\partial^2 r}{\partial \xi^2}\right\|_{0,E}^2 + \left\|\frac{\partial^2 r}{\partial \eta^2}\right\|_{0,E}^2 \right\}^{\frac{1}{2}}.$$

If we now return to the plane x, y and use (21), we obtain

$$\|r - \hat{r}\|_{1,E}^2 \leq C_7 h^2 \{ |u|_{1,e_c}^2 + |u|_{2,e_c}^2 \} .$$

If we now make use of (22), we find that

$$(27) \qquad \|u - \hat{u}\|_{1,e_c}^2 \leq C_2 C_7 h^2 \{ |u|_{1,e_c}^2 + |u|_{2,e_c}^2 \} ,$$

and thus also (20). If we now sum (27) on all the elements, we obtain

$$\|u - \hat{u}\|_{1,\Omega}^2 \leq C_2 C_7 h^2 \{ |u|_{1,\Omega}^2 + |u|_{2,\Omega}^2 \} ,$$

and thus also (19).

Lemma 2. *If* $\omega \in H^2(\Omega) \cap C^1(\bar{\Omega})$ *is such that*

$$(28) \qquad a(\omega, v) = (\zeta, v) \qquad \textit{for each } v \in H^1(\Omega) ,$$

then

$$(29) \qquad \|\omega\|_{2,\Omega} \leq C \|\zeta\|_{0,\Omega} .$$

Proof. From the book of Agmon [**1**], since $\Gamma \in C^2$, it appears that

$$(30) \qquad \|\omega\|_{2,\Omega} \leq C_1 (\|\zeta\|_{0,\Omega} + \|\omega\|_{0,\Omega}) .$$

If we let $v = \omega$ in (28), then since $\alpha(x, y)$ is a continuous positive function on Γ, we obtain $|\omega|^2_{1,\Omega} \leqslant (\zeta, \omega)$ and $\alpha_0\langle\omega, \omega\rangle \leqslant (\zeta, \omega)$ where $\alpha_0 = \min_{(x,y)\in\Gamma} \alpha(x, y)$. From Friedrichs inequality we obtain $\|\omega\|^2_{0,\Omega} \leqslant C_2(\langle\omega, \omega\rangle + |\omega|^2_{1,\Omega})$, which together with the two last inequalities, gives $\|\omega\|_{0,\Omega} \leqslant C_2(1 + 1/\alpha_0)\|\zeta\|_{0,\Omega}$ where we have used the Schwarz inequality. Thus, on combining the last inequality and (30), we obtain (29) with $C = C_1 C_2(1 + 1/\alpha_0)$.

Let us now estimate the global error, that is, the error due to spatial approximation plus that due to temporal approximation. To do this, we use the fact that the function $w(x, y, t)$ which satisfied (9) is such that

$$(31) \qquad \Big(\beta(x, y)\frac{\partial w}{\partial t}, v\Big) + a(w, v) = \langle g(x, v), y\rangle \qquad \forall v \in S^1_h, \quad t \in (0, T].$$

Furthermore, the scalar product (6) defines in $H^1(\Omega)$ a norm equivalent to $\|\cdot\|_{1,\Omega}$ (Necas [**8**]); in other words, we may conclude that

$$(32) \qquad C_1\|v\|^2_{1,\Omega} \leqslant a(v, v) \leqslant C_2\|v\|^2_{1,\Omega} \qquad \forall v \in H^1(\Omega).$$

Given these premisses, we obtain

Theorem 3. *If all the following conditions hold: the exact solution $u(x, y, t)$ to the problem (1)-(3) belongs to $H^2(\Omega)$ when $t \in [0, T]$, $\partial^3 u/\partial t^3$ is continuous when $(x, y, t) \in \bar{\Omega} \times [0, T]$, $\|\dot{u}\|_2 \leqslant C$ when $t \in [0, T]$, $u^0(x, y)$ belongs to $H^2(\Omega)$, the functions $\alpha(x, y)$, $g(x, y)$ are continuous and positive when $(x, y) \in \Gamma$ and $\beta(x, y)$ is a continuously differentiable function bounded from below by a β_0 positive number when $(x, y) \in \bar{\Omega}$, then, when Δt and h are chosen arbitrarily,*

$$(33) \qquad \max_{0 \leqslant n \leqslant T/\Delta t} \|u^n - w^n\|_1 \leqslant C_1 h\|u^0\|_2 + C_2 h + C_3(\Delta t^2 + h^2)$$

and also

$$(34) \qquad \|u^n - w^n\|_{L^\infty(H^1)} \leqslant C_4(h + \Delta t^2),$$

where $u^n = u(x, y, n\Delta t)$, and $w^n = w(x, y, n\Delta t)$ belongs to S^1_h for every $0 \leqslant n \leqslant T/\Delta t$ and satisfies the condition $w(x_i, y_i, n\Delta t) = W^n_i$, $(i = 1, 2, \dots, l)$, and the constants C_1, C_2, C_3 and C_4 are independent of Δt and h.

Proof. For convenience, let us suppose that $u^n = \phi^n + \zeta^n$, where ϕ^n is the orthogonal projection, for every given $0 \leqslant n \leqslant T/\Delta t$, of u^n over S^1_h in the norm $(a(\cdot,\cdot))^{\frac{1}{2}}$. Thus, applying the Lemma 1 and (32), we obtain

$$(35) \qquad \|\zeta^n\|_1 = \|u^n - \phi^n\|_1 \leqslant C_1 h\|u^n\|_2.$$

It is also true that

(36) $$\|\zeta^n\|_0 \leqslant C_2 h \|\zeta^n\|_1 .$$

In fact, if $\omega \in H^2(\Omega) \cap C^1(\bar{\Omega})$ is such that $a(\omega, v) = (\zeta^n, v) \quad \forall v \in H^1(\Omega)$, then, if we choose $v = \zeta^n$,

(37) $$\|\zeta^n\|_0^2 = a(\omega, \zeta^n) = a(\omega - \hat{\omega}, \zeta^n) ,$$

where $\hat{\omega} \in S_h^1$ is the interpolating function of ω. From (37), Lemma 1 and (32) we obtain $\|\zeta^n\|_0^2 \leqslant C_3 \|\omega - \hat{\omega}\|_1 \|\zeta^n\|_1 \leqslant C_4 h \|\omega\|_2 \|\zeta^n\|_1$. For Lemma 2 we find $\|\zeta^n\|_0^2 \leqslant C_2 h \|\zeta^n\|_0 \|\zeta^n\|_1$, which, if simplified, yields (36). Besides this, (35) gives us

(38) $$\|\zeta^n\|_0 \leqslant C_5 h^2 \|u^n\|_2 .$$

From (4) we find that

(39) $$(\beta(x, y) \frac{\partial u^n}{\partial t}, v) + a(\phi^n, v) = \langle g(x, y), v \rangle \quad \forall v \in S_h^1 .$$

If w^n indicates the function w at time $t = n\Delta t$, $n = 0, 1, 2, \ldots$, then the global truncation error at time $t = n\Delta t$ is

(40) $$\|u^n - w^n\|_1 \leqslant \|u^n - \phi^n\|_1 + \|\phi^n - w^n\|_1 \leqslant C_1 h \|u^n\|_2 + \|\phi^n - w^n\|_1 .$$

We must now examine $\|\phi^n - w^n\|_1$. By applying the Crank-Nicolson method and using (31), the following recurring relationship is obtained

(41) $$(\beta(w^{n+1} - w^n), v) + \frac{\Delta t}{2} a(w^{n+1} + w^n, v) = \Delta t \langle g, v \rangle \quad \forall v \in S_h^1 .$$

As the local truncation error in the Crank-Nicolson method is

$$\gamma^n = u^{n+1} - u^n - \frac{\Delta t}{2} (\dot{u}^{n+1} + \dot{u}^n) = \mathcal{O}(\Delta t^3 \max_{t \in [0,T]} |\frac{\partial^3 u}{\partial t^3}|) ,$$

we find that

(42) $$\|\gamma^n\|_0 \leqslant C_6 \Delta t^3$$

and, besides, using (39)

$$(43)\qquad (\beta(\phi^{n+1}-\phi^n), v) + \frac{\Delta t}{2} a(\phi^{n+1}+\phi^n, v)$$

$$= \Delta t\langle g, v\rangle + (\beta\gamma^n, v) - (\beta(\zeta^{n+1}-\zeta^n), v) \qquad \forall v \in S_h^1 .$$

If we subtract (41) from (43), given that $g(x, y)$ is not dependent on t, we obtain

$$(44)\qquad (\beta(\varepsilon^{n+1}-\varepsilon^n), v) + \frac{\Delta t}{2} a(\varepsilon^{n+1}+\varepsilon^n, v) = (\beta\gamma^n, v) - (\beta(\zeta^{n+1}-\zeta^n), v) \qquad \forall v \in S_h^1 ,$$

where $\varepsilon^n = \phi^n - w^n$.

Since $a(\zeta, v) = 0$ for any $v \in S_h^1$ and for every $t \in (0, T]$ then, if it is differentiated with respect to t, we find that $a(\dot\zeta, v) = 0$ $\forall v \in S_h^1$. Thus $\dot\phi = \dot u - \dot\zeta$ is the projection of $\dot u$ over S_h^1 in the norm $(a(\cdot,\cdot))^{\frac{1}{2}}$ and, applying (38), we obtain

$$(45)\qquad \|\dot\zeta\|_0 \leqslant C_7 h^2 \|\dot u\|_2 .$$

In addition, $\zeta^{n+1}-\zeta^n = (\partial\zeta^{\bar n}/\partial t)\Delta t$, where $\bar n = n + \Theta\Delta t$, $0 < \Theta < 1$ and so, using (45), we find that

$$(46)\qquad \|\zeta^{n+1}-\zeta^n\|_0 = \left\|\frac{\partial\zeta^{\bar n}}{\partial t}\right\|_0 \Delta t \leqslant C_7 \Delta t h^2 \|\dot u^{\bar n}\|_2 \leqslant C C_7 \Delta t h^2 .$$

If we choose $v = \varepsilon^{n+1}-\varepsilon^n$ and use the inequality $|ab| \leqslant a^2/2\delta + \delta b^2/2$ where a and b are real and δ is positive and arbitrary, (44) becomes

$$(\beta(\varepsilon^{n+1}-\varepsilon^n), \varepsilon^{n+1}-\varepsilon^n) + \frac{\Delta t}{2} a(\varepsilon^{n+1}+\varepsilon^n, \varepsilon^{n+1}-\varepsilon^n)$$

$$= (\beta\gamma^n, \varepsilon^{n+1}-\varepsilon^n) - (\beta(\zeta^{n+1}-\zeta^n), \varepsilon^{n+1}-\varepsilon^n) ,$$

and also

$$\beta_0\|\varepsilon^{n+1}-\varepsilon^n\|_0^2 + \frac{\Delta t}{2}\left(a(\varepsilon^{n+1}, \varepsilon^{n+1}) - a(\varepsilon^n, \varepsilon^n)\right)$$

$$\leqslant \|\beta\gamma^n\|_0\|\varepsilon^{n+1}-\varepsilon^n\|_0 + \|\beta(\zeta^{n+1}-\zeta^n)\|_0\|\varepsilon^{n+1}-\varepsilon^n\|_0$$

and, in conclusion, indicating the norm $(a(\cdot,\cdot))^{\frac{1}{2}}$ by $|||\cdot|||$, we find that

$$\beta_0 \|\varepsilon^{n+1} - \varepsilon^n\|_0^2 + \frac{\Delta t}{2}(|||\varepsilon^{n+1}|||^2 - |||\varepsilon^n|||^2) \tag{47}$$

$$\leqslant \frac{\delta}{2}\|\beta\gamma^n\|_0^2 + \frac{1}{2\delta}\|\varepsilon^{n+1} - \varepsilon^n\|_0^2 + \frac{\delta}{2}\|\beta(\zeta^{n+1} - \zeta^n)\|_0^2 + \frac{1}{2\delta}\|\varepsilon^{n+1} - \varepsilon^n\|_0^2 .$$

Let us choose $\delta = 1/\beta_0$; then, after simplification, (47) yields

$$\frac{\Delta t}{2}(|||\varepsilon^{n+1}|||^2 - |||\varepsilon^n|||^2) \leqslant \frac{\beta_1^2}{2\beta_0}(\|\gamma^n\|_0^2 + \|\zeta^{n+1} - \zeta^n\|_0^2),$$

where $\beta_1 = \max_{(x,y)\in\bar{\Omega}} \beta(x, y)$.

At this point, application of (42) and (46) yields

$$|||\varepsilon^{n+1}|||^2 \leqslant |||\varepsilon^0|||^2 + C_8(n\Delta t^5 + n\Delta t h^4) .$$

Besides, since $0 \leqslant n \leqslant T/\Delta t$ the final inequality may now be expressed as

$$|||\varepsilon^{n+1}|||^2 \leqslant |||\varepsilon^0|||^2 + C_9(\Delta t^4 + h^4) . \tag{48}$$

Considering that at the initial time $\varepsilon^0 = \phi^0 - w^0 = \phi^0 - \hat{u}^0$ and that ϕ^0 is the orthogonal projection of u^0 over S_h^1 in the norm $|||\cdot|||$, then, using Pythagoras's theorem,

$$|||\varepsilon^0||| \leqslant |||u^0 - \hat{u}^0||| . \tag{49}$$

Substituting (49) in (48), and considering that $|||\cdot|||$ is equivalent to the norm $\|\cdot\|_1$, we now find that

$$|||\varepsilon^n|||_1 \leqslant C_{10}\|u^0 - \hat{u}^0\|_1 + C_{11}(\Delta t^2 + h^2) , \qquad 0 \leqslant n \leqslant T/\Delta t .$$

This, together with (40) and (19), yields

$$\|u^n - w^n\|_1 \leqslant C_1 h\|u^n\|_2 + C_1 C_{10} h\|u^0\|_2 + C_{11}(\Delta t^2 + h^2) , \qquad 0 \leqslant n \leqslant T/\Delta t . \tag{50}$$

Finally, as $\|u^n\|_2 \leqslant C_{12}$ when $0 \leqslant n \leqslant T/\Delta t$, by using (50) we obtain (33), and the theorem has now been proved.

5 - Spectral condition number

We will now demonstrate that system (17) is *stiff*. For a given partition of Ω, we will use $v_1, v_2, \ldots, v_l$ to signify a base of the subspace S_h^1 which possesses the property $v_i(P_j) = \delta_{i,j}$ $(i, j = 1, 2, \ldots, l)$; it is clear that a function $v_i(x, y)$ $(i = 1, 2, \ldots, l)$ may be generated by means of the functions $N_i(\xi, \eta)$ $(i = 1, 2, 3, 4)$ and equations (10) and (11) in such way that it is nonvanishing only for the elements which share the nodal point P_i. We will now postulate the lemmas.

Lemma 3. *The system of functions* $\{v_i\}_{i=1}^{l}$ *has the property*

$$\frac{a(v_i, v_i)}{\|v_i\|_{0,\Omega}^2} \geqslant Ch^{-2} \tag{51}$$

as h *tends to zero.*

Proof. Since the norms $(a(\cdot,\cdot))^{\frac{1}{2}}$ and $\|\cdot\|_1$ are equivalent, instead of (51) we may prove, as h is sufficiently small, the following

$$\frac{|v_i|_{1,\Omega}^2}{\|v_i\|_{0,\Omega}^2} \geqslant C_1 h^{-2}. \tag{52}$$

If $e_{i,1}, e_{i,2}, \ldots, e_{i,d}$ are the elements on which $v_i(x, y)$ is nonvanishing (in the present paper d is equal to 2 or 4)

$$|v_i|_{1,\Omega}^2 = \sum_{j=1}^{d} |v_i|_{1,e_{i,j}}^2. \tag{53}$$

Clearly, from (10) or (11), the restriction of $v_i(x, y)$ on an element $e_{i,j}$ $(j = 1, \ldots, d)$ is equal to

$$\frac{1}{d} N_q[\xi(x, y), \eta(x, y)],$$

where it has been assumed that the nodal i in the element $e_{i,j}$, $(j = 1, \ldots, d)$, corresponds to the vertex q in the unit square E.

We now find that

$$|v_i|_{1,e_{i,j}}^2 = \frac{1}{d^2} \int_E \Big\{\Big(\frac{\partial N_q}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial N_q}{\partial \eta}\frac{\partial \eta}{\partial x}\Big)^2 + \Big(\frac{\partial N_q}{\partial \xi}\frac{\partial \xi}{\partial y} + \frac{\partial N_q}{\partial \eta}\frac{\partial \eta}{\partial y}\Big)^2\Big\} |J| \, d\xi \, d\eta. \tag{54}$$

Considering that $\partial\xi/\partial x = (\partial y/\partial\eta)/J$, $\partial\xi/\partial y = -(\partial x/\partial\eta)/J$ etc., and using (21), it may be concluded that (54) is equal to $\mathcal{O}(1)$ when h tends to zero. Analogously, again using (21), we find that

$$\|v_i\|^2_{0,e_{i,j}} = \frac{1}{d^2}\int_E N_q(\xi,\eta)^2|J|\,d\xi\,d\eta \geq C_2 h^2$$

when h tends to zero.

From (53), since d is equal to 2 or 4, we obtain (52) and (51).

Lemma 4. *Given* $v = z_1 v_1 + z_2 v_2 + \ldots + z_l v_l$, *with* $z_1, z_2, \ldots, z_l$ *real numbers, if* h *is sufficiently small, then*

(55) $$C_1 h^2\|z\|^2 \leq \|v\|^2_{0,\Omega} \leq C_2 h^2\|z\|^2,$$

where $\|z\|^2 = z_1^2 + z_2^2 + \ldots + z_l^2$.

Proof. Let v^e be the restriction of v on e, and $v^E(\xi,\eta) = v^e(x(\xi,\eta), y(\xi,\eta))$ the polynomial which for (10) or (11) corresponds to v^e. Furthermore, let $P_{r_1}, P_{r_2}, P_{r_3}$ and P_{r_4} be the vertices of e. Then, since $v^e(P_{r_i}) = z_{r_i}$ $(i = 1, 2, 3, 4)$, from (13), we find that

$$v^E(\xi,\eta) = N^T(\xi,\eta)\cdot z^e,$$

where $z^e = [z_{r_1}, z_{r_2}, z_{r_3}, z_{r_4}]^T$ and

(56) $$\|v^e\|^2_{0,e} = \int_E v^E(\xi,\eta)^2|J|\,d\xi\,d\eta = (z^e)^T\int_E N(\xi,\eta)N^T(\xi,\eta)|J|\,d\xi\,d\eta\, z^e.$$

Considering that $\|v^e\|^2_{0,e} \geq 0$ and that it is equal to zero if and only if $v^e \equiv 0$, then the symmetric matrix

$$A^e = \int_E N(\xi,\eta)N^T(\xi,\eta)|J|\,d\xi\,d\eta$$

is positive definite.

Also, since the functions $N_i(\xi,\eta)$ $(i = 1, 2, 3, 4)$ are positive in E, we obtain

$$A^e = \int_E |J|\,d\xi\,d\eta\, B^e,$$

where the symmetric matrix B^e is positive definite.

In [**4**]$_2$ has been shown that, when h is sufficiently small,

$$C_3 h^2 \leqslant \int_E |J|\, d\xi\, d\eta \leqslant C_4 h^2 .$$

Then, since the eigenvalues of the matrix B^e are positive and not dependent on h, we get

$$C_3 h^2 \|z^e\|^2 \leqslant \|v^e\|^2_{0,e} \leqslant C_4 h^2 \|z^e\|^2 .$$

Finally, since a vertex is common to a limited number of elements, we obtain (55).

Lemma 5. *For every function* $v \in S^1_h$ *and for small* h

$$\frac{a(v,v)}{\|v\|^2_{0,\Omega}} \leqslant C_3 h^{-2} . \tag{57}$$

Proof. Given (32), it is true that

$$\frac{a(v,v)}{\|v\|^2_{0,\Omega}} \leqslant C_2 \frac{\|v\|^2_{1,\Omega}}{\|v\|^2_{0,\Omega}} = C_2 \left(\frac{|v|^2_{1,\Omega}}{\|v\|^2_{0,\Omega}} + 1 \right) .$$

From (53), we easily see that

$$|v|^2_{1,\Omega} \leqslant C_4 \|z\|^2 ,$$

where C_4 is independent of h and $v = z_1 v_1 + z_2 v_2 + \ldots + z_l v_l$.

Thus, using Lemma 4, we get

$$\frac{|v|^2_{1,\Omega}}{\|v\|^2_{0,\Omega}} = \frac{|v|^2_{1,\Omega}}{\|z\|^2} \frac{\|z\|^2}{\|v\|^2_{0,\Omega}} \leqslant C_5 h^{-2} ,$$

and, as h tends to zero, the (57).

Let us use μ to denote an eigenvalue of the matrix $S^{-1}A$ and z to denote the associated eigenvector, that is

$$\mu = \frac{z^T A z}{z^T S z} = \frac{z^T A z}{z^T z} \frac{z^T z}{z^T S z} . \tag{58}$$

Let $\{\mu_i\}_{i=1}^l$ be the eigenvalues of $S^{-1}A$ and $0 < \mu_1 \leqslant \mu_2 \leqslant \ldots \leqslant \mu_l$; let $\{\lambda_i^A\}_{i=1}^l$ be the eigenvalues of A and $0 < \lambda_1^A \leqslant \lambda_2^A \leqslant \ldots \leqslant \lambda_l^A$; and let $\{\lambda_i^S\}_{i=1}^l$ be the eigenvalues of S and $0 < \lambda_1^S \leqslant \lambda_2^S \leqslant \ldots \leqslant \lambda_l^S$. Where an eigenvalue of multiplicity $\nu > 1$ is repeated ν times with a different index, then, (58) yields

$$\frac{\lambda_l^A}{\lambda_l^S} \leqslant \mu_l \leqslant \frac{\lambda_l^A}{\lambda_1^S} \quad \text{and} \quad \frac{\lambda_1^A}{\lambda_l^S} \leqslant \mu_1 \leqslant \frac{\lambda_1^A}{\lambda_1^S}.$$

Thus, for the spectral condition number μ_l/μ_1, we may now say that

$$\frac{\lambda_1^S}{\lambda_l^S}\,\frac{\lambda_l^A}{\lambda_1^A} \leqslant \frac{\mu_l}{\mu_1} \leqslant \frac{\lambda_l^A}{\lambda_1^A}\,\frac{\lambda_l^S}{\lambda_1^S}. \tag{59}$$

Since norm $(2\beta v, v)^{\frac{1}{2}}$ and norm $\|\cdot\|_{0,\Omega}$ are equivalent, that is

$$2\beta_0 \|v\|_{0,\Omega}^2 \leqslant (2\beta v, v) \leqslant 2\beta_1 \|v\|_{0,\Omega}^2, \tag{60}$$

then, using Lemma 4, we find that $\lambda_l^S/\lambda_1^S = \mathcal{O}(1)$ as h tends to zero. Besides, this calculations similar to those of Lemma 3 show that the elements $a_{i,j} = a(v_i, v_j)$, $(i, j = 1, 2, \ldots, l)$, of the matrix A are $\mathcal{O}(1)$ as h tends to zero and in every row of A there is a finite number of non-zero elements independently of h; if we now use Gerschgorin's theorem, $\lambda_l^A = \mathcal{O}(1)$ as h tends to zero.

The lowest eigenvalue λ_1^A of A is given by

$$\lambda_1^A = \min_{\|z\|\neq 0} \frac{z^T A z}{\|z\|^2} \min_{\|z\|\neq 0} \frac{a(v, v)}{\|v\|_{0,\Omega}^2}\,\frac{\|{}^1v\|_{0,\Omega}^2}{\|z\|^2} \geqslant \min_{\|v\|_{0,\Omega}\neq 0} \frac{a(v, v)}{\|v\|_{0,\Omega}^2} \min_{\|z\|\neq 0} \frac{\|v\|_{0,\Omega}^2}{\|z\|^2},$$

where $v = z_1 v_1 + z_2 v_2 + \ldots + z_l v_l$.

Then, from (60) and using Lemma 4, we find that $\lambda_1^A \geqslant 2\beta_0 C_1 h^2 \mu_1$. Furthermore, choosing z to be the eigenvector corresponding to μ_1 we obtain $\mu_1 = z^T A z / z^T S z$ and hence $\lambda_1^A \leqslant \mu_1 z^T S z / \|z\|^2$. Then equation (60) and Lemma 4 yield $\lambda_1^A \leqslant 2\beta_1 C_2 h^2 \mu_1$.

We may now state

Theorem 4. *Matrix $S^{-1}A$ has real positive eigenvalues and the spectral condition number is given by*

$$\mu_l/\mu_1 = Ch^{-2}, \tag{61}$$

that is, if h is sufficiently small, the differential system (17) *is stiff.*

Furthermore, if σ is an eigenvalue of the matrix $Q = S + (\Delta t/2)A$ and z the corresponding eigenvector, then $Qz = \sigma z$; if we again suppose that

$v = z_1 v_1 + z_2 v_2 + \ldots + z_l v_l$, then

$$\sigma = \frac{z^T S z}{\|z\|^2} + \frac{\Delta t}{2} \frac{z^T A z}{\|z\|^2} = \frac{(2\beta v, v)}{\|z\|^2} + \frac{\Delta t}{2} \frac{a(v, v)}{\|z\|^2}.$$

Then, using (60) and lemmas 4 and 5, and assuming that $0 < \sigma_1 \leqslant \sigma_2 \leqslant \ldots \leqslant \sigma_l$, we find that $\sigma_l \leqslant 2\beta_1 C_2 h^2 + C_1(\Delta t/2)$, and Lemma 4 gives us $\sigma_1 \geqslant 2\beta_0 C_1 h^2$. If we esclude the uninteresting case $\Delta t h^{-2} \to 0$ then we may conclude by asserting

Theorem 5. *The spectral condition number for the matrix Q is such that*

$$\sigma_l/\sigma_1 = \mathcal{O}(\Delta t h^{-2}). \tag{62}$$

For system (18) to be solved, (62) must not be too great; in practice, its value must be between 1 and 10.

Remark. In Caprili and Lazzeretti [**4**]$_1$ a computer program based on the analysis given in the previous sections has been prepared and used to determine the temperature distribution in a cooled rotor blade.

References

[**1**] A. Agmon, *Lectures on elliptic boundary value problems*, Van Nostrand, Reinhold 1965.

[**2**] T. J. Bond, J. H. Swannell, R. D. Henshell and G. B. Warburton, *A comparision of some curved two-dimensional finite elements*, J. Strain Anal., **8** (1973), 182-190.

[**3**] J. H. Bramble and S. R. Hilbert: [•]$_1$ *Estimation of linear functional on Sobolev space with applications to Fourier transform and spline interpolation*, SIAM J. Anal. **7** (1970), 113-124; [•]$_2$ *Bound for a class of linear functional with applications to Hermite interpolation*, Numer. Math. **16** (1971), 362-369.

[**4**] M. Caprili and R. Lazzeretti: [•]$_1$ *Transient temperature distribution in cooled turbine blades*, Fourth International Symposium on Air Breathing Engines, Florida U.S.A., April 1-6 (1979); [•]$_2$ *Curved quadrilateral element in the finite element method*, AIDAA **3** (1976), 135-139.

[**5**] A. Friedman, *Partial differential equations of parabolic type*, Prentice-Hall 1964.

[6] R. McLeod, *Overcoming loss of accurancy when using curved finite element*, in « The Mathematics of Finite Elements and Applications », II, J. R. Whiteman ed., Academic Press, New York and London 1976.

[7] A. R. Mitchell, *Basic functions for curved elements in the mathematical theory of finite elements*, in « The Mathematics of Finite Elements and Applications », II, J. R. Whiteman ed., Academic Press, New York and London 1976.

[8] J. Nečas, *Les méthodes directes en théorie des équations elliptiques*, Masson, Paris 1967.

[9] G. Strang and G. J. Fix, *An analysis of the finite element method*, Prentice-Hall 1973.

Sommario

Si considera il problema della conduzione di calore attraverso un solido con fori di raffreddamento cilindrici. Si descrive un metodo per la risoluzione approssimata del problema fondato sulla decomposizione del solido in elementi quadrilateri curvi. Si trova una stima dell'errore.

* * *