M2AN. Mathematical modelling and numerical analysis - Modélisation mathématique et analyse numérique

S. Finzi Vita

A numerical approach to a class of unilateral elliptic problems of non-variational type

M2AN. Mathematical modelling and numerical analysis - Modélisation mathématique et analyse numérique, tome 25, nº 2 (1991), p. 253-269

<http://www.numdam.org/item?id=M2AN_1991__25_2_253_0>

MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS
MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE

(Vol. 25, n° 2, 1991, p. 253 à 269)

A NUMERICAL APPROACH TO A CLASS OF UNILATERAL ELLIPTIC PROBLEMS OF NON-VARIATIONAL TYPE (*)

by S. FINZI VITA (1)

Communicated by F. BREZZI

Abstract. — *A particular approach is proposed to solve unilateral problems for elliptic operators in non-divergence form. The solution of such non-variational problems is approximated, via a regularization procedure, by the finite element solutions of certain associated variational inequalities. Convergence results and error estimates in the L^∞-norm are proved for such an approximation in the case of Hölder continuous coefficients in the principal part of the operator.*

Résumé. — *On propose une méthode particulière de solution pour des problèmes unilatéraux associés à des opérateurs elliptiques sous forme non divergentielle. La solution est approchée par les solutions aux éléments finis de certaines inéquations variationnelles obtenues par régularisation. On démontre des résultats de convergence et d'estimation d'erreur dans L^∞ sous l'hypothèse d'Hölderianité des coefficients de la partie principale de l'opérateur.*

1. INTRODUCTION AND NOTATIONS

Problems of non-variational type frequently arise in different fields, such as stochastic control theory (see [2], [8]). In these problems, the differential operator cannot be written in divergence form because of the lack of regularity of its coefficients. For example, the stationary unilateral problem we deal with in this note is connected with the optimal stopping time problem. The solution at a certain point x is interpreted as the infimum, in the class of stopping times t, of a functional which represents the « cost » of following the trajectory of a certain stochastic process starting from

(*) Received in January 1989.
(1) Dipartimento di Matematica, Università di Roma « La Sapienza », Piazzale A. Moro 2, 00185 Roma, Italie.

x and stopping at time t. The governing second-order elliptic operator can be seen as the opposite of the infinitesimal generator of a semigroup, which is well-defined under the only assumption of continuity for the coefficients of the principal part (see [2]).

Existence and regularity results for non-variational problems have been proved in many different cases, such as unilateral or bilateral problems with « regular » or « irregular » obstacles, with elliptic or parabolic, and linear or nonlinear operators (see e.g. [10] and [14] and the references quoted therein).

On the other hand, no numerical results are known to us for this kind of problems. In the absence of a variational formulation, a direct discretization of the problem by a classical finite element method is not the right approach, particularly so if one is interested in proving error estimates for the approximate solutions. The approach we propose here is an indirect one making use of a regularization procedure. In order to prove existence and regularity results for non-variational problems, it is rather natural to approximate the « principal » coefficients of the operator with sequences of more regular, say, differentiable, functions ; then, considering the sequence of variational problems associated to such new coefficients, one has to show that the sequence of their solutions converges, in an appropriate sense, to the solution of the original problem.

In the elliptic unilateral case, as approximate solutions we choose the finite element solutions of certain variational inequalities. This allows us to use the theory of variational inequalities, and the numerical results already known for them. In such a way we are able to prove convergence results and error estimates in the uniform norm when the initial coefficients are Hölder continuous. Analogous results easily follow in the case of equations (for a direct investigation, see [7]).

The implementation of this indirect method presents some interesting problems, such as the influence of numerical integration, or the choice of the regularized coefficients. The study of these aspects is not the aim of this paper. Here we limit ourselves to some remarks and comments, to be found in Section 4.

This note is divided in two parts. In Part I, a rather regular (say, $W^{2,p}$) obstacle function is considered, and we study the strong solutions of the corresponding unilateral problem. In particular, in Section 2 we introduce the problem and recall the relevant existence and regularity results for strong solutions. In Section 3 we introduce the indirect method of discretization, and derive convergence results in the uniform norm for the approximate solutions. Being δ $(0 < \delta < 1)$ the Hölder exponent of the coefficients, we show that the error in the approximation is of order $O(h^{2\tilde{\delta}-\varepsilon})$, $\forall \varepsilon > 0$, where $0 < \tilde{\delta} < \delta$, and $\tilde{\delta}$ grows with δ ($\tilde{\delta} \approx 1$ when δ is close to 1), while h is the discretization parameter. This can be proved for different constructions of the regularizing operators (i.e., of the regularized coefficients), as is shown in Section 4.

In Part II we assume the obstacle function to be only Hölder continuous (with exponent β). Therefore, we need to introduce the notion of generalized solution of the unilateral problem. We collect in Section 5 the corresponding existence and regularity results, and in Section 6 we apply again the indirect method of discretization to derive an error estimate in the uniform norm for the approximate solutions. Comparison of this result with that of Section 3 shows that the rate of convergence in the irregular case is reduced by a factor $\beta/2$, as in the variational case (see e.g. [1], [12] and [6]).

For convenience, we list hereafter all the function spaces to be used in the sequel, with the notation adopted for their standard norms.

Let Ω be an open bounded domain in $\mathbb{R}^N$, $N \geq 2$, with sufficiently smooth boundary Γ. We consider the following spaces of functions defined over Ω :

— $C^k(\bar{\Omega})$, $k \in \mathbb{N} \cup \{0\}$ (continuously differentiable functions ; $|||\,.\,|||_k$)
— $C_0^k(\Omega)$ (functions of $C^k(\Omega)$ with compact support in Ω)
— $C^{k,\delta}(\bar{\Omega})$, $0 < \delta < 1$ $[\delta = 1]$ (Hölder [Lipschitz] continuous functions ; $[\cdot]_{k,\delta}$)
— $L^p(\Omega)$, $1 \leq p < +\infty$ (p-integrable functions ; $\|\cdot\|_p$)
— $L^\infty(\Omega)$ (essentially bounded functions ; $\|\cdot\|_\infty$)
— $W^{k,p}(\Omega)$, $k \in \mathbb{N}$, $1 \leq p \leq \infty$ (Sobolev spaces ; $\|\cdot\|_{k,p}$) ; $H^k(\Omega) \equiv W^{k,2}(\Omega)$
— $H_0^1(\Omega)$ (completion of $C_0^1(\Omega)$ in the norm of H^1).

PART I. STRONG SOLUTIONS

2. FORMULATION OF THE PROBLEM. EXISTENCE AND REGULARITY RESULTS

Let us consider the linear second order operator

$$Lv := -\sum_{ij} a_{ij}(x)\,\partial_{ij}v + \sum_i b_i(x)\,\partial_i v + c_0(x)\,v\,, \tag{2.1}$$

(where we have set : $\partial_i v = (\partial v/\partial x_i)$, $\partial_{ij}v = (\partial^2 v/\partial x_i\,\partial x_j)$) ; we assume :

$$\exists \nu > 0 : \quad \sum_{ij} a_{ij}(\cdot)\,\xi_i\,\xi_j \geq \nu\,|\xi|^2, \quad \text{a.e. in } \Omega, \quad \forall \xi \in \mathbb{R}^N\,; \tag{2.2}$$

$$a_{ij} = a_{ji}, \quad a_{ij} \in C^0(\bar{\Omega}), \quad i,j = 1, \dots, N\,; \tag{2.3}$$

$$b_i, c_0 \in L^\infty(\Omega), \quad i = 1, \dots, N\,; \quad c_0(x) \geq k_0 > 0\,. \tag{2.4}$$

Given two functions :

$$f \in L^p(\Omega), \quad 1 < p \leqslant \infty \tag{2.5}$$

$$\psi \in W^{2,p}(\Omega), \psi|_\Gamma \geqslant 0, \quad 1 < p \leqslant \infty\,, \tag{2.6}$$

the unilateral problem we are interested in has the form of the following complementarity system :

$$(P) \qquad \begin{cases} \textit{Find a function } u \textit{ defined over } \Omega \textit{ such that} : \\ u \leqslant \psi,\ Lu \leqslant f,\ \ (u-\psi)(Lu-f) = 0 \quad \text{in } \Omega \\ u|_\Gamma = 0\,. \end{cases}$$

In order to give sense to the formal relations in (P), we introduce the following

DEFINITION 2.1 : *A* strong solution *of the unilateral problem* (P) *is any function* $u \in H_0^1(\Omega) \cap W^{2,p}(\Omega)$ *which solves* (P).

We remark that the a_{ij}'s are not sufficiently smooth to write the operator L in divergence form. For such a reason, problem (P) is not equivalent to a variational inequality ; nevertheless, the following result is known :

THEOREM 2.1 : *Assume* (2.1)-(2.6) *; then there exists a unique (strong) solution* u *of* (P), *continuous if the obstacle* ψ *is continuous, which satisfies the dual inequality*

$$Lu \geqslant L\psi \wedge f\,, \quad \text{a.e. } in\ \Omega\,. \tag{2.7}$$

Proof: See [14]. The basic idea is to replace in L the a_{ij}'s by some differentiable functions a_{ij}^n whose sequences satisfy :

i) $a_{ij}^n \to a_{ij}$ in $C^0(\bar{\Omega})$ when $n \to +\infty$, $\forall i, j = 1, \dots, N$;

ii) (2.2) holds $\forall n$.

In such a way one produces a sequence of « regularized » operators

$$L^n := -\sum_{ij} a_{ij}^n(x)\, \partial_{ij} + \sum_i b_i(x)\, \partial_i + c_0(x) \tag{2.8}$$

which can be written in divergence form. More precisely, if we denote by u^n the solution of the complementarity system

$$(P_n) \qquad \begin{cases} u^n \in H_0^1(\Omega) \cap W^{2,p}(\Omega) \\ u^n \leqslant \psi,\ L^n u^n \leqslant f,\ (u^n - \psi)(L^n u^n - f) = 0 \quad \text{in } \Omega\,, \end{cases}$$

then u^n is also the unique solution of the variational inequality

$$(Q_n) \qquad u^n \in K\,, \quad a_n(u^n, v - u^n) \geqslant (f, v - u^n)\,, \quad \forall v \in K\,,$$

where $K = \{v \in H_0^1(\Omega) : v \leqslant \psi\}$ and

$$(2.9)\quad a_n(v, w) := \int_\Omega \left[\sum_{ij} a_{ij}^n(\partial_i v)(\partial_j w) + \sum_i b_i^n(\partial_i v)\, w + c_0\, v w\right] dx$$

denotes the bilinear, continuous, not symmetric form on $H_0^1 \times H_0^1$ with

$$(2.10)\qquad b_i^n = b_i + \sum_j \partial_j a_{ij}^n\,,\quad i = 1, \dots, N\,.$$

In [14] is proved that the sequence of functions u^n, satisfying

$$(2.11)\qquad L^n u^n \geqslant L^n \psi \wedge f\,,$$

converges weakly in $W^{2,p}(\Omega)$ to a solution u of (P). Furthermore, such solution is unique, and $u^n \to u$ strongly in $L^\infty(\Omega)$. □

Remark 2.1 : The equivalence between problems (P_n) and (Q_n) is proven in [14] using the notion of subsolution together with regularity arguments.

Suppose now (in addition to i) and ii)) that the new coefficients a_{ij}^n in (2.8) satisfy, $\forall i, j = 1, \dots, N$,

$$(2.12)\qquad \left\| a_{ij}^n - a_{ij} \right\|_\infty \leqslant c n^{-\delta}\,,\quad \delta \in (0, 1)\,,$$

$$(2.13)\qquad \left\| a_{ij}^n \right\|_{1,\infty} \leqslant c n^{1-\gamma}\,,\quad \gamma < 1\,,$$

where c, here as in the sequel, denotes different constants independent of n. Then, the result of Theorem 2.1 can be strenghtened :

THEOREM 2.2 : *Assume* (2.1)-(2.6), *and that the coefficients* a_{ij}^n *of* (2.8) *have been constructed so as to satisfy* (2.12) *and* (2.13) ; *then, for sufficiently large* p,

$$(2.14)\qquad \| u^n - u \|_\infty \leqslant c n^{-\delta+\varepsilon} \left(\| f \|_p + \| L\psi \|_p \right), \quad \forall \varepsilon > 2(N+1)/(p+N+1)\,.$$

Proof : It is enough to repeat, in the elliptic unilateral case, the proof of [9] (Theorem 2.1). For later use, we remark that the constant c in (2.14) is independent of ψ. □

COROLLARY 2.1 : *Assume* $p = +\infty$ *in* (2.5), (2.6) *of Theorem* 2.2 ; *then,* $u \in W^{2,q}(\Omega)$ $\forall q < +\infty$, *and* (2.14) *holds* $\forall \varepsilon > 0$.

Remark 2.2 : A natural assumption on the a_{ij}'s which allows to satisfy relations (2.12) and (2.13) (with $\gamma = \delta$) is the following :

$$(2.15)\qquad a_{ij} \in C^{0,\delta}(\bar{\Omega}),\quad \delta \in (0, 1)\,.$$

In Section 4 we will give some examples of sequences a_{ij}^n constructed under this assumption on the a_{ij}'s.

3. DISCRETIZATION AND MAIN RESULT

Let us suppose Ω to be a convex set. We denote by Ω_h a polyhedral domain inscribed in Ω such that the diameter of each « face » does not exceed a positive constant h. For $h \to 0$, we consider over Ω_h a family of « triangulations » T_h, i.e. subdivisions of Ω_h in N-dimensional simplexes τ, such that

$$(3.1)\qquad \begin{cases} \text{i) } \bar{\Omega}_h = \bigcup_{\tau \in T_h} \tau \,; \quad h = \max_{\tau \in T_h} \operatorname{diam}(\tau) \\ \text{ii) } \operatorname{dist}(\Gamma, \partial\Omega_h) \leq ch^2 \\ \text{iii) } \exists \tilde{c}_1, \tilde{c}_2 > 0, \text{ independent of } h, \text{ such that each} \\ \quad \tau \in T_h \text{ contains a ball with radius } \tilde{c}_1 h \\ \quad \text{and is contained in a ball with radius } \tilde{c}_2 h\,. \end{cases}$$

As finite element space, we will consider the subspace of $H_0^1(\Omega)$ defined by

$$(3.2)\qquad V_h = \left\{ v \in C^0(\bar{\Omega}) : v|_\tau \in P_1, \quad \forall \tau \in T_h\,; \quad v \equiv 0 \text{ on } \bar{\Omega} - \Omega_h \right\},$$

where P_1 is the space of first order polynomials in $\mathbb{R}^N$.

We indicate with $\{p_i\}$ $(i = 1, \ldots, s)$ the internal nodes of T_h and with $\{p_i\}$ $(i = s+1, \ldots, \tilde{s})$ its boundary nodes. Correspondingly, let $\{\varphi_i\}$ $(i = 1, \ldots, \tilde{s})$ be continuous piecewise linear functions on Ω such that $\varphi_i(p_j) = \delta_{ij}$. The functions $\{\varphi_i\}$ $(i = 1, \ldots, s)$ form the canonical basis of the space V_h; the interpolation operator $r_h : C^0(\bar{\Omega}) \cap H_0^1(\Omega) \to V_h$ is classically given by

$$(3.3)\qquad r_h[v](x) := \sum_i v(p_i)\, \varphi_i(x)\,.$$

With these notations, the discrete problem associated to (Q_n) can be formulated as

$$(Q_{n,h})\qquad u_h^n \in K_h, \quad a_n(u_h^n, v_h - u_h^n) \geq (f, v_h - u_h^n), \quad \forall v_h \in K_h\,,$$

where

$$K_h = \{ v_h \in V_h : v_h(p_i) \leq \psi(p_i), \quad i = 1, \ldots, s \}\,.$$

At this stage, a crucial assumption is needed, namely,

(3.4) $\exists \mu > 0$ *independent of h such that :*

$$A_{kl} := \sum_{T} \int_{T} \sum_{ij} a_{ij} \, \partial_i \varphi_k \, \partial_j \varphi_l \, dx \leqslant -\mu h^{N-2},$$

for any $k, l = 1, \dots, \tilde{s}$, *with* $k \neq l$ *and* supp $(\varphi_k) \cap$ supp $(\varphi_l) \neq \varnothing$.

This assumption, which is enough for a discrete maximum principle to hold in problems $(Q_{n,h})$, give some restrictions on the differential operators to be considered and on the admissible amplitude of the angles in the triangulations (for more details see [4] and the next section).

Our main result is the following :

THEOREM 3.1 : *Assume* (2.1)-(2.6) *(with* $p = +\infty$), (2.15), (3.1) *and* (3.4). *For any (sufficiently small) h, there exists an integer* $\tilde{n}$ *depending on h such that, if* $\tilde{u}_h$ *denotes the solution of problem* $(Q_{\tilde{n},h})$, *then*

$$\|u - \tilde{u}_h\|_{\infty} \leqslant c h^{2\tilde{\delta} - \varepsilon} |\log h|^2, \quad \forall \varepsilon > 0, \tag{3.5}$$

where u is the solution of the unilateral problem (P), c is a constant independent of h, and $\tilde{\delta}$ *satisfies*

$$0 < \tilde{\delta} < \delta, \quad with \quad \tilde{\delta} \approx 1 \quad when \quad \delta \approx 1 . \tag{3.6}$$

To prove Theorem 3.1 we need some preliminary results. First of all, we want to study the convergence (in $L^{\infty}(\Omega)$ and for a fixed n) of the finite element solutions u_h^n of $(Q_{n,h})$ to the solution u^n of (Q_n). Let us begin with :

LEMMA 3.1 : *Assume* (2.1)-(2.4), (2.15), (3.1) *and* (3.4). *Then, for n large enough, the regularized coefficients* a_{ij}^n *of* (2.8) *can be constructed such that, for a sufficiently small step h,*

(3.7) $\exists \tilde{\mu} > 0 \quad (\tilde{\mu} < \mu)$ *such that :*

$$A_{kl}^n := a_n(\varphi_l, \varphi_k) \leqslant -\tilde{\mu} h^{N-2},$$

for any $k, l = 1, \dots, \tilde{s}$, *with* $k \neq l$ *and* supp $(\varphi_k) \cap$ supp $(\varphi_l) \neq \varnothing$;

$$\sum_{l} A_{kl}^n > 0, \quad k = 1, \dots, \tilde{s} . \tag{3.8}$$

Remark 3.1 : Inequalities (3.7) and (3.8) imply that the stiffness matrix (A_{kl}^n) is a M-matrix (since it is strictly diagonally dominant and $A_{kk}^n > 0 \ \forall k$, see e.g. [11]).

Proof : From (2.9) :

$$A_{kl}^n = \sum_\tau \int_\tau \left[\sum_{ij} a_{ij}^n \, \partial_i \varphi_k \, \partial_j \varphi_l + \sum_i b_i^n \, \partial_i \varphi_k \, \varphi_l + c_0 \, \varphi_k \, \varphi_l \right] dx \, ;$$

then, from assumption (3.1),

$$\begin{aligned} A_{kl}^n &\leq A_{kl} + \max_{ij} \; \| a_{ij}^n - a_{ij} \|_\infty \sum_\tau \int_\tau \sum_{ij} | \partial_i \varphi_k \, \partial_j \varphi_l | \, dx \\ &\quad + \max_i \; \| b_i^n \|_\infty \sum_\tau \int_\tau \sum_i | \partial_i \varphi_k \, \varphi_l | \, dx + \| c_0 \|_\infty \int_\Omega \varphi_k \, \varphi_l \, dx \\ &\leq (-\mu + cn^{-\delta}) \, h^{N-2} + cn^{1-\delta} h^{N-1} + ch^N , \end{aligned}$$

since (2.15) implies (2.12), (2.13) [with $\gamma = \delta$] ; if n is large enough, (3.7) then follows by taking a sufficiently small h.

Condition (3.8) is easily verified since, if $v_h = \sum_l V_l \, \varphi_l$, then

$$a_n(v_h, \varphi_k) = \sum_l A_{kl}^n \, V_l \, ,$$

and, if $v_h \equiv 1$,

$$\sum_l A_{kl}^n = a_n(1, \varphi_k) = \int_\Omega c_0 \, \varphi_k \, dx > 0 \text{ (from (2.4))} . \quad \square$$

THEOREM 3.2 : *Assume in* (Q_n) $a_{ij}^n \in C^{0,1}(\bar{\Omega})$, *together with* (2.2)-(2.6) [$p = +\infty$], (3.1), (3.7)-(3.8). *Then, for a sufficiently small h,*

$$\| u^n - u_h^n \|_\infty \leq k_n \, h^2 |\log h|^2 , \tag{3.9}$$

where k_n *is a constant, independent of h, which depends on n through the norms* $\| a_{ij}^n \|_{1,\infty}$, $\| b_i^n \|_\infty$.

Proof : We are in a position to use the results of [5]. The idea is the following : since the form (2.9) is not coercive (see Remark 2.1), we choose a positive and sufficiently large constant λ^n such that the new bilinear form

$$b_n(v, w) := a_n(v, w) + \lambda^n \int_\Omega vw \, dx$$

is coercive, i.e. :

$$b_n(v, v) \geq \alpha_0 (\| v \|_{1,2})^2 , \quad \forall v \in H_0^1(\Omega) , \tag{3.10}$$

with α_0 a given positive constant ($\alpha_0 < \nu$). If we take, for example,

$$\lambda^n = \left(\max_i \|b_i^n\|_\infty\right)^2 / 4(\nu - \alpha_0)\,, \tag{3.11}$$

the inequality (3.10) is satisfied. Then, problems (Q_n) and $(Q_{n,h})$ can be transformed into the equivalent coercive problems of implicit type :

$$(Q_n)' \quad u^n \in K\,, \quad b_n(u^n, v - u^n) \geq (f + \lambda^n u^n, v - u^n), \quad \forall v \in K\,,$$
$$(Q_{n,h})'' \quad u_h^n \in K_h\,, \quad b_n(u_h^n, v_h - u_h^n) \geq (f + \lambda^n u_h^n, v_h - u_h^n), \quad \forall v_h \in K_h\,.$$

If we introduce now the two auxiliary problems

$$(Q_{n,h})' \quad w_h^n \in K_h\,, \quad b_n(w_h^n, v_h - w_h^n) \geq (f + \lambda^n u^n, v_h - w_h^n), \quad \forall v_h \in K_h\,,$$
$$(Q_{n(h)})'' \quad w^{n(h)} \in K\,, \quad b_n(w^{n(h)}, v - w^{n(h)}) \geq (f + \lambda^n u_h^n, v - w^{n(h)}), \; \forall v \in K\,,$$

it is easy to get, following the proof in [5], the inequality :

$$\|u^n - u_h^n\|_\infty \leq (1 + \lambda^n) \max\left(\|u^n - w_h^n\|_\infty, \|w^{n(h)} - u_h^n\|_\infty\right); \tag{3.12}$$

since $(Q_n)' - (Q_{n,h})'$ and $(Q_{n(h)})'' - (Q_{n,h})''$ represent two pairs of coercive continuous and discrete variational inequalities, we recover the situation typical of the coercive case. The basic tool of the proof is a discrete maximum principle, which holds if the stiffness matrix $B^n = (B_{kl}^n) := (b_n(\varphi_l, \varphi_k))$ is a M-matrix. This is the case, since (using (3.7), (3.8) and the strong coercivity of $b_n(\,.\,,\,.\,)$), we have :

i) $B_{kl}^n = A_{kl}^n + \lambda^n \displaystyle\int_\Omega \varphi_k \varphi_l \, dx \leq -\tilde{\mu} h^{N-2} + c\lambda^n h^N \leq 0$

for any $k, l = 1, \ldots, \tilde{s}$, with $k \neq l$ and $\operatorname{supp}(\varphi_k) \cap \operatorname{supp}(\varphi_l) \neq \varnothing$ (and h sufficiently small) ;

ii) $\sum_l B_{kl}^n \geq \sum_l A_{kl}^n > 0, \quad \forall k = 1, \ldots, \tilde{s}$;

iii) B^n is positive definite.

As to the coercive case, it can be proved that

$$\|u^n - w_h^n\|_\infty \leq c_n h^2 |\log h|^2 \left(\|L^n u^n\|_\infty + \|u^n\|_\infty\right), \tag{3.13}$$

(a similar estimate holds for the second norm in the right-hand side of (3.12)). In order to prove (3.13) it is sufficient to extend some known results on the finite element approximation of coercive variational inequalities, keeping track of the relevant constants. If we denote by $R_h u^n$ the Ritz projection of u^n associated to the form $b_n(\,.\,,\,.\,)$, i.e., the solution of

$$R_h u^n \in V_h\,, \quad b_n(R_h u^n - u^n, v_h) = 0, \quad \forall v_h \in V_h\,,$$

then (3.13) is a consequence of the three inequalities :

$$(3.14) \qquad \|u^n - w_h^n\|_\infty \leq c\,\|u^n - R_h u^n\|_\infty \quad (c \text{ indep. of } h \text{ and } n\,)$$

$$(3.15) \qquad \|u^n - R_h u^n\|_\infty \leq c_{1,n}\, h^{2-N/p} |\log h|\, \|u^n\|_{2,p}, \quad \forall p < +\infty\,,$$

$$(3.16) \qquad \|u^n\|_{2,p} \leq c_{2,n}\, p(\|L^n u^n\|_p + \|u^n\|_p), \quad \forall p < +\infty\,,$$

where $c_{1,n} \approx \lambda^n$, and $c_{2,n}$ depends on n only through the modulus of continuity $\omega(a_{ij}^n)$ of the coefficients a_{ij}^n. (The general approach leading to inequalities (3.14) and (3.15) can be found in [13], together with a complete reference list ; for (3.16), see e.g. [3]).

It is enough to take $p = \log(1/h)$ in (3.16), and to combine it with the two previous inequalities, to get (3.13), and then the thesis (from (3.12)), with

$$(3.17) \qquad k_n = c(\omega(a_{ij}^n))\, \lambda^n (1 + \lambda^n)(\|L^n u^n\|_\infty + \|u^n\|_\infty)\,. \quad \square$$

COROLLARY 3.1 : *In the assumptions of Theorem* 3.2, *let us suppose that, for* $n \to +\infty$, *the coefficients* a_{ij}^n *converge to some functions* a_{ij} *in the way prescribed by* (2.12), (2.13). *Then, the constant* k_n *in* (3.9) *will diverge with rate*

$$(3.18) \qquad k_n \approx n^{4(1-\gamma)}\,.$$

Proof: From the dual inequality (2.11) and (2.14) we get $\|L^n u^n\|_\infty + \|u^n\|_\infty \leq c$, uniformly in n ; moreover, from (2.12), there exists a constant c independent of n such that $\omega(a_{ij}^n) \leq c$, $\forall i, j = 1, \ldots, N$; the thesis then follows from (3.17), (3.11), (2.10) and (2.13). $\square$

We are now in a position to prove Theorem 3.1.

Proof of Theorem 3.1 : Lemma 3.1 implies that, if h is small enough, conditions (3.7), (3.8) are satisfied for any sufficiently large n. Therefore, Theorem 3.2 holds true together with Corollary 3.1 (with $\gamma = \delta$). Then, for any pair of such n and h, we have (from (2.14), (3.9), (3.18)) :

$$(3.19) \qquad \begin{aligned} \|u - u_h^n\|_\infty &\leq \|u - u^n\|_\infty + \|u^n - u_h^n\|_\infty \\ &\leq c n^{-\delta+\varepsilon} + c n^{4(1-\delta)} h^2 |\log h|^2, \quad \forall \varepsilon > 0\,. \end{aligned}$$

It is now sufficient to choose n as a function of h in such a way as to balance the terms in the right-hand side of (3.19). More precisely, if we consider a function $\tilde{n}(h) : (0, 1) \to \mathbb{N}$ implicitly defined by the inequalities

$$(3.20) \qquad h^{-2/(4-3\delta)} \leq \tilde{n}(h) \leq 1 + h^{-2/(4-3\delta)}\,,$$

then (3.5) follows, with

$$\tilde{\delta} = \delta/(4 - 3\,\delta)\,. \quad \square \tag{3.21}$$

Remark 3.2 : The estimate (3.5) is not optimal, since the interpolation error for a function in $W^{2,p}(\Omega)$ $\forall p < +\infty$ is a $O(h^{2-\varepsilon})$, $\forall \varepsilon > 0$. It has to be said that the exponent $\tilde{\delta}$ given by (3.21) does not seem to be the best possible. However, if $\delta = 1$ (i.e., in the particular case of Lipschitz continuous coefficients), our result coincides with that of [1], since the operator L can be written in divergence form, and we fall down into the variational case.

4. REMARKS ON NUMERICAL INTEGRATION

The results of the previous section give a theoretical justification to the indirect method of approximation introduced for problem (P). However, this method is not completely satisfactory to handle in practice. Here we just wish to point out these practical difficulties, since the way of overcoming them looks like an interesting problem by itself.

a) The discrete maximum principle

A large use has been made in Section 3 of the discrete maximum principle in order to prove convergence and error estimate results. As we already said, condition (3.4) puts severe limitations on the type of differential operators which can be considered. For example, it is known ([4]) that, for $a_{ij} = \delta_{ij}$ (i.e., $L = -\Delta$) and $N = 2$, all the angles θ in the triangulations T_h cannot exceed $\pi/2 - \varepsilon$ in order to verify (3.4) (ε being a fixed positive constant independent of h). Moreover, for general constant coefficients, it is necessary to take $\theta < \pi/2 - \eta$, where η is an angle which grows with the absolute value of the ratio between the maximum and the minimum eigenvalue of the matrix $A = (a_{ij})$. This means that, if A is not well-conditioned, it becomes impossible to construct a family of triangulations for which (3.4) is satisfied.

A fortiori, this is true for variable coefficients and N whatever. Not only that, if μ is very close to zero, we are forced to choose an extremely small h if we want to verify (3.7) and the analogous condition for B^n needed in Theorem 3.2.

b) The construction of the a_{ij}^n

In the literature (see e.g. [10], [14]), the usual choice consists in constructing the new coefficients by convolution with the initial ones :

$$a_{ij}^n(x) = n^N \int_{\mathbb{R}^N} \rho(n(x-y))\, a_{ij}(y)\, dy, \quad i,j = 1, \ldots, N\,, \tag{4.1}$$

where $\rho \in C_0^\infty(\mathbb{R}^N)$ is a given function such that $\rho \geq 0$, $\rho(x) = 0$ when $|x| \geq 1$, and $\int \rho \, dx = 1$. If the a_{ij}'s satisfy (2.15), then the functions in (4.1) verify (2.12) and (2.13) [10].

From a numerical point of view this choice does not seem the most convenient one. Convolution integrals are not easy to handle, whereas C^∞ regularity for the a_{ij}^n is, in a certain sense, much more than what is needed. Indeed, all the existence and regularity results we recalled in Section 2 hold under the minimal hypotheses (such as Lipschitz continuity) allowing to write the operator L^n in divergence form.

For that reason, let us consider a different approach. With the notations of Section 3, for a fixed $n \in \mathbb{N}$, we introduce a polyhedral domain $\Omega_{1/n}$, a regular triangulation $T_{1/n}$ of « size » $1/n$ and internal nodes $\{q_k\}$, a finite element space $V_{1/n}$ defined as in (3.2), whose basis functions we now indicate by w_k. Then, following definition (3.3), we set

$$a_{ij}^n(x) = r_{1/n}[a_{ij}](x) = \sum_k a_{ij}(q_k)\, w_k(x), \quad i,j = 1, \ldots, N . \tag{4.2}$$

It is then easy to prove :

THEOREM 4.1 : *Let $v \in C^{0,\delta}(\bar{\Omega})$, $\delta \in (0,1)$; then $v^n = r_{1/n}[v]$ satisfies :*

$$\| v - v^n \|_\infty \leq c n^{-\delta} [v]_{0,\delta} , \tag{4.3}$$

$$\| v^n \|_{1,\infty} \leq c n^{1-\delta} [v]_{0,\delta} . \tag{4.4}$$

Proof : Estimate (4.3) is a maximum norm interpolation result already known for Hölder continuous functions (see e.g. [6]).

On the other hand, (4.4) can be derived by interpolation from the two easy estimates

i) $\| v^n \|_{1,\infty} \leq cn \, |\!|\!| v |\!|\!|_0$ (by the inverse inequality) ;

ii) $\| v^n \|_{1,\infty} \leq c \| v \|_{1,\infty}$ (by the uniform Lipschitz continuity of the interpolating functions of a Lipschitz continuous function). □

For a fixed n, $a_{ij}^n \in W^{1,\infty}(\Omega)$; all the results of Section 3 remain true, while the computation of the terms of the stiffness matrix reduces to a sum of straightforward integrals.

For the sake of completeness, we mention a third admissible choice for the a_{ij}^n, which is a combination of the previous two. In order to achieve a strong regularity of the coefficients of L^n without loosing « too much » in the calculation, we set :

$$a_{ij}^n(x) = n^N \int_{\mathbb{R}^N} \rho(n(x-y))\, r_{1/n}[a_{ij}](y)\, dy, \quad i,j = 1, \ldots, N , \tag{4.5}$$

that is, we regularize by convolution the interpolating functions of the initial coefficients. Then, it is easy to prove :

THEOREM 4.2 : *Let* (2.15) *hold ; then the functions defined in* (4.5) *have the following properties :*

i) $a_{ij}^n \in C_0^\infty(\mathbb{R}^N)$;

ii) (2.12) *and* (2.13) *hold* [*with* $\gamma = \delta$] ;

iii) $a_{ij}^n(x) = \sum_k a_{ij}(q_k)\, w_k^n(x)$, $\quad \partial_l a_{ij}^n(x) = \sum_k a_{ij}(q_k)\, \partial_l w_k^n(x)$,

$(i, j, l = 1, \ldots, N)$, *where* $w_k^n(x) = n^N \int_{\mathbb{R}^N} \rho(n(x-y))\, w_k(y)\, dy$.

In other words, the construction of the new coefficients, and hence, of the stiffness matrix, reduces to computing, once and for all the convolution integrals of the basis functions and of their derivatives.

Unfortunately, the connection between n and h needed for the validity of Theorem 3.1 in general implies that, from definitions (4.2) and (4.5), the a_{ij}'s must be interpolated on a grid which is much finer than the one inducted by T_h, and which is not in a simple relation with it. This contrasts with the usual ideas of numerical integration.

PART II. GENERALIZED SOLUTIONS

5. FORMULATION OF THE PROBLEM. EXISTENCE AND REGULARITY RESULTS

Let us assume now, in problem (P) :

(5.1) $$\psi \in C^0(\bar{\Omega}), \quad \psi|_\Gamma \geq 0 ,$$

instead of (2.6) : the existence of strong solutions is no more guaranteed, and we need define solutions of (P) in a weaker sense.

Let us denote by $S(\psi, f)$ the set of all subsolutions of (P), i.e., the set of functions $w \in H^1(\Omega)$ such that

(5.2) $$\begin{cases} w \leq \psi & \text{a.e. in } \Omega \\ Lw \leq f & \text{in the sense of } H^{-1}(\Omega) \\ w \leq 0 & \text{on } \Gamma . \end{cases}$$

DÉFINITION 5.1 : *A* generalized solution *of the unilateral problem* (P) *is the function*

$$u = \sup S(\psi, f) .$$

For the sake of conciseness, we shall indicate in the sequel by $\sigma(A, \varphi)$ the solution of a unilateral problem of type (P) with operator

A and obstacle φ (in a either strong or generalized sense, according to the smoothness of φ).

The following results are known :

THEOREM 5.1 : *Assume* (2.1)-(2.5) (*with* $p > N/2$), (5.1). *Then, there exists a generalized solution* $u = \sigma(L, \psi)$ *of* (P) (*of course unique*), *such that* $u \in C^0(\bar{\Omega}) \cap H^2(\Omega')$, $\forall \Omega' \subset\subset D_\psi = \{x \in \bar{\Omega} : u(x) < \psi(x)\}$.

THEOREM 5.2 : *If assumptions* (2.3) *and* (5.1) *in Theorem* 5.1 *are replaced respectively by* (2.15) *and*

$$\psi \in C^{0,\beta}(\bar{\Omega}),\ \beta \in (0, 1)\,, \tag{5.3}$$

and, moreover, p *in* (2.5) *is taken large enough, then* $u \in C^{0,\beta'}(\bar{\Omega})$, *with* $\beta' = \delta \wedge \beta$.

Theorem 5.1 is proved in [14]; Theorem 5.2 is proved in [10] for the bilateral evolution problem. In both cases, the obstacle ψ is approximated by a sequence of smooth functions $\{\psi^m\}$. For the sequence of strong solutions $u^m = \sigma(L, \psi^m)$, it is possible to prove that

$$u^m \to u \quad \text{uniformly in } L^\infty(\Omega)\,, \tag{5.4}$$

$$\|u^m - u\|_\infty \leq \|\psi^m - \psi\|_\infty\,, \tag{5.5}$$

where $u = \sigma(L, \psi)$ is the generalized solution of (P). In such a way, the case studied in the Section 3 is recovered, since, for every fixed m, $u^m := \lim u^{m,n}$ (for $n \to +\infty$), where $u^{m,n} = \sigma(L^n, \psi^m)$ are the solutions of the variational inequalities with obstacle ψ^m for the regularized operators L^n.

A sharp estimate of the convergence in (5.4) is achieved in [10] via the following lemma.

LEMMA 5.1 : *Let* ψ *be a function satisfying* (5.3). *Then, a sequence* $\{\psi^m\}$ *can be constructed (e.g., by convolution) in order to have*

$$\psi^m \in C^2(\bar{\Omega})\,, \tag{5.6}$$

$$\|\psi^m + \psi\|_\infty \leq c m^{-\beta}[\psi]_{0,\beta}\,, \tag{5.7}$$

$$|||\psi^m|||_2 \leq c m^{2-\beta}[\psi]_{0,\beta}\,, \tag{5.8}$$

where the constants c *are independent of* m.

From (5.5) and (5.7) we get now the estimate :

$$\|u - u^m\|_\infty \leq c m^{-\beta}[\psi]_{0,\beta}\,, \tag{5.9}$$

with c independent of m. Estimate (5.9) allows to evaluate the rate of convergence in n and m of the strong solutions $u^{m,n}$ for the regularized variational problems to the generalized solution u for the non-variational problem (P). Indeed, from (5.9), (2.14), (5.8), we have that

$$\begin{aligned} \|u - u^{m,n}\|_\infty &\le \|u - u^m\|_\infty + \|u^m - u^{m,n}\|_\infty \\ &\le cm^{-\beta} + cn^{-\delta+\varepsilon}\|L\psi^m\|_p \\ &\le cm^{-\beta} + cn^{-\delta+\varepsilon} m^{2-\beta} ; \end{aligned} \tag{5.10}$$

(notice that we have denoted by c different constants independent of m and n).

6. ERROR ESTIMATE IN THE CASE OF GENERALIZED SOLUTION

With the notation of Section 3, let $u_h^{m,n} = \sigma_h(L^n, \psi^m)$ be the solution of the discrete variational inequality

$$u_h^{m,n} \in K_h^m , \quad a_n(u_h^{m,n}, v_h - u_h^{m,n}) \geq (f, v_h - u_h^{m,n}), \ \forall v_h \in K_h^m , \tag{6.1}$$

where $K_h^m := \{ v_h \in V_h : v_h(p_i) \le \psi^m(p_i), i = 1, \dots, s \}$.

THEOREM 6.1 : *Let us consider* $u^m = \sigma(L, \psi^m)$ *in the assumptions* (2.1)-(2.5) (*with* $p = +\infty$), (2.15), (5.3), *together with* (3.1) *and* (3.4) *for the discretization. Then, for any sufficiently small* h, *there exists an integer* $\tilde{n}$ *depending on* h *such that*

$$\left\| u^m - u_h^{m,\tilde{n}} \right\|_\infty \le ch^{2\tilde{\delta}-\varepsilon}|\log h|^2(\|f\|_\infty + \|\psi^m\|_{2,\infty}), \quad \forall \varepsilon > 0 , \tag{6.2}$$

where c *is a constant independent of* h *and* m, *while* $\tilde{\delta}$ *is the constant given in Theorem* 3.1.

Proof: We proceed as in the proof of Theorem 3.1. For every $n \in \mathbb{N}$,

$$\left\|u^m - u_h^{m,n}\right\|_\infty \le \left\|u^m - u^{m,n}\right\|_\infty + \left\|u^{m,n} + u_h^{m,n}\right\|_\infty . \tag{6.3}$$

Now, in view of Theorem 2.2 :

$$\begin{aligned} \|u^m - u^{m,n}\|_\infty &\le cn^{-\delta+\varepsilon}(\|f\|_\infty + \|L\psi^m\|_\infty) \\ &\le cn^{-\delta+\varepsilon}(\|f\|_\infty + \|\psi^m\|_{2,\infty}) . \end{aligned} \tag{6.4}$$

Moreover, from Theorem 3.2 and Corollary 3.1 we get

$$\begin{aligned} \left\|u^{m,n} - u_h^{m,n}\right\|_\infty &\le cn^{4(1-\delta)}h^2|\log h|^2 (\|L^n u^{m,n}\|_\infty + \|u^{m,n}\|_\infty) \\ &\le cn^{4(1-\delta)}h^2|\log h|^2 (\|f\|_\infty + \|\psi^m\|_{2,\infty}) ; \end{aligned} \tag{6.5}$$

(again we have made use of a dual inequality such as (2.11), and of the regularity assumptions). If, as was done in Theorem 3.1, we define the function $\tilde{n}(h)$ by the inequalities (3.20), the thesis follows from (6.3), (6.4) and (6.5). □

Finally, as to the approximation of the solution of problem (P), we have.

THEOREM 6.2 : *Assume in* (P) (2.1)-(2.5) (*with* $p = +\infty$), (2.15), (5.3), *together with* (3.1) *and* (3.4) *for the discretization. Moreover, let* $\{\psi^m\}$ *be a sequence of functions satisfying* (5.6)-(5.8). *Then, for any sufficiently small h, there exist two integers* $\tilde{m}$ *and* $\tilde{n}$ *(uniquely determined by h) such that, if we call* $\tilde{u}_h = u_h^{\tilde{m},\tilde{n}}$ *the solution of problem* (6.1) *for such values of m and n, the following estimate holds true :*

$$\| u - \tilde{u}_h \|_\infty \leq c h^{\beta\tilde{\delta} - \varepsilon} |\log h|^2, \quad \forall \varepsilon > 0 , \tag{6.6}$$

c being a constant independent of h.

Proof: For any fixed m, Theorem 6.1, (5.9) and (5.8) imply :

$$\begin{aligned} \| u - {}_h^{m,\tilde{n}} \|_\infty &\leq \| u - u^m \|_\infty + \| u^m - u_h^{m,\tilde{n}} \|_\infty \\ &\leq c m^{-\beta} + c(1 + m^{2-\beta}) h^{2\tilde{\delta} - \varepsilon} |\log h|^2, \quad \forall \varepsilon > 0 . \end{aligned}$$

The thesis then follows if m is suitably choosen in dependence of h, i.e., if we define $\tilde{m}$ through the inequalities

$$h^{-\tilde{\delta}} \leq \tilde{m} < 1 + h^{-\tilde{\delta}} . \quad \square$$

Remark 6.1 : In the case of Lipschitz continuous coefficients ($\delta = 1$ in (2.15)), the estimate (6.6) yields the rate of convergence $O(h^{\beta-\varepsilon})$, which is the known result for variational inequalities with Hölder continuous obstacles (see [6]).

REFERENCES

[1] C. BAIOCCHI, Estimation d'erreur dans L^∞ pour les inéquations à obstacle, *Lecture Notes in Math., 606* (1977), 27-34.

[2] A. BENSOUSSAN-J. L. LIONS, Applications des inéquations variationnelles en contrôle stochastique, *Dunod,* 1978.

[3] M. CHICCO, Solvability of the Dirichlet problem in $H^{2,p}(\Omega)$ for a class of linear second order elliptic partial differential equations, *Boll. UMI, 4* (1971), 374-387.

[4] P. G. CIARLET-P. A. RAVIART, Maximum principle and uniform convergence for the finite element method, *Comput. Methods Appl. Mech. Engrg., 2* (1973), 17-31.

[5] Ph. CORTEY-DUMONT, Sur les inéquations variationnelles à opérateur non coercif, *R.A.I.R.O.-M²AN, 19* (1985), 195-212.

[6] S. FINZI VITA, L^{∞}-error estimates for variational inequalities with Hölder continuous obstacles, *R.A.I.R.O.-Anal. Numér., 16* (1982), 27-37.

[7] S. FINZI VITA, Sul trattamento numerico di problemi ellittici di tipo non variazionale, *Atti del Convegno di Analisi Numerica* (Sorrento, 1985), *De Frede,* Napoli (1986), 195-205.

[8] A. FRIEDMAN, Stochastic differential equations and applications (II), *Academic Press,* 1975.

[9] M. G. GARRONI-M. A. VIVALDI, Approximation results for bilateral nonlinear problems of non-variational type, *Nonlinear Anal., 8* (1984), 301-312.

[10] M. G. GARRONI-M. A. VIVALDI, Bilateral evolution problems of non-variational type : existence, uniqueness, Hölder regularity and approximation of solutions, *Manuscripta Math., 48* (1984), 39-69.

[11] J. ORTEGA-W. RHEINBOLDT, Iterative solutions of nonlinear equations in several variables, *Academic Press,* 1970.

[12] J. NITSCHE, L^{∞}-convergence of finite element approximation, *Lecture Notes in Math., 606* (1977), 261-274.

[13] R. H. NOCHETTO, Approximacion de problemas elipticos de frontera libre, *Universidad Complutense de Madrid* (internal report, 1985).

[14] G. M. TROIANIELLO, Some unilateral problems of the non-variational type, *Math. Nachr., 106* (1982), 47-62.

Saisie MATHOR — Photocomposition PHOTOMAT

JOUVE, 18, rue Saint-Denis, 75001 PARIS

Dépôt légal 1991 — Éditeur : 029 — CPPP 63420

N° 13879, Février 1991 *Imprimé en France*

Le Directeur de la publication : J. LISSARRAGUE