Comp. & Maths. with Appls. Vol. 2, pp. 17–40. Pergamon Press, 1976. Printed in Great Britain.

A DUAL ALGORITHM FOR THE SOLUTION OF NONLINEAR VARIATIONAL PROBLEMS VIA FINITE ELEMENT APPROXIMATION

DANIEL GABAY
Centre National de la Recherche Scientifique, Laboratoire d'Analyse Numérique, L. A. 189, 4, Place Jussieu, 75230 Paris Cedex 05, France

BERTRAND MERCIER*
IRIA, Laboria, Domaine de Voluceau, 78150, Le Chesnay, France

Communicated by R. Glowinski

(*Received May*, 1975)

Abstract—For variational problems of the form

$$\operatorname*{Inf}_{v\in V}\{f(Av)+g(v)\},$$

we propose a dual method which decouples the difficulties relative to the functionals f and g from the possible ill-conditioning effects of the linear operator A.

The approach is based on the use of an Augmented Lagrangian functional and leads to an efficient and simply implementable algorithm. We study also the finite element approximation of such problems, compatible with the use of our algorithm. The method is finally applied to solve several problems of continuum mechanics.

1. INTRODUCTION

Many problems, in Physics, Mechanics, and Mathematical Economics, can be formulated as the following variational problem:

$$(\mathcal{P})\quad \operatorname*{Inf}_{v\in V}\{f(Av)+g(v)\}.$$

V and Y are two Hilbert spaces, endowed with the norm topology; A is a continuous linear operator from V into Y; f and g are convex functions defined respectively on Y and V, and taking their values in $(-\infty, +\infty]$. This formulation, first used by Rockafellar[1] to extend Fenchel's duality theory, includes, in particular, the ordinary problems of convex programming. We call v^* the solution of $(\mathcal{P})$, when it exists. Let

$$\phi = f\circ A + g.$$

A natural approach to solve $(\mathcal{P})$ consists in searching directly for the minimum of $\phi(v)$ over V. If ϕ is differentiable iterative techniques based on the use of its gradient are available to construct a sequence of points in V, converging to v^*. But the speed of convergence is slow if the operator A is ill-conditioned[2].

This difficulty may disappear if we use the following device: in a first time, we introduce the additional variable $y\in Y$, linked to the original variable $v\in V$ by the constraint $Av-y=0$, and consider the constrained problem $(\mathcal{P}_c)$ on $V\times Y$, obviously equivalent to $(\mathcal{P})$:

$$(\mathcal{P}_c)\qquad \operatorname{Inf}\{f(y)+g(v)\,|\,(v,y)\in V\times Y,\ Av-y=0\}.$$

We then eliminate the artificial constraint, just introduced, by a dual approach to solve $(\mathcal{P}_c)$.

*Present address: Ecole Polytechnique, Centre de Mathématiques Appliquées, 91120, Palaiseau, France.

In the following, $(.,.)$ and $((.,.))$ denote the scalar product on respectively Y and V, while $|.|$ and $\|.\|$ are the corresponding norms. V' and Y' are the dual spaces of V and Y. We denote the duality between V and V' by $\langle .,. \rangle$, while we identify Y' to Y.

The classical dual approach to $(\mathcal{P}_c)$ consists in introducing a Lagrange multiplier $\lambda \in Y'$ corresponding to the constraint and the Lagrangian functional $\mathcal{L}$, defined on $(V \times Y) \times Y'$ with values in $(-\infty, +\infty]$

$$\mathcal{L}(v,y;\lambda) = g(v) + f(y) + (\lambda, Av - y). \tag{1.1}$$

Duality theory[3, 1] tells us that, if there exists a *saddle point* $(v^*, y^*; \lambda^*)$ of $\mathcal{L}$ on $(V \times Y) + Y'$,† then (v^*, y^*) is solution of $(\mathcal{P}_c)$. In this case, we have also the equality between

$$\operatorname*{Inf}_{(v,y)\in V\times Y} \{\operatorname*{Sup}_{\lambda\in Y'} \mathcal{L}(v, y; \lambda)\},$$

where we recognize the primal problem $(\mathcal{P}_c)$, and

$$(\mathcal{D}) \quad \operatorname*{Sup}_{\lambda\in Y'} \{ \operatorname*{Inf}_{(v,y)\in V\times Y} \mathcal{L}(v, y; \lambda)\},$$

which constitutes the dual problem.

In order to guarantee the existence of a unique solution to the inner minimization problem in $(\mathcal{D})$ under the same hypotheses that insure the existence and uniqueness of a solution of $(\mathcal{P})$, it is convenient, as originally proposed by Glowinski and Marrocco[4], to define the dual problem with the *Augmented Lagrangian, introduced by Hestenes* [5] *and Powell* [6], is obtained by adding to the standard Lagrangian $\mathcal{L}(v, y; \lambda)$, expressed in (1·1), a term, depending on a positive parameter r, penalizing for the violation of the constraint; in our case

$$\mathcal{L}_r(v, y; \lambda) = f(y) + g(v) + (\lambda, Av - y) + \frac{r}{2}|Av - y|^2. \tag{1.2}$$

This formulation presents also the advantage over the ordinary penalty function (which corresponds to the case where λ is maintained equal to 0) that a minimizing sequence $\{v_n\}$ for $(\mathcal{P})$ can be generated without making $r \to +\infty$, thus avoiding the well-known ill-conditioning of ordinary penalty methods. For any arbitrary fixed $r > 0$, if $(v^*, y^*; \lambda^*)$ is a saddle point of $\mathcal{L}_r(v, y; \lambda)$, then v^* is solution of $(\mathcal{P})$.

In practice, we must solve a sequence of unconstrained minimization problems

$$\operatorname*{Inf}_{(v,y)\in V\times Y} \mathcal{L}_r(v, y; \lambda^n), \tag{1.3}$$

where the multipliers λ^n form a maximizing sequence of the dual fuctional

$$d_r(\lambda) = \operatorname*{Inf}_{(v,y)\in V\times Y} \mathcal{L}_r(v, y; \lambda). \tag{1.4}$$

Let (v^{n+1}, y^{n+1}) be the solution of (1.3). The concave functional d_r is differentiable and its gradient is given by

$$\nabla d_r(\lambda^n) = Av^{n+1} - y^{n+1}.$$

We can use the gradient method to maximize $d_r(\lambda)$ by generating the multipliers λ^n according to the iteration

$$\lambda^{n+1} = \lambda^n + \rho(Av^{n+1} - y^{n+1}). \tag{1.5}$$

A subsequence of $\{\lambda^n\}$ converges to λ^*, provided that the stepsize ρ is chosen $0 < \rho < 2r$ [10].

†$(v^*, y^* \lambda^*)$ is a saddle point of $\mathcal{L}$ iff

$$\mathcal{L}(v^*, y^*; \lambda) \leq \mathcal{L}(v^*, y^*; \lambda^*) \leq \mathcal{L}(v, y; \lambda^*) \quad \forall (v, y), \lambda \in (V \times Y) \times Y'$$

This dual algorithm due to Uzawa[7], is mentioned in the pioneering paper of Hestenes with $\rho = r$, and is used in [8] to solve the non-linear Dirichlet's problem. The convergence results for convex programming problems, in finite dimensional spaces, have been extended, by Rockafellar[9] and Bertsekas[10, 11], to the case where the minimization in (1.3) is performed approximately at each cycle and is only asymptotically exact.

An alternative formulation of the dual function (1.4), provided by Fortin[12], consists in writing

$$d_r(\lambda) = \operatorname*{Inf}_{v\in V}\left\{g(v) + \operatorname*{Inf}_{p\in Y}\left[f(Av+p) + \frac{r}{2}|p|^2 - (\lambda, p)\right]\right\}.$$

It is then possible to explicit the problem in p with the proximation operator of Y relatively to f[13]. The resulting problem is still non-linear in v and Av, and is of the same type as the original problem $(\mathcal{P})$. This approach is particularly interesting when f is a support function of a closed convex set of Y, since the non-differentiable unconstrained problem $(\mathcal{P})$ is replaced by a differentiable sup-inf problem. A classical dual method has been presented earlier for the same type of situation in[14, 15]. See also[16] for an approach similar to[12].

We present, in this paper, a modification of Uzawa's algorithm, hinted in[4], and experimented in[8], without proof of convergence.

In this modification, the minimization in (1.3) is neither performed exactly nor subject to a termination criterion like in [10, 9]. This proposal is based on the observation that problem (1.3) involves a problem of minimization in v coupled with a problem in y by the term $-r(y, Av)$. An approximate solution (v^{n+1}, y^{n+1}) to (1.3) is therefore provided by the following algorithm:

Given v^n, y^n, λ^n:

$$\text{select } v^{n+1}\text{, solution of } \operatorname*{Inf}_{v\in V}\left\{g(v) + \frac{r}{2}\|Av\|^2 + \langle\lambda^n - ry^n, Av\rangle\right\}; \tag{1.6}$$

$$\text{select } y^{n+1}\text{, solution of } \operatorname*{Inf}_{y\in Y}\left\{f(y) + \frac{r}{2}\|y\|^2 - \langle\lambda^n + rAv^{n+1}, y\rangle\right\}; \tag{1.7}$$

which we complete by iteration (1.5) for the multiplier λ^{n+1}.

The algorithm is interesting for computation only if problems (1.6) and (1.7) are relatively easy to solve. If $g(.)$ is a linear functional or a quadratic form, (1.6) becomes a quadratic problem. In the finite dimensional case, it can be solved by direct methods of matrix inversion, *which are less sensitive to the ill-conditioning effect of* A than gradient methods. We observe also that (1.7) is a non-linear problem in y which is not sensitive at all to the ill-conditioning of A. We have thus achieved a *decoupling of* f *and* A which constitutes a serious improvement upon $(\mathcal{P})$.

We have applied our algorithm to several problems of *continuum mechanics*, namely the minimal hypersurfaces problem, the problem of visco-plastic flow in a cylindrical pipe, and two problems of elasto-plasticity: torsion and equilibrium. All these problems are of the special form $(\mathcal{P})$ where $g(.)$ is a linear functional, and to which we shall restrict our analysis in the following. In 5.1, for example, V is taken as the Sobolev space $H_0^1(\Omega)$ (see [17]), where Ω is a regular open subset of $\mathbb{R}^N$ ($N = 1, 2$ or 3). Y is taken as $[L^2(\Omega)]^N$ and A is the gradient operator. Finally, the convex function f can be represented by an integral over Ω.

Although our algorithm can be applied to problem $(\mathcal{P})$ in this functional framework, numerical computations must, in practice, be performed in finite dimensional spaces. We thus construct internal approximations $V_h \subset V$ and $Y_h \subset Y$ of finite dimension[18] *via a finite element method*[19]. We introduce a regular triangulation $\mathcal{T}_h$ of Ω in a finite number of "triangles" T of size $\leq h$. V_h is chosen, for instance, as a space of continuous functions which are piecewise polynomial. Since we are not able, in general, to calculate exactly the integral defining f, we must use a formula for numerical integration with a certain number of integration nodes. Naturally, the approximation A_h of the operator A is defined by: the values taken by Av_h at each integration node. We thus choose for Y_h a space of piecewise constant functions of dimension $\mathcal{N} = N$ times the number of integration nodes. We establish that the solution v^*_h of the approximate problem

$$(\mathcal{P}_h)\quad \operatorname*{Inf}_{v_h\in V_h}\{f(A_hv_h) - \langle b, v_h\rangle\}$$

converges towards v^* as $h \to 0$, if the numerical integration is precise enough. Using our dual algorithm to solve $(\mathcal{P}_h)$, we benefit of a powerful decomposition property, [8], since each iteration consists in a quadratic problem (1.6) in V_h and $\mathcal{N}$ non-linear problems (1.7) in $\mathbb{R}^N$ independent of each other. The iteration (1.5) can also be performed component by component.

The article is organized as follows: in Section 2, after a brief review of duality theory, we study the properties of the augmented Lagrangian, and, in particular, the existence and uniqueness of a saddle point. In Section 3, we state a dual algorithm to compute such a saddle point, and we establish its convergence. In Section 4, we turn to the approximation of infinite dimensional problems. A convergence result is given and applied to a finite element approximation compatible with our dual approach. We use this approximation for several problems of continuum mechanics, stated in Section 5, and apply our algorithm. Numerical results are reported in Section 6, and compared with direct methods of solution.

2. CONVEXITY PROPERTIES AND THE AUGMENTED LAGRANGIAN

Duality theory and saddle-points

Duality theory has recently received an elegant formulation [1, 20], based on the consideration of a family of perturbed problems $(\mathcal{P}_p)$associated to a problem $(\mathcal{P})$: $\operatorname{Inf}_{v \in V} \phi(v)$. We consider a general bifunction $\Phi: V \times Y \to (-\infty, +\infty]$ such that

$$\Phi(v, 0) = \phi(v)$$

and the problem depending on a perturbation $p \in Y$

$$(\mathcal{P}_p) \quad \operatorname*{Inf}_{v \in V} \Phi(v, p).$$

In this framework, we define a Lagrangian function $\Lambda: V \times Y' \to [-\infty, +\infty]$ associated to $(\mathcal{P})$ by the relation

$$\Lambda(v; \lambda) = \operatorname*{Inf}_{p \in Y} \{\Phi(v, p) - (\lambda, p)\}. \tag{2.1}$$

We must immediatly observe that this is different from the classical Lagrangian function defined for constrained minimization problem like $(\mathcal{P}_c)$, since this function Λ is directly associated to the unconstrained problem $(\mathcal{P})$. Λ depends, however, on the choice of the perturbation bifunction Φ.

We can verify that

$$\operatorname*{Sup}_{\lambda \in Y'} \Lambda(v; \lambda) = \Phi(v, 0);$$

hence the problem

$$\operatorname*{Inf}_{v \in V} \operatorname*{Sup}_{\lambda \in Y'} \Lambda(v; \lambda)$$

is nothing else than problem $(\mathcal{P})$ and is independent of Φ. Parallel to the duality theory for convex programming in terms of mini-max, we define a dual problem to $(\mathcal{P})$

$$(\mathcal{D}) \quad \operatorname*{Sup}_{\lambda \in Y'} \operatorname*{Inf}_{v \in V} \Lambda(v; \lambda).$$

From now on, let

$$(\mathcal{P}) \quad \operatorname*{Inf}_{v \in V} \{f(Av) - \langle b, v\rangle\} \quad \text{with} \quad b \in V'.$$

We assume that f is the sum of two lower semi-continuous convex functions f_1 and f_2 from Y into $(-\infty, +\infty]$,

$$f = f_1 + f_2 \tag{2.2}$$

.f_1 *is* C^1.*Gateaux-differentiable* and its gradient f_1' is weakly continuous on finite dimensional subspaces of V and *strongly monotone*: i.e. there exists $\gamma > 0$ such that

$$\forall y, z \in Y (f_1'(y) - f_1'(z), y - z) \geq \gamma |y - z|^2. \tag{2.3}$$

.*the interior in* Y *of dom* f_2 *is non empty.* (2.4)

We assume moreover that

.the operator $A'A$ is an isomorphism from V onto V'; i.e. there exists $\alpha > 0$ such that

$$|Av|^2 \geq \alpha^2 \|v\|^2. \tag{2.5}$$

The strict convexity of $f \circ A$ follows.

We consider the specific perturbation bifunction

$$\Phi(v, p) = f(Av + p) - \langle b, v \rangle.$$

The dual problem is defined by

$$\operatorname*{Sup}_{\lambda \in Y'} \operatorname*{Inf}_{v \in V} \{ -\langle b, v \rangle + \operatorname*{Inf}_{p \in Y} [f(Av + p) - (\lambda, p)] \} = \operatorname*{Sup}_{\lambda \in Y'} \operatorname*{Inf}_{(v, y) \in V \times Y} \{ f(y) + (\lambda, Av - y) - \langle b, v \rangle \}, \tag{2.6}$$

after the change of the variable $y = Av + p$. Inside the brackets, we recognize the expression (1.1) of the standard Lagrangian functional $\mathscr{L}(v, y; \lambda)$ corresponding to the constrained problem $(\mathscr{P}_c)$. We have thus established the equivalence between the choice of the perturbation bifunction Φ and the introduction of the artificial variable y in problem $(\mathscr{P}_c)$.

We study, now, the existence of solutions to $(\mathscr{P})$ and $(\mathscr{D})$.

PROPOSITION 2.1 *Under* (2.3), (2.5), *there exists a unique solution* v^* *to* $(\mathscr{P})$.

Proof. There exists $v_0 \in V$ such that $f_2(Av_0) = (f_2 \circ A)(v_0) < +\infty$, otherwise $(\mathscr{P})$ has no meaning.

Let now $z = Av_0$ and $\phi(t) = f_1(z + t(y - z))$ for $y \in Y$; using (2.3) and the equality $\phi(1) = \phi(0) + \int_0^1 \phi'(t) dt$, we deduce

$$f_1(y) \geq f_1(z) + (f_1'(z), y - z) + \frac{\gamma}{2} |y - z|^2. \tag{2.7}$$

Since f_2 is a proper convex function, even if (2.4) if not satisfied, it has a continuous affine lower bound:
there exists $y_2 \in Y$ and $\beta \in IR$ such that:

$$f_2(y) \geq (y_2, y) + \beta \quad \forall y \in Y \tag{2.8}$$

Then, applying (2.5), we deduce the coercivity of $(\mathscr{P})$: $(\phi(v) \to +\infty$ if $\|v\| \to +\infty)$, and then, the existence of v^*[20]. Uniqueness follows from strict convexity of $f \circ A$.■

We now prove the existence of a saddle point of $\mathscr{L}(v, y; \lambda)$. We recall first that the subgradient of a function $\phi : V \to (-\infty, +\infty]$ at a point $u \in V$ is the set (possibly empty)

$$\partial \phi(u) = \{ \mu \in V' | \forall v \in V \quad \phi(v) - \phi(u) \geq \langle \mu, v - u \rangle \}. \tag{2.9}$$

THEOREM 2.1. *Any saddle point* $(v^*, y^*; \lambda^*)$ *of* $\mathscr{L}(v, y; \lambda)$ *over* $(V \times Y) \times Y'$ *satisfies*:

v^* *is solution of* $(\mathscr{P})$, $y^* = Av^*$ *and* $\lambda^* \in \partial f(Av^*)$ *with* $A'\lambda^* = b$.†

Conversely, if (2.3), (2.4), (2.5), (2.10) *hold, there exists at least one such saddle point.*

(2.10)

Proof. We first assume that $(v^*, y^*; \lambda^*)$ is a saddle point of

$$\mathscr{L}(v, y; \lambda) = f(y) + (\lambda, Av - y) - \langle b, v \rangle,$$

i.e. which satisfies

$$\mathscr{L}(v^*, y^*; \lambda) \leq \mathscr{L}(v^*, y^*; \lambda^*) \quad \forall \lambda \in Y' \tag{2.11}$$

†A' is the continuous linear operator from $Y' \to V'$, dual of A.

$$\mathscr{L}(v^*, y^*; \lambda^*) \leq \mathscr{L}(v, y; \lambda^*) \quad \forall v, y \in V \times Y. \tag{2.12}$$

(2.11) implies immediately $y^* = Av^*$. Let $v = v^*$ in (2.10). We get

$$f(y) - f(Av^*) \geq (\lambda^*, y - Av^*) \quad \forall y \in Y,$$

hence the necessary condition

$$\lambda^* \in \partial f(Av^*). \tag{2.13}$$

Let $y = Av$ in (2.12). We get, this time,

$$f(Av) - \langle b, v\rangle \geq f(Av^*) - \langle b, v^*\rangle \quad \forall v \in V,$$

which shows that v^* must be solution of $(\mathscr{P})$. It can also be written

$$f(Av) - f(Av^*) \geq \langle b, v - v^*\rangle,$$

which means that

$$b \in \partial(f \circ A)(v^*). \tag{2.14}$$

We now prove that such a triple $(v^*, y^*; \lambda^*)$ exists. Proposition 2.1 concludes to the existence of v^* since (2.3), (2.5) hold. The *qualification hypothesis* (2.4) guarantees that there exists a point of Y where f is finite and continuous. Then ([20] ch. 1, PR 5.7)

$$\partial(f \circ A)(v) = A' \partial f(Av) \quad \forall v \in V;$$

Therefore (2.14) shows that $\partial f(Av^*)$ is non empty; we can find a $\lambda^* \in f(Av^*)$ such that

$$A'\lambda^* = b,$$

since A' is an operator onto Y' by (2.5). We verify easily that $(v^*, y^*; \lambda^*)$ is actually a saddle point of $\mathscr{L}$.■

Remark 1. Since f_1 is Gateaux-differentiable and because of the qualification hypothesis on f_2, we have ([20], ch. 1, PR 5.6)

$$\partial f = \partial f_1 + \partial f_2.$$

We can therefore replace the condition $\lambda \in \partial f(Av)$ by

$$\lambda^* - f_1'(Av^*) \in \partial f_2(Av^*).■ \tag{2.15}$$

Remark 2. In the general case where the convex function $g(.)$ is not restricted to be a linear functional, we can prove the existence of a saddle point of the Lagrangian $\Lambda(v, \lambda)$ and hence of $\mathscr{L}(v, y; \lambda)$ under the stronger qualification hypothesis: there exists $v_0 \in \operatorname{dom} g$ such that $Av_0 \in \operatorname{int} \operatorname{dom} f$. ([20], ch. 3)■

The augmented Langrangian

Although $(\mathscr{D})$ has been shown to have a solution λ^*, the inner minimization problem in (2.6) may not have a bounded solution for every $\lambda \in Y'$. For this reason, we switch to the augmented Lagrangian. Like the standard Lagrangian was arising from the consideration of a particular perturbation bifunction, we can derive the *augmented Lagrangian* from the class of perturbations defined by [34, 12]

$$\Phi_r = \Phi(v, p) + \frac{r}{2}|p|^2 = g(v) + f(Av + p) + \frac{r}{2}|p|^2. \tag{2.16}$$

We define now

$$\Lambda_r(v;\lambda) = \operatorname*{Inf}_{p\in Y}\left\{\Phi(v,p)+\frac{r}{2}|p|^2-(\lambda,p)\right\}; \tag{2.17}$$

the dual problem

$$\operatorname*{Sup}_{\lambda\in Y'}\operatorname*{Inf}_{v\in V}\Lambda_r(v;\lambda) \tag{2.18}$$

can be written, after the change of variable $y = Av + p$

$$(\mathscr{D}_n)\quad \operatorname*{Sup}_{\lambda\in Y'}\operatorname*{Inf}_{(v,y)\in V\times Y}\left\{f(y)-\langle b,v\rangle+(\lambda,Av-y)+\frac{r}{2}|Av-y|^2\right\},$$

which agrees with $\mathscr{L}_r(v,y;\lambda)$ of (1.2).

The virtues of the augmented Lagrangian result from the following properties.

THEOREM 2.2. *If* r > 0, *any saddle point of the augmented Lagrangian* $\mathscr{L}_r$ is saddle point of the standard Lagrangian $\mathscr{L}$ *and conversely.*

(Hence, if (2.3), (2.4), (2.5) hold, according to Theorem 2.1, there exists a saddle point $(v^*, y^*; \lambda^*)$ of $\mathscr{L}_r$).

Proof. By definition (1.2), $\mathscr{L}_r$ differs from $\mathscr{L}$ by a non-negative term; hence

$$\mathscr{L}(v,y;\lambda)\leq\mathscr{L}_r(v,y;\lambda),$$

with equality if $y = Av$.

If $(v^*, y^*; \lambda^*)$ is saddle point of $\mathscr{L}$, we have $y^* = Av^*$. Therefore

$$\mathscr{L}_r(v^*,y^*;\lambda)=\mathscr{L}(v^*,y^*;\lambda)\leq\mathscr{L}(v^*,y^*;\lambda^*)=\mathscr{L}_r(v^*,y^*;\lambda^*)\leq\mathscr{L}(v,y;\lambda^*)\leq\mathscr{L}_r(v,y;\lambda^*) \tag{2.19}$$

and $(v^*, y^*; \lambda^*)$ is also a saddle point of $\mathscr{L}_r$. Conversely, if $(v^*, y^*; \lambda^*)$ is a saddle point of $\mathscr{L}_r$, $y^* = Av^*$, then:

$$\mathscr{L}(v^*,y^*;\lambda^*)=\mathscr{L}_r(v^*,y^*;\lambda^*)\leq\mathscr{L}_r(v,y;\lambda^*)\quad\forall(v,y)\in V\times Y.$$

Thus, returning to the definition (1.2) of $\mathscr{L}_r$:

$$f(y^*)-\langle b,v^*\rangle\leq f(y)-\langle b,v\rangle+(\lambda^*,Av-y)+\frac{r}{2}|Av-y|^2. \tag{2.20}$$

Choosing, first, $v = v^*$ in (2.20), we get, since $y^* = Av^*$:

$$f(y)-f(y^*)+(\lambda^*,y^*-y)+\frac{r}{2}|y-y^*|^2\geq 0\quad\forall y\in Y; \tag{2.21}$$

For any $z \in Y$, we take $y = \theta z + (1-\theta)y^*$ with $\theta \in]0, 1[$ in (2.21). We get, using the convexity of f:

$$\theta(f(z)-f(y^*))-\theta(\lambda^*,z-y^*)+\frac{r}{2}\theta^2|z-y^*|^2\geq 0; \tag{2.22}$$

Dividing by θ, and making $\theta\to 0$, we get:

$$f(y^*)-(\lambda^*,y^*)\geq f(y)-(\lambda^*,y)\quad\forall y\in Y. \tag{2.23}$$

On the other hand, choosing $y = y^*$ in (2.20), we get, since $y^* = Av^*$:

$$\frac{r}{2}|A(v-v^*)|^2+(\lambda^*,A(v-v^*))-\langle b,v-v^*\rangle\geq 0\quad\forall v\in V; \tag{2.24}$$

In the same way, taking $v = \theta w + (1-\theta)v^*$ in (2.24) with $\theta \in]0, 1[$, we get, as $\theta \to 0$:

$$(\lambda^*, Av^*) - \langle b, v^* \rangle \leq (\lambda^*, Aw) - \langle b, w \rangle \quad \forall w \in V; \tag{2.25}$$

Finally, adding (2.23) and (2.25) we obtain the desired result:

$$\mathcal{L}(v^*, y^*; \lambda^*) = f(y^*) - \langle b, v^* \rangle \leq f(y) + (\lambda^*, Av - y) - \langle b, v \rangle = \mathcal{L}(v, y; \lambda^*) \quad \forall (v, y) \in V \times Y$$

the other inequality being obvious.■

The augmented Lagrangian is useful to evaluate the dual functional:

$$d_r(\lambda) = \operatorname*{Inf}_{(v,y) \in V \times Y} \mathcal{L}_r(v, y; \lambda) \tag{2.26}$$

THEOREM 2.3. *Under* (2.3), (2.5), *for every* $\lambda \in Y'$, *there exists a unique solution* (v_λ, y_λ) *to the minimization problem in* (2.24).

Proof. From (2.5), (2.7) and (2.8), we see that $\mathcal{L}_r(v, y; \lambda) \to +\infty$ if v or $y \to +\infty$, so that the problem (2.26) is coercive. Then, it has a solution for any $\lambda \in Y'$.

Uniqueness follows from the strict convexity of $\mathcal{L}_r$, for $r > 0$.■

3. A DUAL ALGORITHM

Description

From now on, we assume that the hypotheses (2.2) to (2.5) hold. To solve $(\mathcal{D}_r)$, we can maximize the differentiable function $d_r(\lambda)$ using a gradient algorithm. This forms the basis for Uzawa's algorithm[7]:

Uzawa's Algorithm

Let $\lambda^0 \in Y$. *By induction* λ^n *being given*

Step 1 *Find* v^{n+1}, y^{n+1} *minimizing on* $V \times Y$

$$\mathcal{L}_r(v, y; \lambda^n) = f(y) - \langle b, v \rangle + (\lambda^n, Av - y) + \frac{r}{2}|Av - y|^2. \tag{3.1}$$

Step 2 *Make*

$$\lambda^{n+1} = \lambda^n + \rho(Av^{n+1} - y^{n+1}).■$$

Because of the convexity of f, the problem in Step 1 is equivalent to the variational inequality:

$$f(y) - f(y^{n+1}) + (\lambda^n + rAv^{n+1}, Av - (y - y^{n+1})) - \langle b, v \rangle \geq 0 \quad \forall v \in V, \quad \forall y \in Y;$$

which decomposes itself into a variational equation in V

$$r(Av^{n+1}, Av) = (ry^{n+1} - \lambda^n, Av) - \langle b, v \rangle \quad \forall v \in V, \tag{3.2}$$

and a variational inequality in Y, equivalent to

$$(f_1'(y^{n+1}), y - y^{n+1}) + f_2(y) - f_2(y^{n+1}) + (ry^{n+1} - rAv^{n+1} - \lambda^n, y - y^{n+1}) \geq 0 \quad \forall y \in Y, \tag{3.3}$$

since f_1 is Gateaux-differentiable. These two problems are coupled to each other by the terms (y^{n+1}, Av) in (3.2) and $(Av^{n+1}, y - y^{n+1})$ in (3.3). A simple way to decouple these problems consists in replacing, for instance, y^{n+1} by y^n in (3.2). The resulting system is much simpler to solve and provides an approximation to the minimization in Step 1. We state the new algorithm:

A modified dual algorithm (3.4)

Let $(y^0, \lambda^0) \in Y \times Y'$. *By induction* (y^n, λ^n) *being given*

Step 1: *Find* v^{n+1} *such that*:

$$r(Av^{n+1}, Av) = (ry^n - \lambda^n, Av) - \langle b, v \rangle \quad \forall v \in V. \tag{3.5}$$

Step 2: *Find* y^{n+1} *such that*:

$$(f_1'(y^{n+1}), y - y^{n+1}) + f_2(y) - f_2(y^{n+1}) + (ry^{n+1} - \lambda^n - rAv^{n+1}, y - y^{n+1}) \geq 0 \quad \forall y \in Y. \tag{3.6}$$

Step 3: *Make*

$$\lambda^{n+1} = \lambda^n + \rho(Av^{n+1} - y^{n+1}). \blacksquare \tag{3.7}$$

Remark 1. (3.5) is the variational formulation of the linear system

$$rA'Av^{n+1} = rA'y^n - A'\lambda^n - b,$$

which is invertible by (2.5). At each iteration, Step 1 consists of the resolution of this system for a different second member. ■

Remark 2. (3.6) is a non-linear variational problem in y independent of A. An interesting decomposition occurs if Y is finite dimensional and can be viewed as the product of k spaces Y_i such that f separates itself in

$$f(y) = \sum_{i=1}^{k} f_i(y_i), \text{ with } y = (y_1, \dots, y_k);$$

(3.6) is then equivalent to k variational problems

$$(f_{1i}'(y_i^{n+1}), y_i - y_i^{n+1}) + f_{2i}(y_i) - f_{2i}(y_i^{n+1}) + (ry_i^{n+1} - s_i^n, y_i - y_i^{n+1}) \geq 0 \quad \forall y_i \in Y_i$$

where s_i^n is the component on Y_i of $s^n = \lambda^n + rAv^{n+1}$. ■

Convergence

According to Theorem 2.2, there exists a saddle point $(v^*, y^*; \lambda^*)$ satisfying (2.8). It verifies also

$$r(Av^*, Av) = (ry^* - \lambda^*, Av) - \langle b, v \rangle \quad \forall v \in V, \tag{3.8}$$

$$(f_1'(y^*), y - y^*) + f_2(y) - f_2(y^*) + (ry^* - \lambda^* - rAv^*, y - y^*) \geq 0 \quad \forall y \in Y \tag{3.9}$$

$$y^* = Av^*. \tag{3.10}$$

Subtracting (3.8) from (3.5) yields:

$$r(A(v^{n+1} - v^*), Av) = (r(y^n - y^*) - (\lambda^n - \lambda^*), Av) \quad \forall v \in V. \tag{3.11}$$

We introduce the *projection operator* P from Y *onto* the range of A, $R(A)$. Since $R(A)$ is closed, we have $Y = R(A) \oplus R(A)^{\perp}$. Given any $y \in Y$, Py is the unique element of $R(A)$ such that

$$(Py, Av) = (y, Av) \quad \forall v \in V.$$

This allows to write (3.11) under the explicit form

$$A(v^{n+1} - v^*) = P(y^n - y^*) - \frac{1}{r} P(\lambda^n - \lambda^*). \tag{3.12}$$

Adding (3.9) with $y = y^{n+1}$ to (3.6) with $y = y^*$, we get

$$(f_1'(y^{n+1}) - f_1'(y^*), y^{n+1} - y^*) + (r(y^{n+1} - y^*) - (\lambda^n - \lambda^*) - rA(v^{n+1} - v^*), y^{n+1} - y^*) \leq 0$$

where the terms in $f_2(y^{n+1})$ and $f_2(y^*)$ have cancelled out. Using the expression (3.12) of

$A(v^{n+1}-v^*)$, we obtain

$$(f_1'(y^{n+1})-f_1'(y^*),y^{n+1}-y^*)+r|y^{n+1}-y^*|^2 \\ \leqslant r(y^n-y^*,P(y^{n+1}-y^*))+(\lambda^n-\lambda^*,(I-P)(y^{n+1}-y^*)) \tag{3.13}$$

where I is the identity operator on Y. A majorant for the first term is given by Schwartz' inequality:

$$(y^n-y^*,P(y^{n+1}-y^*))=(P(y^n-y^*),P(y^{n+1}-y^*))\leqslant\frac{1}{2}\{|P(y^n-y^*)|^2+|P(y^{n+1}-y^*)|^2\}. \tag{3.14}$$

We deduce from (3.7), (3.10), (3.12)

$$\lambda^{n+1}-\lambda^*=\lambda^n-\lambda^*-\rho\left[y^{n+1}-y^*-P(y^n-y^*)+\frac{1}{r}P(\lambda^n-\lambda^*)\right].$$

The projection of the equality of $R(A)$ and $R(A)^\perp$ yields

$$P(\lambda^{n+1}-\lambda^*)=\left(1-\frac{\rho}{r}\right)P(\lambda^n-\lambda^*)-\rho P(y^{n+1}-y^n), \tag{3.15}$$

$$(I-P)(\lambda^{n+1}-\lambda^*)=(I-P)(\lambda^n-\lambda^*)-\rho(I-P)(y^{n+1}-y^*) \tag{3.16}$$

and squaring the norm of both members of (3.16)

$$|(I-P)(\lambda^{n+1}-\lambda^*)|^2=|(I-P)(\lambda^n-\lambda^*)|^2+\rho^2|(I-P)(y^{n+1}-y^*)|^2 \\ -2\rho(\lambda^n-\lambda^*,(I-P)(y^{n+1}-y^*)). \tag{3.17}$$

We now use the strong monotonicity of f_1' (2.3) to obtain

$$(f_1'(y^{n+1})-f_1'(y^*),y^{n+1}-y^*)\geqslant\gamma|y^{n+1}-y^*|^2; \tag{3.18}$$

Combining (3.14), (3.17), (3.18) with (3.13) gives

$$\gamma|y^{n+1}-y^*|^2+\left(r-\frac{\rho}{2}\right)|(I-P)(y^{n+1}-y^*)|^2+\frac{r}{2}|P(y^{n+1}-y^*)|^2+\frac{1}{2\rho}|(I-P)(\lambda^{n+1}-\lambda^*)|^2 \\ \leqslant\frac{r}{2}|P(y^n-y^*)|^2+\frac{1}{2\rho}|(I-P)(\lambda^n-\lambda^*)|^2. \tag{3.19}$$

Adding up (3.19) for $n=0,1,\ldots,N$:

$$\gamma\sum_{n=0}^{N}|y^{n+1}-y^*|^2+\left(r-\frac{\rho}{2}\right)\sum_{n=0}^{N}|(I-P)(y^{n+1}-y^*)|^2+\frac{1}{2\rho}|(I-P)(\lambda^{N+1}-\lambda^*)|^2 \\ \leqslant\frac{r}{2}|P(y^0-y^*)|^2+\frac{1}{2\rho}|(I-P)(\lambda^0-\lambda^*)|^2. \tag{3.20}$$

This shows that for any choice of ρ such that $0<\rho\leqslant 2r$, and for any N, the series

$$\sum_{n=0}^{N}|y^{n+1}-y^*|^2$$

remains bounded and so, converges; therefore

$$\lim_{n\to+\infty}|y^n-y^*|^2=0$$

which proves the strong convergence of $\{y^n\}$ to y^* in Y.

Squaring the norm of (3.15) and setting $\theta = (\rho / r)$, we get

$$|P(\lambda^{n+1} - \lambda^*)|^2 = (1-\theta)^2 |P(\lambda^n - \lambda^*)|^2 + \rho^2 |P(y^{n+1} - y^n)|^2 - 2\rho(1-\theta)(P(\lambda^n - \lambda^*), P(y^{n+1} - y^n)). \tag{3.21}$$

We use Schwartz' inequality under the form

$$(P(\lambda^n - \lambda^*), P(y^{n+1} - y^n)) \leq \frac{1}{2}\left\{\epsilon |P(\lambda^n - \lambda^*)|^2 + \frac{1}{\epsilon}|P(y^{n+1} - y^n)|^2\right\}$$

where ϵ is an arbitrary small positive number; (3.21) gives

$$|P(\lambda^{n+1} - \lambda^*)|^2 \leq ((1-\theta)^2 + \rho\epsilon|1-\theta|)|P(\lambda^n - \lambda^*)|^2 + \left(\rho^2 + |1-\theta|\frac{\rho}{\epsilon}\right)|P(y^{n+1} - y^n)|^2. \tag{3.22}$$

Consider the sequence of positive scalars u_n defined by

$$u_{n+1} = a u_n + b w_n \tag{3.23}$$

with $0 \leq a < 1$, $b > 0$, $w_n > 0$ and $\sum_{n=0}^{\infty} w_n < +\infty$. It is easy to verify that for any N, $\sum_{n=0}^{N} u_n < +\infty$ and $u_n \to 0$. We observe that the sequence

$$w_n = |P(y^{n+1} - y^n)|^2$$

satisfies the previous condition since

$$\sum_0^{\infty} w_n \leq 2 \sum_0^{\infty} |P(y^n - y^*)|^2 \leq 2 \sum_0^{\infty} |y^n - y^*|^2,$$

which has been shown to be bounded. We choose $b = [\rho^2 + |1-\theta|(\rho/\epsilon)]$ and ρ such that

$$a = \left(1 - \frac{\rho}{r}\right)^2 + \rho\epsilon\left|1 - \frac{\rho}{r}\right| < 1. \tag{3.24}$$

It is easy to check that, if $u_0 = |P(\lambda^0 - \lambda^*)|^2$, $|P(\lambda^n - \lambda^*)|^2 \leq u_n$ for every n and $|P(\lambda^n - \lambda^*)| \to 0$. Since ϵ is arbitrary, the inequality (3·24) is satisfied for all $0 < \rho < 2r$.

(3.12) allows us to conclude that $|A(v^{n+1} - v^*)| \to 0$ and, applying (2.5), we see that $\|v^{n+1} - v^*\| \to 0$.

We have thus established that the sequence $\{v^n\}$, constructed by algorithm (3.4), converges strongly to v^* in V. However, we have only proved that $|P(\lambda^n - \lambda^*)| \to 0$; we can deduce from (3.20) that $|(I-P)(\lambda^n - \lambda^*)|$ remains bounded. In fact λ^* may not be unique.

THEOREM 3.1. *For every stepsize* ρ *such that* $0 < \rho < 2r$, *the sequence* $\{(v^n, y^n)\}$, *constructed by algorithm* (3.4), *converges strongly in* $V \times Y$ *to* (v^*, Av^*) *where* v^* *is the unique solution of* $(\mathcal{P})$. *The sequence* $\{\lambda^n\}$ *remains bounded in* Y.■

Remark 1. Assuming only the monotonicity of f_1' instead of (2.3), we obtain a modification of (3.20):

$$\left(r - \frac{\rho}{2}\right)\sum_{n=0}^{N} |(I-P)(y^{n+1} - y^*)|^2 + \frac{r}{2}|P(y^{N+1} - y^*)|^2 + \frac{1}{2\rho}|(I-P)(\lambda^{N+1} - \lambda^*)|^2$$
$$\leq \frac{r}{2}|P(y^0 - y^*)|^2 + \frac{1}{2\rho}|(I-P)(\lambda^0 - \lambda^*)|^2;$$

thus, for all $0 < p < 2r$, the sequence $\{y^n\}$ generated by algorithm (3.4) remains bounded.

To prove theorem 3.1, it is therefore enough to assume the strong monotonicity of f_1' on every bounded set of Y.■

Remark 2. Convergence of the algorithm can be shown for $\rho = r$ under the weaker hypothesis on f_1: suppose that for all C positive, there exists a family of forcing functions δ_C: $\mathbb{R}^+ \to \mathbb{R}^+$ strictly monotone increasing with $\delta_C(0) = 0$ such that

$$(f_1'(y) - f_1'(z), y - z) \geq \delta_C(|y - z|) \ \forall y, z \in Y \text{ with } |y| \leq C \text{ and } |z| \leq C. \tag{3.25}$$

We can still prove that $y^n \to y^*$; for $\rho = r$. (3.12) and (3.15) give immediately the convergence of $v^n \to v^*$.

This assumption is satisfied in the important case where Y is finite dimensional and f_1 strictly convex [21].■

Remark 3. For a differentiable problem ($f_2 = 0$), we have the existence and uniqueness of λ^*. Moreover $\lambda^n \to \lambda^*$.■

4. APPROXIMATION VIA FINITE ELEMENTS

It is in practice convenient to approximate the infinite dimensional problem $(\mathcal{P})$ by a finite dimensional one.

Let $V_h \subset V$ be a family of internal approximation of V:

$$\forall v \in V, \textit{there exists } v_h \in V_h \textit{ such that } v_h \to v \textit{ when } h \to 0$$

(In fact, it is sufficient to check it for a dense subset of V).

In the same way, we shall approximate Y by $Y_h \subset Y$.

Since $A(v_h)$ may not be included in Y_h, we must approximate A by a continuous linear operator A_h: $V_h \to Y_h$. We assume that A_h satisfies:

(i) $\alpha'\|v_h\| \leq |A_h v_h| \leq M\|v_h\| \quad \forall v_h \in V_h$;
(ii) *If* $v_h \rightharpoonup v$ *weakly in* V, *then* $A_h v_h \rightharpoonup Av$ *weakly in* Y; (4.1)
(iii) *If* $v_h \to v$ strongly in V, *then* $A_h v_h \to Av$ *strongly* Y;
(iv) $|A_h v_h - A v_h| \to 0$ *when* $\|v_h\| \leq C$ *and* $h \to 0$.

We restrict, from now on, our analysis to the simplified case where

$$f(y) = \frac{1}{2}|y|^2 + f_2(y). \tag{4.2}$$

f_2 must be compatible with the approximation in the following sense:

$$\forall v \in V, \exists v_h \in V_h \textit{ with } v_h \to v \textit{ and } f_2(A_h v_h) \to f_2(Av) \tag{4.3}$$

we verify that the hypothesis (2.3) is satisfied in this case. According to proposition (2.1), there exists a unique solution v^* of $(\mathcal{P})$.

We approximate $(\mathcal{P})$ by the following problem:

$$(\mathcal{P}_h) \ \operatorname*{Inf}_{v_h \in V_h} \{f(A_h v_h) - \langle b, v_h\rangle\}$$

We now establish the convergence of the approximation in the following sense:

THEOREM 4.1 *Under the hypothesis* (4.1), (4.2), (4.3), *there is a unique solution* v^*_h *to* $(\mathcal{P}_h)$ *which converges to the solution* v^* *of* $(\mathcal{P})$:

$$v^*_h \to v^* \text{ in } V_h \textit{ strongly}.$$

Proof. (4.1) (i) shows the existence and uniqueness of v_h^*. Moreover v_h^* satisfies the variational inequality (equivalent to $(\mathscr{P}_h)$)

$$(A_h v_h^*, A_h(v_h - v_h^*)) + f_2(A_h v_h) - f_2(A_h v_h^*) \geq \langle b, v_h - v_h^* \rangle \quad \forall v_h \in V_h. \tag{4.4}$$

for $v_h = 0$, we obtain

$$|A_h v_h^*|^2 \leq f_2(A_h v_h^*) - f_2(0) + \langle b, v_h^* \rangle. \tag{4.5}$$

Applying (2.8) and (4.1) (i) to (4.5), we deduce that

$$\alpha'^2 \|v_h^*\|^2 \leq (M|y_2| + \|b\|_{V'})\|v_h^*\| + \beta - f_2(0);$$

hence $\|v_h^*\| \leq C$.

Thus, we can extract, from v_h^*, a subsequence (which we still call v_h^*) converging weakly to $w \in V$, as $h \to 0$. From (4.1) (ii) and (4.4) we have, choosing, for any $v \in V$, $v_h \in V_h$ as in (4.3):

$$\begin{aligned} |Aw|^2 + f_2(Aw) &\leq \liminf_{h \to 0} |A_h v_h^*|^2 + f_2(A_h v_h^*) \\ &\leq f_2(Av) + (Aw, Av) + \langle b, w - v \rangle \quad \forall v \in V \end{aligned} \tag{4.7}$$

which shows that $w = v^*$ solution of $(\mathscr{P})$. Thus, $v_h^* \rightharpoonup v^*$ in V weakly (without extraction of any subsequence). In fact, it converges strongly: by (4.1) (i)

$$\alpha' |v_h^* - v^*| \leq |A(v_h^* - v^*)| \leq |Av_h^* - A_h v_h^*| + |A_h v_h^* - Av^*| \tag{4.8}$$

(4.1) (iv) implies that the first term tends to zero. On the other hand, (4.4) implies that, for any $v \in V$ with $v_h \to v$ in V strongly:

$$\begin{aligned} X_h = |A_h v_h^* - Av^*|^2 &= |A_h v_h^*|^2 - 2(A_h v_h^*, Av^*) + |Av^*|^2 \\ &\leq (A_h v_h^*, A_h v_h) + f_2(A_h v_h) - f_2(A_h v_h^*) - \langle b, v_h - v_h^* \rangle - 2(A_h v_h^*, Av^*) + |Av^*|^2. \end{aligned}$$

Since f_2 is convex l.s.c., we know that $f_2(Av^*) \leq \liminf f_2(A_h v_h^*)$ and choosing v_h as in (4.3), we have:

$$f_2(A_h v_h) \to f_2(Av).$$

Then, taking the lim sup of both members in (4.9), we get:

$$\limsup X_h \leq (Av^*, Av) + f_2(Av) - f_2(Av^*) - \langle b, v - v^* \rangle - |Av^*|^2. \tag{4.10}$$

Choosing $v = v^*$, we get:

$$X_h \to 0 \tag{4.11}$$

which, by (4.8) implies the strong convergence of v_h^* toward f^*. ■

Remark 1. If f_2 has finite values, it is continuous (see [20]). Then, applying (4.1) (iv), we see that (4.3) holds. ■

2. Suppose that $f_2 = \delta_K$, the indicator function of some closed convex set $K \subset Y$:

$$\delta_K(y) = \begin{cases} 0 \text{ if } y \in K \\ +\infty \text{ otherwise} \end{cases} \tag{4.12}$$

Let us define K_0 and K_h:

$$K_0 = \{v \in V | Av \in K\}, \tag{4.13}$$

$$K_h = \{v \in V_h | A_h v_h \in K\}; \tag{4.14}$$

Hypothesis (4.3) can be formulated as in [18]:

$$\forall v \in K_0, \exists v_h \in K_h \text{ such that } v_h \to v \text{ in } V \text{ strongly.} \blacksquare \tag{4.15}$$

Remark 3. One could think of approximating problem $(\mathcal{P})$ by

$$\operatorname*{Inf}_{v_h \in V_h} \frac{1}{2}|A_h v_h|^2 + f_2(B_h v_h) - \langle b, v_h\rangle$$

where A_h and B_h are two approximations of A, satisfying (4.1). We see that, *in order to apply the penalty-duality algorithm* (3.4), *it is necessary that* $A_h = B_h$. Thus, when f_2 is the indicator function of some convex K of Y corresponding to a convex K_0 of V as in (4.13), the *choice of* A_h automatically determines K_h by (4.14), and we have

$$\delta_{K_h}(v_h) = \delta_K(A_h v_h). \blacksquare$$

Finite element approximation

We consider now a open set Ω in $\mathbb{R}^N$ and specify

$$V = H_0^1(\Omega), \quad Y = [L^2(\Omega)]^N$$

and

$$(Av)_i = \sum_{j=1}^{N} a_{ij}\frac{\partial v}{\partial x_j} \tag{4.16}$$

where the coefficients a_{ij} are smooth functions.

We introduce a triangulation $\mathcal{T}_h$ of Ω consisting of a *finite* family of "triangles" $T \subset \Omega$, sufficiently regular, and of size less or equal to the positive parameter h [22]. Let P_k denote the space of polynomials of degree less or equal to k and define the following spaces: V_h is the space of continuous functions on $\bar{\Omega}$ which are equal to polynomials of P_k on each triangle of $\mathcal{T}_h$, C_h is the space of vectorial functions, continuous on each triangle of $\mathcal{T}_h$, and $Z_h \subset C_h$ is the space of vectorial functions, the restriction of which belongs to $(P_{k-1})^N$ on each triangle of $\mathcal{T}_h$,

$$\begin{aligned} V_h &= \{v_h \in C^0(\bar{\Omega}) \mid \quad v_h|_T \in P_k \text{ for all } T \in \mathcal{T}_h\}, \\ C_h &= \prod_{T \in \mathcal{T}_h} (C^0(T))^N, \\ Z_h &= \{z_h \in Y \mid \quad z_h|_T \in (P_{k-1})^N \text{ for all } T \in \mathcal{T}_h\}. \end{aligned} \tag{4.17}$$

We notice that $A(V_h) \subset C_h$. But C_h is not finite dimensional. If the functions a_{ij} are constant, $A(V_h) \subset Z_h$, and the approximation defined by (V_h, Z_h, A) is convergent, according to theorem 4.1. If the a_{ij} are not constant, $A(V_h) \not\subset Z_h$; we now need an approximation formula for the numerical integration of

$$\int_\Omega |Av_h|^2 \mathrm{d}x.$$

Such a formula is also necessary in any case, if the function f_2 is defined by an integral over Ω.

This formula is defined by L integration nodes $\hat{a}_l$ on a reference element $\hat{T}$ of $\mathcal{T}_h$; then, for any triangle $T \in \mathcal{T}_h$, we consider the affine mapping J_T on Ω, transforming $\hat{T}$ into T. To the reference nodes $\hat{a}_l$ correspond on T the integration nodes $a_{T,l} = J_T(\hat{a}_l)$. We thus approximate the integral $\int_\Omega \phi(x)\mathrm{d}x$ by

$$\sum_{T \in \mathcal{T}_h} |\det J_T| \sum_{l=1}^{L} \hat{\omega}_l \phi(a_{T,l}).$$

We suppose that the formula is exact for the polynomials of $P_{k'}$, $k' \geq 0$. The use of this integration formula can be interpreted as the introduction of an approximate scalar product $(.,.)_h$ on C_h defined in the following way: for any $y_1, y_2 \in C_h$

$$(y_1, y_2)_h = \sum_{T\in\mathcal{T}_h} |\det J_T| \sum_{l=1}^{L} \hat{\omega}_l y_1(a_{T,l}) y_2(a_{T,l}). \tag{4.18}$$

Since $k' \geq 0$, the formula is exact for the constant functions

$$\sum_{l=1}^{L} \hat{\omega}_l = \text{meas}\,(\hat{T});$$

we can partition $\hat{T}$ into $\hat{T}_e$, $l = 1, \ldots, L$ such that

$$\bigcup_{l=1}^{L} \hat{T}_l = \hat{T},\ \hat{a}_l \in \hat{T}_l,\ \text{measure}\ (\hat{T}_l) = \hat{\omega}_l.$$

We extend this partitioning on every triangle T of $\mathcal{T}_h$, defining

$$T_l = J_T(\hat{T}_l). \quad l = 1, \ldots, L.$$

We are now able to define

$$Y_h = \{y_h \in Y| \quad y_h|_{T_l} \in (P_0)^N \text{ for } l = 1, \ldots, L \text{ and all } T \in \mathcal{T}_h\}. \tag{4.19}$$

Even though $Y_h \not\subset C_h$, we can easily extend the scalar product $(.,.)_h$ on Y_h and moreover

$$\forall y_1, y_2 \in Y_h \quad (y_1, y_2)_h = (y_1, y_2).$$

We now construct the mapping p_h: $C_h \to Y_h$, defined for any $z_h \in C_h$ by the restrictions of $p_h(z_h)$ to each T_l:

$$\forall z_h \in C_h \quad p_h(z_h)|_{T_l} = z_h(a_{T,l}) \text{ for } l = 1, \ldots, L \text{ and all } T \in \mathcal{T}_h. \tag{4.20}$$

We obviously have

$$\forall y_h, z_h \in C_h \quad (y_h, z_h)_h = (p_h(y_h), p_h(z_h)).$$

We finally define

$$A_h = p_h \circ A, \tag{4.21}$$

Which is an operator from V_h (4·17) into Y_h(4·19), such that

$$\forall v_h, w_h \in V_h \quad (Av_h, Aw_h)_h = (A_h v_h, A_h w_h). \tag{4.22}$$

We now show the convergence of this approximation in the sense of theorem 4.1. This results from the following proposition:

PROPOSITION 4.1. *If* $k' > 2k - 3$, *the operator* A_h, *defined in* (4.21) *satisfies* (4.1) *and provides us with a converging approximation.*

Proof. Following Nédelec[23], we show that:

$$\|Av_h^*|_h^2 - |Av_h^*|^2| \leq C_1 h^{k'-(2k-3)} \|v_h^*\|^2, \tag{4.23}$$

and, for all $z_h \in Z_h$,

$$|(Av_h^*, z_h)_h - (Av_h^*, z_h)| \leq C_2 h^{k'-(2k-3)} \|v_h\| |z_h|. \tag{4.24}$$

Then, if $k' > 2k - 3$, (4.23) implies (4.1) (i) for h sufficiently small.

To prove (4.1) (ii), suppose that $v_h \rightharpoonup v$, we want to prove that $(A_h v_h, y) \to (Av, y)\ \forall y \in Y$. But y being given, since $\mathcal{T}_h$ is a regular family of triangulations, we can choose a particular $z_h \in Z_h$ such that z_h is constant on each $T \in \mathcal{T}_h$ and $z_h \to y$ in Y strongly. (In fact, $z_h \in Y_h$ too). Since $z_h = p_h(z_h)$, $(A_h v_h, z_h) = (Av_h, z_h)_h$. Then, we can write:

$$|(A_h v_h, y) - (Av, y)| \leq |(A_h v_h, y - z_h)| + |(Av_h, z_h)_h - (Av_h, z_h)| + |(Av_h, z_h) - (Av, y)|.$$

Since $v_h \rightharpoonup v$, $|A_h v_h|$ is bounded and the first term tends to zero. It is the same for the second term applying (4.24), and for the third one, since A is continuous; (4.1) (ii) is then established.

Taking into account (4.23), (4.1) (iii) and (iv) follow immediately.■

5. APPLICATION TO SOME PROBLEMS IN CONTINUUM MECHANICS

In the following, Ω is a regular open bounded set of $\mathbb{R}^N$ of boundary Γ. ($N = 1, 2, 3$).

5.1 *Bingham fluids*

We consider, following Mossolov–Miasnikov[24] and Duvaut–Lions[25], the flow of a visco-plastic fluid in a cylindrical pipe of cross section Ω. Let u denote the component of the velocity of the fluid parallel to the axis, and b be the pressure drop per unit length of the pipe. The problem is to find the function $u(x) \in H_0^1(\Omega)$ minimizing

$$\frac{\nu}{2}\int_\Omega |\text{grad } v|^2 dx + g\int_\Omega |\text{grad } v| dx - \int_\Omega bv\, dx \tag{5.1}$$

where ν is the viscosity of the material in the fluid regions and g is the plasticity threshold.

This is a non-differentiable problem of the form $(\mathcal{P})$, where $V = H_0^1(\Omega)$, $Y = (L^2(\Omega))^2$, $A = \text{grad}$, $f_1(y) = (\nu/2)\int_\Omega |y(x)|^2 dx$ and $f_2(y) = g\int_\Omega |y(x)| dx$. We check that (2.2) to (2.5) hold.

The algorithm. Algorithm (3.4) can be made explicit as follows:

By induction, y^n, λ^n *being given*

Step 1: *Find* v^{n+1} *solution of* $\begin{cases} -r\Delta v^{n+1} + \text{div}(ry^n - \lambda^n) = b \text{ on } \Omega \\ v^{n+1}|_\Gamma = 0 \end{cases}$

Step 2: *Find* y^{n+1} *solution of the variational inequality*

$$(\nu + r)\int_\Omega y^{n+1}(y - y^{n+1})dx + g\int_\Omega (|y| - |y^{n+1}|)dx$$
$$\geq \int_\Omega (\lambda^n + r\,\text{grad } v^{n+1})(y - y^{n+1})dx \quad \forall y \in [L^2(\Omega)]^2 \tag{5.3}$$

Step 3: *Make*

$$\lambda^{n+1} = \lambda^n + \rho(\text{grad } v^{n+1} - y^{n+1}).\blacksquare \tag{5.4}$$

In fact, (5.3) is equivalent to

$$(\nu + r)y^{n+1} - s^n \in \partial f_2(y^{n+1}) \text{ a.e. in } \Omega, \tag{5.5}$$

where $s^n = \lambda^n + r\,\text{grad } v^{n+1}$. Since almost everywhere:

$$\partial f_2(Y) = \begin{cases} \dfrac{gy}{|y|} & \text{if } y \neq 0 \\ gB_2(0;1)\dagger & \text{if } y = 0 \end{cases} \tag{5.6}$$

(5.3) can be solved explicitly;

$$y^{n+1} = \frac{s^n}{\nu + r}\,\text{Max}\left(0, 1 - \frac{g}{|s^n|}\right) \text{a.e.} \tag{5.7}$$

The algorithm reduces to a sequence of Dirichlet problems (5.2), accompanied by the updating

†$B_2(0,1) = \{z \in R^2 \| z \leq 1\}$.

formulae (5.7), (5.4) for y^{n+1} and λ^{n+1}. Since (2.2) to (2.5) hold, the sequence $\{v_n\}$, constructed by this algorithm, converges to the unique solution v^* of (5.1).

Approximation. We approximate $(\mathcal{P})$ by $(\mathcal{P}_h)$ as described in Section 4. Since Y_h is a space of piecewise constant functions, formulae (5.5) to (5.7) are still valid, and the algorithm applied to problem $(\mathcal{P}_h)$ still consists of iterations formed by (5.2), (5.7), (5.4). Proposition 4.1 shows that, if $k' > 2k-3$, $v_h^* \to v^*$ as $h \to 0$.

5.2. *Minimal hypersurfaces problem*

We want to minimize the area of an hypersurface supported by a given contour in $\mathbb{R}^{N+1}$. Given $\Omega \subset \mathbb{R}^N$ and a function g defined on Γ (the contour being $\{(x,y) \in \Gamma \times \mathbb{R} \mid y = g(x)\}$), the problem is to find a solution u to:

$$\operatorname*{Inf}_{\substack{v \in V \\ v|_\Gamma = g}} \int_\Omega \sqrt{(1 + |\operatorname{grad} v|^2)}\, dx. \tag{5.8}$$

With a slight modification due to the presence of the non-homogeneous boundary condition, this can be formalized as a problem $(\mathcal{P})$ with $V = W_0^{1,1}(\Omega)$, $Y = [L^1(\Omega)]^N$, $A = \operatorname{grad}$, $f_1(y) = \int_\Omega \sqrt{(1 + |y(x)|^2)}\, dx$, $f_2 \equiv 0$, $b = 0$.

We see that V and Y are non-reflexive Banach spaces, and we cannot apply the theory we have developed so far. We refer to [20] for the study of the existence of solutions or generalized solutions to (5.8).

Approximation. Formulated in a finite dimensional space V_h, these difficulties disappear: f_1' is strongly monotone, there exists a unique bounded solution v_h^* to $(\mathcal{P}_h)$ and the algorithm (3.4) is convergent. We use to this purpose an approximation of the problem in the functional framework, by piecewise linear finite elements. We refer to [25] for further details and for convergence results on the approximation.

Algorithm. Algorithm (3.4) can be made explict as previously. We notice that Step 2 decomposes itself into $\mathcal{N}$ variational inequalities in y_i^{n+1} where y_i^{n+1} is the component of y^{n+1} on the ith element T_i. Each of these problems is equivalent to

$$\left(r + \frac{1}{\sqrt{(1 + |y_i^{n+1}|^2)}}\right) y_i^{n+1} = s_i^n \quad \begin{array}{l} \text{for } i = 1, \ldots, \mathcal{N} \\ \text{for } y_i^{n+1} \in \mathbb{R}^N, \end{array}$$

where s^n is, as before, given by $s^n = \lambda^n + r \operatorname{grad} v^{n+1}$. This system of $\mathcal{N}$ equations in N variables can be reduced to a system of $\mathcal{N}$ equations in one variable $\theta \geq 0$

$$\left(r + \frac{1}{\sqrt{(1+\theta^2)}}\right)\theta = |s_i^n| \text{ for } i = 1, \ldots, \mathcal{N}.$$

This non-linear equation can be solved efficiently by Newton's method. Finally, the algorithm is as follows:

Choose y^0, λ^0. *By induction,* y^n, λ^n *being given*

Step 1: *Find* v^{n+1} *solution in* V_h *of the non-homogeneous Dirichlet problem*

$$\begin{cases} -r\Delta v^{n+1} + \operatorname{div}(ry^n - \lambda^n) = 0 \text{ on } \Omega \\ v^{n+1}|_\Gamma = g \end{cases} \tag{5.9}$$

Step 2: *For* $l = 1, \ldots, L$ *and for* $T \in \mathcal{T}_h$, *solve the non-linear equation*

$$\left(r + \frac{1}{\sqrt{(1+\theta^2)}}\right)\theta = |\sigma^n|$$

with $\sigma^n = (\lambda^n + r \operatorname{grad} v^{n+1})|_{T_l}$
and set $y^{n+1}|_{T_l} = \theta \,.\, \dfrac{\sigma^n}{|\sigma^n|}$.

Step 3: *Make*

$$\lambda^{n+1} = \lambda^n + \rho(Av^{n+1} - y^{n+1}). \blacksquare$$

5.3. *Elasto-plastic torsion of a cylindrical bar*

We consider a cylindrical pipe of section Ω. We are led to the following minimization problem [25]:

$$\operatorname*{Inf}_{v\in K_0} \frac{1}{2}\int_\Omega |\text{grad } v|^2 - \int_\Omega bv\,dx \tag{5.10}$$

where K_0 is the closed convex set:

$$K_0 = \{v \in H_o^1(\Omega) \mid |\text{grad } v| \leq 1 \text{ a.e.}\}, \tag{5.11}$$

Let $Y = (L^2(\Omega))^N$; $V = H_0^1(\Omega)$; $A = \text{grad}$ and

$$K = \{y \in (L^2(\Omega))^N \mid |y| \leq 1 \text{ a.e.}\}. \tag{5.12}$$

We see that, if δ_K is the indicator function of K (see (4.12)), (5.10) can be written;

$$\operatorname*{Inf}_{v\in V} \frac{1}{2}\int_\Omega |\text{grad } v|^2 dx + \delta_K(\text{grad } v) - \langle b, v\rangle. \tag{5.13}$$

It is a particular case of $(\mathcal{P})$ with $f_1(y)_0 = \frac{1}{2}|y|^2$ and $f_2 = \delta_K$. We check that (2.3) and (2.5) hold, but K being empty in Y, f_2 is not continuous in any point, and (2.4) is not satisfied.

We approximate (5.10) by finite elements, and to prove convergence, we must check that (4.14) holds. This is performed in [18] for the particular case of piecewise linear finite elements ($k = 1$).

We then use algorithm (3.4) to solve the approximate problem $(\mathcal{P}_h)$. It converges to the unique solution of (5.10). Step 1 is again a Dirichlet problem while step 2 is

$$(1+r)(y^{n+1}, z - y^{n+1}) - (\lambda^n + rAv^{n+1}, z - y^{n+1}) \geq 0 \quad \forall z \in K \tag{4.12}$$

which can be solved explicitly in Y_h: in each integration node, we have;

$$y^{n+1} = (\lambda^n + rAv^{n+1}) \min\left(\frac{1}{1+r}, \frac{1}{|\lambda^n + rAv^{n+1}|_{\mathbb{R}^N}}\right) \tag{4.13}$$

5.4. *Elasto-plastic displacement problem*

Hencky's law results in displacement problem [25, 26], which is, again, a particular case of $(\mathcal{P})$ with

. $V = \{v \in (H^1(\Omega))^N \mid v = 0 \text{ on a part } \Gamma_u \text{ of } \Gamma\}$

. $Y = \{e_{ij} \in L^2(\Omega) \mid i, j = 1, \ldots, N \mid e_{ij} = e_{ji}\}$

with the scalar product (e, τ): $\sum_{i,j=1}^{n} \int_\Omega \tau_{ij} e_{ij}\,dx$

. $A = \epsilon$ "Strain operator", where, by definition,

$$\epsilon_{ij}(v) \equiv \frac{1}{2}\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right) \tag{5.14}$$

.Finally $f_2 = 0$, and f_1 is defined in the following way. (We suppose for the sake of simplicity that the matrix of the elasticity coefficients is equal to identity; otherwise, see [26])

$$f_1(e) = \sup_{\tau\in K} (\tau, e) - \frac{1}{2}|\tau|^2 \tag{5.15}$$

where K is the plasticity convex set

$$K = \{\tau \in Y \mid \mathcal{F}(\tau) \leq 0 \text{ a.e.}\} \tag{5.16}$$

and $\mathcal{F}$ is a convex function on $\mathbb{R}^{N(N+1)/2}$ called plasticity criterion (it may be Von Mísès, or Tresca Criterion).

It can be checked that f_1 is differentiable and its gradient is $f_1'(e) = P_K(e)$. Hypothesis (2.3) does not hold. In fact, f_1 is not even strictly convex.

We refer to[27] where approximation by piecewise linear finite elements has been studied, and convergence has been proved using the dual problem (Stress problem) which is coercive. Algorithm (3.4) can again be used to solve the approximate problem. Step 2 of the algorithm can be solved explicitly, element by element, and the algorithm is then reduced to the solution of a sequence of elastic problems (linear systems, the matrix of which is fixed). The method is particularly performing in this case, as we can see in the numerical results, for which we refer to[27], because the elasticity matrix is ill-conditioned.

5.5. *Non-linear Dirichlet problem* [4]

This is the following problem: find $u \in W_0^{1,p}(\Omega)$ such that

$$-\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left(|\nabla u|^{p-2}\frac{\partial u}{\partial x_i}\right) = b \tag{5.17}$$

Once again, algorithm (3.4) can be applied, with A = grad and $f(y) = \frac{1}{p}\int_\Omega |y|^p\, dx$: this has been performed by Glowinski–Marrocco[28], to which we refer for details and numerical results.

6. NUMERICAL RESULTS

Let $\epsilon = (1/r)$ in the following

6.1. *Mossolov problem*

We treat a 2-dimensional example, the exact solution of which is known[18]; Ω is the disk

$$\Omega = \{(x, y)\} \in R^2 | x^2 + y^2 < R^2\}.$$

We used a triangulation with 512 triangles and 225 internal nodes (Fig. 1) corresponding to piecewise linear finite elements.

The exact solution is constant on a disk of radius $R/2$ and its value is 0·25. The value of the approximate solution is $0{\cdot}2506 \pm 5 \,.\, 10^{-5}$. With $\epsilon = 1$ and $\rho = 1$, this value is reached already at the second iteration of the algorithm.

However, the convergence of the algorithm is much slower with respect to the multiplier λ (in fact, it is not proved). This can be seen on the quantity $|y^{n+1} - Av^{n+1}|$ which tends to zero quite slowly (Tables 1 and 2).

Little improvements can be obtained by a variation of ρ (Table 1). The choice of ϵ is more important, even though it is easy: taking $\epsilon < 1$ accelerates slightly the convergence in λ, but gives

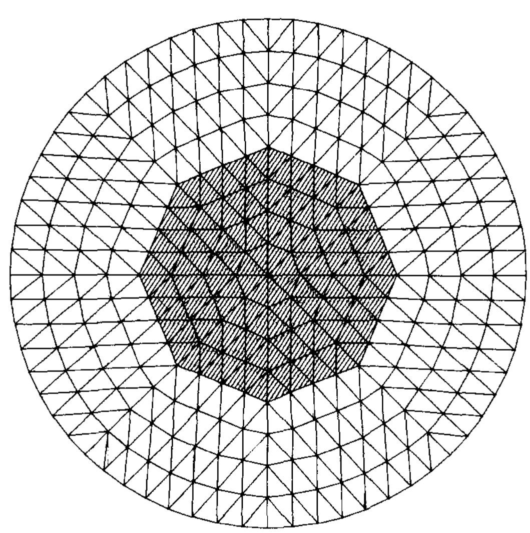

Fig. 1. Triangulation chosen for the unit disk [18] (512 nodes, 225 internal nodes). (dard regions are plastic).

Table 1. Influence of ρ on the speed of convergence. $\epsilon = 1$, 225 independent variables. (1) Norm chosen for the stopping test (2) n: number of performed iterations (3) U: value of the solution at the center of the disk (4) n^*: number of iterations necessary for the convergence of U

ρ	$\sum_{T \in \mathcal{T}_h} \lvert y_T^n - Av_T^n \rvert$	n	U	n^*
0.33	0.7	5	0.467	
0.66	0.1	5	0.2725	
1	0.03	5	0.2506	2
1.1	0.0022	10	0.2506	7
1.2	0.0020	10	0.2505	8
1.3	0.0018	10	0.2505	9
1.33	0.05	5	0.2574	
1.66	0.5	5	0.4093	

Table 2. Influence of ϵ on the speed of convergence. $\rho = (1/\epsilon)$. (Columns as for Table 1)

ε	$\sum_{T \in \mathcal{T}_h} \lvert y_T^n - Av_T^n \rvert$	n	U	n^*
0.5	4.10^{-4}	20	0.2505	19
1	7.10^{-4}	20	0.2506	2
1.5	$1.2\ 10^{-3}$	20	0.2506	10

reverse effect on u. (This can be seen on Table 2, where n^* is the number of iterations from which $v^n = 0{\cdot}2506 \pm 5\,.\,10^{-5}$). Conversely, when $\epsilon > 1$, both convergence in λ and u are slower.

A good compromise seems $\epsilon = 1$ and $\rho = 1$. For this value, the computing time is 2 sec on IBM 370/168. Comparison with duality methods [18] or direct non-differentiable methods of Davidon type [29, 30] shows the efficiency of the algorithm (5.2), (5.3), (5.4). For quadratic approximations with numerical integrations, it would be interesting to compare with Bristeau [31].

6.2. *Minimal surface problem*

We treat, as [32], a 2-dimensional example where Ω is the circular crown bounded by two concentrical circles Γ_1 and Γ_2, of radius 1 and 4 (Fig. 2). We take $g = 0$ on Γ_2 and $g = C$ (constant) on Γ_1. As noted by [32], for $C \geqslant 2.07$, the solution has some vertical parts and is no more in $W^{1,1}(\Omega)$.

We used, first, a regular triangulation with 192 triangles and 72 internal nodes. Then, dividing each triangle in 4 triangles, we got a new triangulation with 768 triangles and 336 variables (Fig. 3). At last, as the solution is irregular near Γ_1, we also used a triangulation, isomorphic to the previous one, but refined around Γ_1 (Fig. 4).

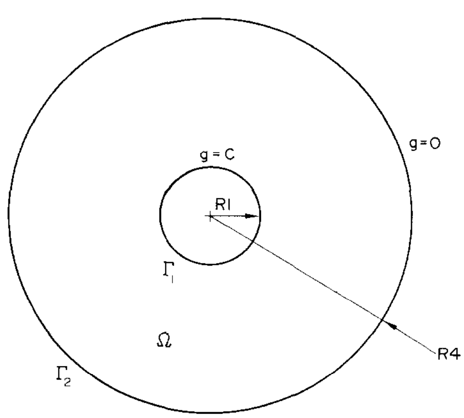

Fig. 2. Domain for the minimal surface problem.

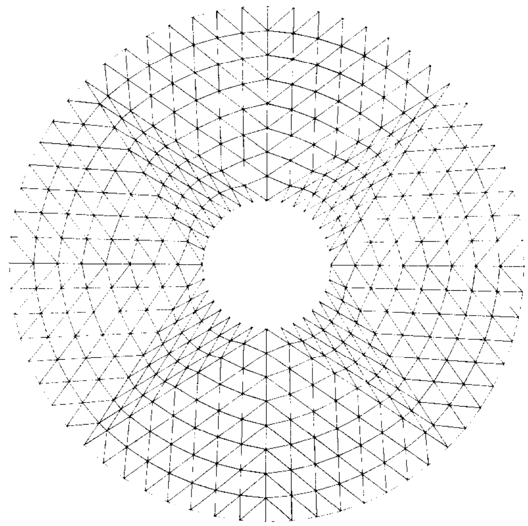

Fig. 3. Normal triangulation. 768 triangles, 336 variables.

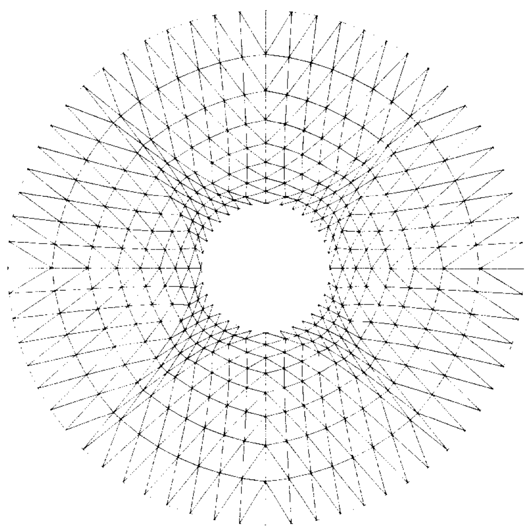

Fig. 4. Refined triangulation. 768 triangles, 336 variables.

The stopping test, chosen for the algorithm, is

$$E_n = \sum_{T \in \mathcal{T}_h} |y_T{}^n - Av_T{}^n| \leq \alpha;$$

For $\alpha = 10^{-6}$ and 72 variables, the variation of the number of iterations (n) with respect to ϵ is small (Fig. 5).

For $\epsilon = 1.8$, little improvements can be obtained by a variation of ρ (Table 3).

For 336 variables, the number of iterations increases within reasonable proportions for the regular triangulation, but more for the refined one (Table 4).

The computing time is 6 sec, for 72 variables, and 20 sec, for 336 variables (IBM 370/168), which is reasonable compared to more classical methods like non linear successive over relaxation[32] or non-linear conjugate gradient[28].

For another finite element approximation, see Jouron[33].

6.3. *Elasto-plastic torsion*

We treat, for the sake of simplicity, a one-dimensional example: $\Omega = [0, 1]$ and a discretization by piecewise linear finite elements on a regular mesh, the step of which is h.

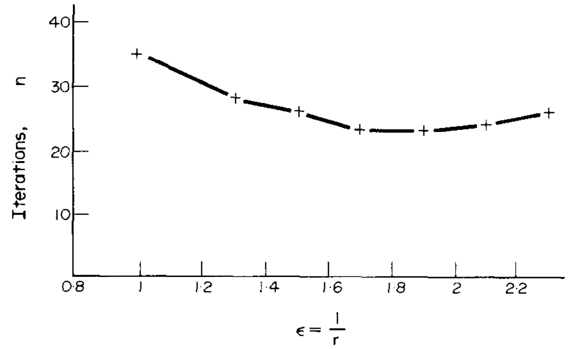

Fig. 5. Variation of n (number of iterations) with respect to ϵ. $\alpha = 10^{-6}$, 72 variables.

Table 3. Variation of n with respect to ρ. $\alpha = 10^{-7}$, 72 variables for $\epsilon = 1{\cdot}8$

$\rho . \varepsilon$	n
1	27
1.1	26
1.2	27
1.3	27

Table 4. Results with 336 variables

type of the triangulation	ε	C	E_n	n
normal	1.8	2	10^{-7}	43
	2	2	10^{-7}	39
	2.2	2	10^{-7}	36 (12")
	2.2	3	2.10^{-5}	66
refined	1.8	2	2.10^{-4}	46
	2.2	2	3.10^{-4}	26
	2.2	3	10^{-7}	140 (45")

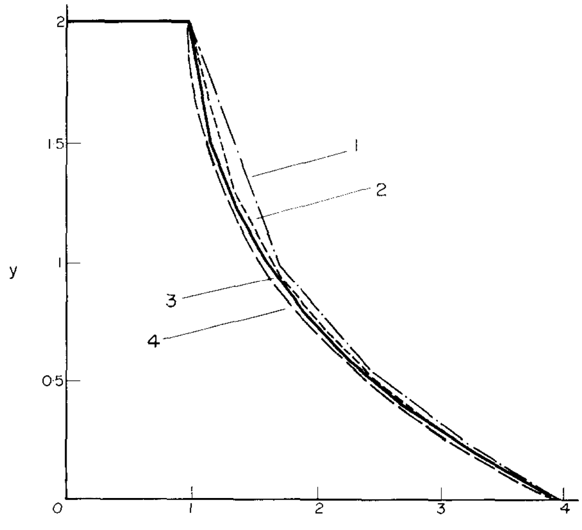

Fig. 6. Radial section of the solution (minimal surface) 1. 72 variables. 2. 336 variables (normal triangulation) 3. 336 (refined triangulation) 4. exact[32].

When b = constant, we have an exact solution which is parabolic on the elastic zone $]\frac{1}{2}-\frac{1}{b}, \frac{1}{2}+\frac{1}{b}[$ and linear elsewhere on Ω.

Numerical results show that for $b = 10$, starting from $y^0 = 0$ and $\lambda^0 = 0$, the best choice for ϵ is the biggest one, limited, however, by the appearance of round-off errors when ϵ is to big (Table 5). For ρ, a good choice seems to be $\rho = (1/\epsilon)$ (Table 6).

In that case, results are quite independent on b, even when $b < 2$, which corresponds to a quadratic problem.

A theorical study of the speed of convergence would show that starting from $\lambda^0 = 0$, it is better to take ϵ very big, this resulting from the fact that $\lambda^0 - \lambda^*$ is eigenvector of $A(A'A)^{-1}A'$.

On the contrary, if we start from $\lambda^0 = (1, 1, \ldots, 1)$, when b is small (quadratic problem), a good choice is $\epsilon = 1$ (Table 7).

Table 5. Choice of ϵ for the elasto-plastic torsion with $h = (1/20)$, $b = 10$ and $\lambda^0 = 0$

ε	n nb of iterations	E_n stopping test
2	10	10^{-4}
3	10	10^{-5}
10	10	7. 10^{-9}
10^2	7	10^{-10}
10^3	5	10^{-10}
10^4	5	10^{-11}
10^6	4	10^{-13}
10^{10}	4	10^{-13}
10^{12}	4	10^{-11}
10^{20}	10(round- off-errors)	10^3

Table 6. Choice of ρ with $h = (1/10)$, $b = 2$, $\lambda^0 = 0$, $\epsilon = 20$

ρ	n	E_n	$U(\frac{1}{2})$ (Uex=0.25)
0.5	10	0.2	0.26
0.9	10	10^{-7}	0.25
1	8	10^{-8}	0.25
1.1	10	10^{-7}	0.25
1.8	10	12	-0.35
2	10	90.9	-4.29

Table 7. Choice of ϵ when $\lambda^0 = (1, 1 \ldots, 1)$. $h = (1/10)$, $b = 2$

ε	n	E_n
10^{-3}	10	Cv.very slow
0.7	10	1.6 10^{-2}
1	10	9. 10^{-3}
10^3	10	7.5 } no Cv.
10^{12}	10	7.5

CONCLUSION

From the numerical point of view, we can summarize the properties of the penalty-duality algorithm.

It is very easy to implement, the main part of the algorithm being common to every problem and consisting of the solution of a linear system, the matrix of which can be factorized once for all, at the beginning. The only change, from one problem to another, is relative to step 2 of the algorithm, which consists in solving $\mathcal{N}$ non-linear equations, generally in one variable.

To select ρ and ϵ, the best is to choose $\rho = (1/\epsilon)$, and to try $\epsilon = 1$ or 2 at the beginning then, other values of ϵ may be tried.

It is better to start from $y^0 = 0$ and $\lambda^0 = 0$, at least at the beginning.

The algorithm has revealed to be very efficient, and the comparison with other classical methods[18, 30] shows that it is particularly useful when the matrix $A'A$ is ill-conditioned.

Acknowledgements—The authors are greatly indebted to Professor Glowinski, who originates the method, for his friendly encouragements and for very stimulating discussions during the research.

REFERENCES

1. R. T. Rockafellar, *Convex Analysis*. Princeton (1970).
2. D. G. Luenberger, *Introduction to linear and nonlinear programming* Addison-Wesley, Mass. (1973).
3. D. Geoffrion, Duality in nonlinear programming. A simplified applications-oriented development. *SIAM Rev.* **13**, 1–37 (1971).
4. R. Glowinski and A. Marrocco, Sur l'approximation, par éléments finis d'ordre 1, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non-linéaires. *C.R. hebd. Séanc. Acad. Sci., Paris.* **278**, Sér. A, 1649–1652 (1974).
5. M. Hestenes, Multiplier and gradient methods. *JOTA* **4**(5), 303–320 (1969).
6. M. J. D. Powell, A method for nonlinear optimization in minimization problems. *Optimization* (R. Fletcher Ed.) Academic Press (1969).
7. K. J. Arrow, L. Hurwicz and U. Uzawa, *Studies in Non-linear Programming*. Standford University Press (1968).
8. R. Glowinski and A. Marrocco, Rapport No 74023. Laboratoire d'Analyse Numérique, L.A. 189, Université de Paris VI. Paris France (to appear in R.A.I.R.O.).
9. R. T. Rockafellar, The multiplier Method of Hestenes and Powell applied to Convex Programming. *JOTA*. **12**(6), 555–562 (1973).
10. D. P. Bertsekas, On the method of multipliers for convex programming. *I.E.E.E. Trans.* AC-**20**, 385–388.
11. B. V. Kort and D. Bertsekas, Combined primal-dual and penalty methods for convex programming, *SIAM J.* (To appear).
12. M. Fortin, Minimization of some non-differentiable functionals by the Augmented Lagrangian method of Hestenes and Powell. To appear.
13. J. J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. math. Fr.* **93**, 273–299. (1965).
14. D. G. Luenberger, Control problems with kinks. *I.E.E.E. Trans.* **15**, 570–575 (1970).
15. J. Cea, R. Glowinski and J. C. Nedelec, Minimisation de fonctionnelles non différentiables. Conférence on Numerical Analysis, Dundee (Morris, ed) *Lectures Notes in Mathematics* Springer Verlag 19–38 (1971).
16. D. P. Bertsekas, Non-differentiable optimization via approximation. *Math. Prog. Study* **3**. North Holland (1976) To appear.
17. J. L. Lions, *Quelques méthodes de résolution des problèmes non-linéaires*, Dunod, Paris (1969).
18. R. Glowinski, J. L. Lions and R. Tremolieres, *Analyse numérique des inéquations variationnelles*. Dunod, Paris. To appear.
19. Zienkiewicz, *The Finite Element Method in Engineering Science*. McGraw-Hill (1970).
20. I. Ekeland and R. Temam, *Analyse convexe et problèmes variationnels* Dunod, Paris (1971).
21. J. Céa and R. Glowinski, Sur des méthodes d'optimisation par relaxation, R.A.I.R.O. *R*3, 5–32 (1973).
22. P. G. Ciarlet and P. A. Raviart, The combined effect of curved boundaries and numerical integration in isoparametric finite element method. In *The Mathematical Foundations of the Finite Element Method* (A. K. Aziz, ed) Academic Press, 409–474 (1972).
23. J. C. Nédelec, Approximation par éléments finis des équations de Riccati. Université de Rennes, *Journées Eléments Finis* (1974).
24. P. P. Mosolov and V. P. Miasnikov, Variational methods in the theory of the fluidity of a visco-plastic medium. *PMM* **29**, 468–492 (1965).
25. G. Duvaut and J. L. Lions, *Les inéquations en Mécanique et en Physique*, Dunod, Paris (1972).
26. B. Mercier, Approximation par éléments finis et résolution par un algorithme de pénalisation-dualité d'un problème d'élasto-plasticité. *C.R. hebd. Séanc. Acad. Sci., Paris.* **280**, Sér. A, 287–290 (1975).
27. B. Mercier, A finite Element Aptroximation of Elasto-Plastic Problem. *International Computing Symposium* (E. Gelenbe and D. Potier, Eds.) North Holland, 23–31 (1975).
28. R. Glowinski and A. Marrocco, On the solution of a class of non-linear Dirichlet problems by a penalty–duality method and finite element of order one. *Proceedings* 6th *Congress IFIP-TC7* (*Optimization*) Springer Verlag (1975).
29. C. Lemaréchal, An extension of Davidon methods to nondifferentiable functions. *Math. Prog. Study* **3**, North Holland (1976). To appear.
30. C. Lemaréchal, Thèse d'Etat, Paris (To appear).
31. M. O. Bristeau, Thèse de 3ème cycle, Paris (1975).
32. A. Vanende-Morreeuw, Thèse de 3ème cycle, Paris (1972).
33. C. Jouron, Resolution numérique du problème des surfaces minima. *Archs Ration. Mech. Analysis* (To appear).
34. R. T. Rockafellar, A dual approach to solving nonlinear programming problems by unconstrained optimization *Math. Prog.* **5** 354–373 (1973).