

Matiérisation | code matière AMS (1980) 65N., 47H. .  
mots-clefs

**PUBLICATIONS**

**MATHEMATIQUES**

**D'ORSAY**

80.01

INEQUATIONS VARIATIONNELLES DE LA MECANIQUE

B. MERCIER

Université de Paris-Sud  
Département de Mathématique

Bât. 425

91405 **ORSAY** France

# INEQUATIONS VARIATIONNELLES DE LA MECANIQUE

B. MERCIER  
Orsay, 1979

## Plan

- A - Analyse numérique des inéquations stationnaires
- B - Origine des inéquations variationnelles
- C - Analyse numérique des inéquations d'évolution

L'avantage de ce plan est d'introduire dans la partie A les outils nécessaires à la partie B.

Le signe # indique une fin de démonstration où une fin d'unité logique.

## A - ANALYSE NUMERIQUE ET CADRE ABSTRAIT

### Plan

1. Exemples et cadre abstrait
  - 1.1. Problème de l'obstacle
  - 1.2. Problème de Mosolov
  
2. Opérateurs maximaux monotones
  - 2.1 Monotonie
  - 2.2. Maximalité
  - 2.3 Algorithmes pour résoudre  $Au \ni 0$
  - 2.4. Contractions fermes
  - 2.5. Application
  
3. Somme d'opérateurs maximaux monotones  $A+B$ 
  - 3.1. Difficulté
  - 3.2. Cas B lipschitzien
  - 3.3. Algorithme I :  $u^{n+1} = J_A^\lambda(I - \lambda B)u^n$
  - 3.4. Algorithme II :  $u^{n+1} = J_A^\lambda J_B^\lambda u^n$  et régularisation
  - 3.5. Applications
  - 3.6. Algorithme III - Douglas-Rachford
  - 3.7. Algorithme IV - Peaceman-Rachford
  
4. Dualité
  - 4.1. Fonctions convexes conjuguées
  - 4.2. Problèmes en dualité
  - 4.3. Faut-il résoudre le primal ou le dual ?
  - 4.4. Exemples - Problème de Mosolov - Problème Elasto-plastique
  
5. Approximation des problèmes  $\inf_{v \in V} \phi(\Lambda v)$  où  $\Lambda : V \rightarrow Y$ 
  - 5.1. Cas où  $\phi$  est coercive
  - 5.2. Cas où  $\phi^*$  est coercive
  - 5.3. Influence de la discrétisation sur la vitesse de convergence

## 1. EXEMPLES ET CADRE ABSTRAIT

### 1.1. Problème de l'obstacle

Soit  $f \in L^2(\Omega')$  une fonction donnée sur un ouvert  $\Omega \subset \mathbb{R}^n$ . On veut résoudre le problème suivant

$$(1-1) \quad -\Delta u \leq f \quad \text{dans } \Omega$$

$$(1-2) \quad u \leq \psi \quad \text{dans } \Omega$$

$$(1-3) \quad (\Delta u + f) = 0 \quad \text{là où } u < \psi \quad \text{dans } \Omega$$

$$(1-4) \quad u = 0 \quad \text{sur } \Gamma, \text{ frontière de } \Omega.$$

où  $\psi \in H^2(\Omega)$  est une fonction donnée, appelée obstacle, supposée positive sur  $\Gamma$  pour que le problème admette une solution) ;  $H^1(\Omega)$  désigne l'espace de Sobolev

$$\{v \in L^2(\Omega) : \frac{\partial v}{\partial x_i} \in L^2(\Omega), i=1, \dots, n\}$$

et

$$H^2(\Omega) = \{v \in H^1(\Omega) : \frac{\partial^2 v}{\partial x_i \partial x_j} \in L^2(\Omega), i, j=1, 2, \dots, n\}$$

Le sens à donner à la 3ème équation est le suivant : dans une zone  $\Omega^-$  où  $u < \psi$ , on a  $-\Delta u = f$ , alors que dans le complémentaire de  $\Omega^-$  on a seulement l'inéquation.

La frontière de  $\Omega^-$  étant inconnue, le problème est dit "à frontière libre".

Mécaniquement, en dimension d'espace  $n=2$ , la solution  $u$  représente le déplacement transversal d'une membrane élastique occupant initialement un domaine  $M = \{x \in \mathbb{R}^3 : (x_1, x_2) \in \Omega, x_3 = 0\}$ , fixée au bord (cf (4)), soumise à une force transversale  $f$ , et se heurtant à un obstacle  $\mathcal{O} = \{x \in \mathbb{R}^3 : (x_1, x_2) \in \Omega \text{ et } x_3 = \psi(x_1, x_2)\}$ .

Pour résoudre ce problème, nous suivons LIONS-STAMPACCHIA<sup>(1)</sup> et en donnons une formulation variationnelle.

(1) C.P.A.M. 20 (1967) 493-519

Soit

$$a(u,v) \equiv \int_{\Omega} \nabla u \nabla v dX$$

$$(f,v) \equiv \int_{\Omega} f v dX.$$

et

$$K = \{v \in H_0^1(\Omega) : v \leq \psi \text{ p.p. sur } \Omega\} \quad (1)$$

Nous affirmons que  $u$  solution de (1-1) à (1-4) est aussi solution du problème

Trouver  $u \in K$  tel que

$$(1-5) \quad a(u, v-u) - (f, v-u) \geq 0 \quad v \in K.$$

En effet (sous réserve que  $u$  soit assez régulier, par exemple  $u \in C^0(\Omega)$ ), sur  $\Omega^-$  on a

$$(-\Delta u - f)(v-u) = 0$$

pour tout  $v \in K$ , sur le complémentaire de  $\Omega^-$ ; on a

$$v \leq \psi = u \Rightarrow (-\Delta u - f)(v-u) \geq 0$$

finalement on a (si  $\Delta u \in L^2(\Omega)$ )

$$(-\Delta u - f, v-u) \geq 0$$

d'où (1-5) en intégrant par parties. #

Réciproquement, si  $u$  est solution de (1-5), on choisit  $\phi \in \mathcal{D}(\Omega)$  et  $\phi \geq 0$  sur  $\Omega$  (i.e.,  $\phi \in \mathcal{D}_+(\Omega)$ ) d'où  $v = u - \phi \in K$ .

On obtient donc

$$\begin{aligned} a(u, \phi) &\leq (f, \phi) & \phi \in \mathcal{D}_+(\Omega) \\ \text{c'est-à-dire} & \quad -\Delta u \leq f & \text{dans } \mathcal{D}'(\Omega) \end{aligned}$$

(1)  $H_0^1(\Omega) = \{v \in H^1(\Omega) : v=0 \text{ sur } \Gamma\}$  est un sous espace fermé de  $H^1(\Omega)$  :  
cf ADAMS, Sobolev Spaces, Academic Press 1975, pour tous les détails concernant les espaces de Sobolev.

On choisit alors  $\phi \in \mathcal{D}(\Omega^-)$  et

$$\mu = \min_{x \in S} (\psi(x) - u(x)) / \max_{x \in \Omega^-} |\phi(x)|$$

où  $S$  désigne le support de  $\phi$ . On choisit alors dans (1-5)

$$v = u \pm \mu \phi \in K.$$

d'où

$$a(u, \phi) = f, \phi) \quad \forall \phi \in \mathcal{D}(\Omega^-)$$

c'est-à-dire  $-\Delta u = f$  sur  $\Omega^-$  (au sens des distributions), ce qui complète l'équivalence, en précisant son sens. #

### Cadre abstrait

Plus généralement soit  $V$  un espace de Hilbert,  $K \subset V$ , un convexe fermé,  $a(\cdot, \cdot)$  une forme bilinéaire continue sur  $V \times V$ , et  $f \in V'$  dual de  $V$ , le problème (1-5) est alors appelé inéquation stationnaire de type I.

Si  $a$  est symétrique :

$$a(u, v) = a(v, u)$$

Le problème (1-5) est équivalent au problème d'optimisation

$$(1-6) \quad \inf_{v \in K} J(v) \quad (1)$$

en effet la condition d'optimalité pour (1-6) est

$$(J'(u), v - u) \geq 0 \quad \forall v \in K. \quad \#$$

<sup>(1)</sup> Une fois pour toutes : cette écriture veut dire  
Trouver  $u \in K$  tel que  $J(u) \leq J(v) \quad \forall v \in K$

Lorsque  $a$  est coercif

$$(1-7) \quad a(v,u) \geq \alpha \|v\|_V^2, \quad \forall v \in V$$

L'existence d'une solution résulte du théorème très général suivant.

Théorème 1.1 : Soit  $V$  un espace de Banach réflexif et  $J : V \rightarrow \mathbb{R}$  une fonction convexe semi-continue inférieurement vérifiant

$$(1-8) \quad J(v) \rightarrow +\infty \text{ si } \|v\|_V \rightarrow +\infty$$

alors le problème (1-6) où  $K \subset V$  est un convexe fermé de  $V$  admet au moins une solution.

Si  $J$  est strictement convexe, la solution est unique.

Lorsque  $a$  est non symétrique, on a le résultat suivant dû à LIONS-STAMPACCHIA loc.cit. :

Théorème 1.2 : Si  $K$  est un convexe fermé non vide et si la forme bilinéaire  $a$  est continue et coercive, le problème (1-5) admet une solution unique  $u$ .

Démonstration :

On va montrer que  $u$  est le point fixe d'une contraction stricte  $T$ . Cette méthode est constructive, car  $u$  sera la limite de la suite récurrente

$$(1-9) \quad u^{n+1} = T u^n.$$

On définit donc  $z \equiv T y$  comme l'unique solution de

$$(1-10) \quad b(z, v-z) - L(v-z) \geq 0 \quad \forall v \in K$$

$$\text{où } L(v) \equiv b(y, v) - \rho(a(y, v) - (f, v))$$

et  $b(\cdot, \cdot)$  est une forme bilinéaire continue coercive et symétrique sur  $V \times V$  ce qui garantit l'existence de  $z$  solution de (1-10).

On vérifie que pour tout  $\rho > 0$ ,  $u$  est point fixe de  $T$ . Le choix standard pour  $b$  est le produit scalaire de  $V$  ; et on peut toujours se ramener à ce cas quitte à changer le produit scalaire de  $V$ , puisque  $b$  est symétrique. On a donc

$$b(y,v) \equiv (y,v)_V$$

On pose alors  $A:V \rightarrow V$  et  $u_f$  satisfaisant

$$(Au,v)_V \equiv a(u,v) \quad \forall v \in V$$

$$(u_f,v)_V = (f,v)$$

(l'existence de  $Au$  et de  $u_f$  découle de l'isomorphisme entre  $V$  et  $V'$ ). Avec ces notations on peut écrire (1-10) sous la forme

$$(z - (y - \rho(Ay - u_f)), v - z)_V \geq 0$$

$$z = P_K(y - \rho(Ay - u_f))$$

où  $P_K : V \rightarrow K$  désigne la projection sur  $K$ .

Soit  $z_i = Ty_i$   $i = 1, 2$ , on a

$$\|z_2 - z_1\|_V \leq \|(y_2 - \rho(Ay_2 - u_f)) - (y_1 - \rho(Ay_1 - u_f))\|_V$$

$$= \|y_2 - y_1 - \rho(Ay_2 - Ay_1)\|_V$$

où l'on a appliqué le fait que  $P_K$  est une contraction.

En élevant au carré et en développant, on a

$$\|z_2 - z_1\|_V^2 \leq \|y_2 - y_1\|_V^2 - 2\rho(Ay_2 - Ay_1, y_2 - y_1) + \rho^2 \|Ay_2 - Ay_1\|_V^2$$

Mais

$$(Ay_2 - Ay_1, y_2 - y_1)_V = a(y_2 - y_1, y_2 - y_1) \geq \alpha \|y_2 - y_1\|_V^2$$

$$\text{et } \|Ay_2 - Ay_1\|_V^2 \leq C^2 \|y_2 - y_1\|_V^2$$

où  $C$  est la constante de continuité de  $a$ , d'où

$$(1-11) \quad \|z_2 - z_1\|_V^2 \leq (1 - 2\rho\alpha + \rho^2 C^2) \|y_2 - y_1\|_V^2$$

qui montre que  $T$  est une contraction stricte pour

$$0 < \rho < \frac{2\alpha}{C^2}$$

Remarque 1.1 : La relation (1-11) donne même une évaluation de la vitesse de la convergence de l'algorithme (1-9), qui converge donc d'autant plus rapidement que  $b$  aura été choisi "proche" de  $a$  (pour que  $\alpha$  et  $C$  soient assez proches de 1). Noter que pour calculer  $u^{n+1}$  à partir de  $u^n$  dans l'algorithme (1-9) il faut tout de même résoudre une inéquation variationnelle. Cela étant, tous les algorithmes étudiés par la suite ressembleront plus ou moins à cet algorithme fondamental.

### 1.2. Problème de Mosolov

Nous donnerons directement la formulation variationnelle du problème qui est

$$(1-12) \quad \begin{array}{l} \text{Trouver } u \in V \text{ tel que} \\ a(u, v-u) + j(v) - j(u) \geq (f, v-u) \quad \forall v \in V \end{array}$$

$$\text{où } a(u, v) = \nu \int_{\Omega} \nabla u \nabla v \, dX$$

$$j(v) = \int_{\Omega} g |\nabla v| \, dX$$

$$V = H_0^1(\Omega)$$

$$f \in L^2(\Omega)$$

L'interprétation mécanique du problème est la suivante :  $u$  est le champ de vitesse d'un fluide de Bingham en écoulement laminaire stationnaire dans

une conduite cylindrique de section  $\Omega$  (le vecteur vitesse est orthogonal à  $\Omega$  et a pour module  $u(x)$  en un point  $x$ ). Nous aurons l'occasion de revenir en détail sur l'origine de l'inéquation (1-12) dans la partie B (§8). Les paramètres  $\nu$  et  $g$  positifs sont respectivement la viscosité et le seuil de plasticité du fluide de Bingham;  $f(x)$  désigne la perte de charge.

Nous reviendrons également plus loin sur l'interprétation de ce problème en termes d'équations aux dérivées partielles multivoques (§ 4).

Le problème (1-12) est appelé une inéquation stationnaire de type II.

#### Cadre abstrait

La nouveauté du problème (1-12) par rapport au problème (1-5) c'est la fonctionnelle  $j$  qui est convexe s.c.i. mais non différentiable. En revanche il n'y a pas ici de convexe  $K$ .

Théorème 1.3 <sup>(1)</sup> : Soit  $V$  un espace de Hilbert,  $j : V \rightarrow \mathbb{R}$  une fonction convexe s.c.i.,  $a$  une forme bilinéaire continue et coercive sur  $V \times V$ ,  $f \in V'$ , alors le problème (1-12) admet une solution unique.

(Adapter la démonstration du Théorème 1.2).

#### Interprétation en termes de problème d'optimisation.

Dans le cas où  $a$  est symétrique, le problème (1-12) est équivalent au problème d'optimisation

$$(1-13) \quad \inf_{v \in V} [J(v) + j(v)]$$

$$\text{où } J(v) = \frac{1}{2} a(v, v) - (f, v).$$

En effet la condition d'optimalité pour (1-13) s'écrit

$$(J'(u), v - u) + j(v) - j(u) \geq 0$$

et est visiblement équivalente à (1-12).

---

(1) Démontré dans un cadre plus général par LIONS, *Quelques Méthodes de Résolution des problèmes Non linéaires*, Dunod, Paris 1969.

Remarque : A condition d'étendre le champ de définition de  $j$  de  $V$  dans  $\mathbb{R} \cup \{+\infty\}$ <sup>(1)</sup>, le problème (1-6) est un cas particulier de (1-13) avec  $j=I_K$  fonction indicatrice de  $K$  définie par

$$I_K(v) = \begin{cases} 0 & \text{si } v \in K \\ +\infty & \text{sinon} \end{cases}$$

Bien entendu le problème (1-13) a un sens même si  $J$  n'est pas quadratique. Si  $J$  est coercif il aura néanmoins une solution (cf. Théorème 1.1).

### Orientation

Nous allons maintenant montrer que les inéquations stationnaires de type I et II sont en fait des cas particuliers de problèmes du type

$$Au + Bu \ni 0$$

où  $A$  et  $B$  sont des opérateurs multivoques.

Cette remarque nous servira pour définir des algorithmes de résolution utilisant la décomposition en une somme de deux opérateurs.

---

(1) EKELAND-TEMAN, *Analyse Convexe et problèmes Variationnels*, Dunod, Paris 1973.

## 2. OPERATEURS MAXIMAUX MONOTONES <sup>(1)</sup>

### 2.1. Opérateurs monotones

Soit  $H$  un espace de Hilbert. On note  $|\cdot|$  sa norme et  $(\cdot, \cdot)$  son produit scalaire. Un opérateur multivoque  $A$  est un sous ensemble (un graphe) de  $H \times H$ . Il est monotone si

$$(2-1) \quad (z-y, v-u) \geq 0 \quad \forall \{u, y\} \text{ et } \{v, z\} \in H \times H.$$

Pour  $u \in H$  on appelle  $A(u)$  l'ensemble (éventuellement vide) des  $y \in H$  tels que  $\{y, u\} \in A$ .

On appelle domaine de  $A$  l'ensemble

$$D(A) = \{u \in H : A(u) \neq \emptyset\}$$

et image de  $A$  l'ensemble

$$R(A) = \{y \in H : \exists v \in H \text{ tel que } \{v, y\} \in A\}$$

L'opérateur  $A^{-1}$  est celui dont le graphe est symétrique :

$$v \in A^{-1}(z) \Leftrightarrow z \in A(v).$$

### Exemple : Sous-différentielles de fonctions convexes

Soit  $\phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$  une fonction convexe, on définit son sous-gradient  $\partial\phi$  par

$$\partial\phi(u) = \{y \in H : \phi(v) - \phi(u) \geq (y, v-u) \quad \forall v \in H\}$$

alors  $\partial\phi$  est un opérateur monotone (multivoque).

En effet soit  $z \in \partial\phi(v)$  et  $y \in \partial\phi(u)$  on a

$$\phi(v) - \phi(u) \geq (y, v-u)$$

$$\phi(u) - \phi(v) \geq (z, u-v)$$

d'où la monotonie, par addition.

---

(1) Pour tout ce paragraphe la référence de base est H. BREZIS, Opérateurs Maximaux Monotones, North Holland, Lecture Note

Dans le cas où  $\partial\phi$  est univoque ( $\partial\phi(u)$  réduit à un élément)  $\partial\phi(u)$  coïncide avec le gradient de Gateaux  $\phi'(u)$  (cf EKELAND-TEMAM, loc. Cit.).  
On appelle

$$D(\phi) = \{u \in H : \phi(u) < +\infty\}$$

le domaine de  $\phi$ . Il faut noter que  $D(\partial\phi) \subset D(\phi)$  ; avec inclusion stricte en général (penser par exemple à  $H=L^2(\Omega)$ )

$$\phi(v) = \frac{1}{2} \int_{\Omega} |\nabla v|^2 dx \quad \text{avec } D(\phi) = H_0^1(\Omega) . \quad \text{On a alors}$$

$$\partial\phi = -\Delta \quad \text{et } D(\partial\phi) = H^2(\Omega) \cap H_0^1(\Omega).$$

Bien entendu il y a des opérateurs monotones qui ne sont pas des sous-gradients. Par exemple, dans  $\mathbb{R}^2$ , une rotation d'angle  $\leq \frac{\pi}{2}$  autour de l'origine est un opérateur monotone (linéaire) mais pas un sous-gradient.

## 2.2. Opérateurs maximaux monotones

L'inclusion des ensembles (dans  $H \times H$ ) fournit une relation d'ordre dans l'ensemble des opérateurs multivoques. On dit qu'un opérateur monotone  $A$  est maximal s'il est maximal dans l'ensemble des opérateurs monotones :

$$B \supset A, B \text{ monotone} \Rightarrow B = A.$$

Autrement dit  $(y-\eta, x-\xi) \geq 0 \quad \forall \{\xi, \eta\} \in A \Rightarrow y \in Ax$  (sinon  $A \cup \{x, y\}$  serait encore monotone et  $\supset A$ ).

Pour illustrer cette notion, considérons une application croissante  $f : \mathbb{R} \rightarrow \mathbb{R}$  univoque : elle est monotone.

Mais l'application  $\tilde{f} : x \in \mathbb{R} \rightarrow \tilde{f}(x) = [f(x_+), f(x_-)] \cap \mathbb{R}$  est aussi monotone. Elle est en général multivoque et contient strictement  $f$  : dans ce cas  $f$  ne peut être maximale monotone (en fait  $\tilde{f}$  est maximal monotone comme on le vérifierait à l'aide du résultat suivant).

Proposition 2.1 : Un opérateur A de H est maximal monotone si et seulement si pour tout  $\lambda > 0$  on a  $R(I+\lambda A) = H$ . #

On va seulement démontrer que  $R(I+A) = H \Rightarrow A$  maximal monotone. Soit en effet  $B \supset A$  monotone et  $y \in Bx$ . Par hypothèse  $\exists x' \in D(A)$  tel que  $x'+Ax' \ni x+y \Rightarrow x+y-x' \in Bx'$  et  $y \in Bx \Rightarrow x=x'$  d'après la monotonie de B.

La réciproque utilise l'axiome de Zorn : cf. BREZIS, loc. Cit., p. 23 #

Corollaire 2.1 : L'opérateur  $J_A^\lambda \equiv (I+\lambda A)^{-1}$  appelé résolvante de A est une contraction définie sur H tout entier pour tout  $\lambda > 0$ .

La monotonie de A entraîne en effet immédiatement que  $J_A^\lambda$  est une contraction : soit  $y_i \in H$  et  $u_i = J_A^\lambda y_i$ ,  $i=1,2$  ; on a

$$y_i = u_i + \lambda a_i$$

où  $a_i \in A(u_i)$ ,  $i = 1,2$ . On a

$$\begin{aligned} |u_2 - u_1|^2 &= (y_2 - y_1, u_2 - u_1) - \lambda(u_2 - u_1, a_2 - a_1) \\ &\leq (y_2 - y_1, u_2 - u_1) \leq |y_2 - y_1| |u_2 - u_1| \end{aligned} \quad \#$$

Exemple 2.1 : Sous-différentielles de fonctions convexes s.c.i.

Si  $\phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$  est convexe s.c.i., alors  $\partial\phi$  est maximal monotone. En effet d'après le théorème 1.1 pour  $y \in H$ , donné, le problème

$$(2-2) \quad \inf_{v \in H} \left[ \frac{1}{2\lambda} |v-y|^2 + \phi(v) \right]$$

admet une solution unique. On conclut à l'aide du Lemme suivant que  $u + \lambda \partial\phi(u) \ni y$  admet une solution pour tout  $\lambda > 0$  et donc que  $\partial\phi$  est maximal monotone. #

Lemme 2.1 : Si  $\phi$  est convexe et  $\lambda > 0$ , u solution de (2-2) équivaut à  $u + \lambda \partial\phi(u) \ni y$ .

En effet, si  $\frac{1}{\lambda} (y-u) \in \partial\phi(u)$ , on a  $\phi(u) < +\infty$  et

$$\phi(v) - \phi(u) \geq \frac{1}{\lambda} (y-u, v-u) \geq \frac{1}{2\lambda} |v-y|^2 - \frac{1}{2\lambda} |u-y|^2, \quad \forall v \in H.$$

Réciproquement, en prenant  $w = (1-\theta)u + \theta v$ ; avec  $0 < \theta < 1$ , on a par convexité

$$\begin{aligned} \theta(\phi(v) - \phi(u)) &\geq \phi(w) - \phi(u) \geq \frac{1}{2\lambda} |u-y|^2 - \frac{1}{2\lambda} |w-y|^2 \\ &= \frac{1}{2\lambda} |u-y|^2 - \frac{1}{2\lambda} |(1-\theta)u + \theta v - y|^2 \end{aligned}$$

en divisant par  $\theta$  et en faisant  $\theta \rightarrow 0$ , on obtient

$$\phi(v) - \phi(u) \geq \frac{1}{\lambda} (y-u, v-u) . \quad \#$$

Exemple 2.2 : Soit  $V \subset H \subset V'$  un espace de Banach réflexif dense dans  $H$ , de norme  $\|\cdot\|$ . Soit  $A : V \rightarrow V'$  un opérateur monotone univoque défini sur  $V$ , hémicontinu et coercif :

$$(2-3) \quad \frac{\langle Au, u \rangle}{\|u\|} \rightarrow +\infty \quad \text{si} \quad \|u\| \rightarrow +\infty$$

Alors l'opérateur  $A_H$  restriction de  $A$  à  $H$  de domaine

$$D(A_H) = \{v \in V : Av \in H\}$$

est maximal monotone.

Ceci résulte d'un théorème de MINTY <sup>(1)</sup> : l'équation  $u + \lambda Au = y$  admet une solution unique pour tout  $y \in V'$ , d'où si  $y \in H$ ,  $Au \in H$ , qui entraîne  $R(I + \lambda A_H) = H$ . #

### 2.3 Algorithmes pour résoudre $Au \ni 0$

Si  $A$  est univoque, on peut utiliser l'algorithme suivant

$$(2-4) \quad u^{n+1} = (I - \lambda A)u^n.$$

(1) cf. LIONS, *loc. cit.* p. 171.

Cet algorithme est en fait un cas particulier de l'algorithme envisagé au théorème 2. On voit que dans les mêmes conditions : si A est coercif et lipschitzien:

$$(Av - Au, v-u) \geq \alpha |v-u|^2$$
$$|Av - Au|^2 \leq M |v-u|^2$$

alors  $I-\lambda A$  est une contraction stricte pour  $0 < \lambda < \frac{\alpha}{M^2}$ , et donc que la suite  $u^n$  générée par l'algorithme (2-4) converge fortement vers u solution (unique en l'occurrence) de  $Au = 0$ .

Lorsque  $A = \partial\phi$  cet algorithme est connu sous le nom d'algorithme de gradient.

Mais évidemment les hypothèses assurant sa convergence sont très fortes. D'autre part si  $\lambda$  est trop grand, l'algorithme (2-4) est instable : il se produit numériquement une croissance exponentielle de la solution produisant des dépassements de capacité. Cette propriété est bien connue pour les schémas explicites de résolution des équations différentielles, ce qui est justement le cas : l'algorithme

$$\frac{u^{n+1} - u^n}{\lambda} + Au^n = 0$$

visiblement équivalent à (2-4), est bien un schéma explicite pour résoudre l'équation différentielle

$$\frac{du}{dt} + Au = 0$$

$$u(0) = u^0. \quad (\text{cf. Partie C, § 3})$$

Le schéma implicite donne l'algorithme suivant :

$$(2-5) \quad u^{n+1} = J_A^\lambda u^n$$

qui a quant à lui l'avantage d'être défini même pour A multivoque. On va voir qu'il converge sous des hypothèses beaucoup plus faibles.

### 2.4 Contractions fermes

Définition 2.1 : Soit  $C \subset H$  un convexe fermé et  $T : C \rightarrow C$  un opérateur univoque. On dit que  $T$  est une contraction ferme si et seulement si

$$(2-6) \quad |Tx - Ty|^2 \leq (Tx - Ty, x - y) \quad \forall x, y \in C. \quad \#$$

Une contraction ferme est en particulier une contraction, mais il y a un peu mieux. Supposons que  $y = Ty$  soit un point fixe de  $T$ , alors si  $T$  est une contraction,  $Tx$  appartient à la boule de centre  $y$  et de rayon  $|x - y|$ . Mais si  $T$  est une contraction ferme,  $Tx$  appartient à la boule de centre  $\frac{x+y}{2}$  et de rayon  $\frac{1}{2}|x - y|$  (donc de diamètre  $|x - y|$ ).

Exemple : Si  $A$  est maximal monotone, sa résolvante  $J_A^\lambda$  est une contraction ferme (c'est ce qu'on a en fait démontré au Corollaire 2.1). En particulier si  $A = \partial I_K$ , où  $K \subset H$  est un convexe fermé, on a  $J_A^\lambda = P_K$  projection sur  $K$  qui est aussi contraction ferme.  $\#$

Le résultat suivant montre l'intérêt de la notion de contraction ferme. On note, dans ce qui suit,  $F(T)$  l'ensemble des points fixes de  $T$  <sup>(1)</sup>.

Théorème 2.1 : Soit  $T : H \rightarrow H$  une contraction ferme. Si  $F(T)$  est non vide, alors

$$T^n x \rightarrow \xi \in F(T)$$

lorsque  $n \rightarrow \infty$ .

Démonstration :

Soit  $y \in F(T)$  et  $x^n \equiv T^n x^0$  pour  $n \geq 0$ . On a d'après (2-6)

$$|x^{n+1} - y|^2 \leq (x^{n+1} - y, x^n - y) = \frac{1}{2} |x^{n+1} - y|^2 + \frac{1}{2} |x^n - y|^2 - \frac{1}{2} |x^{n+1} - x^n|^2$$

Soit

$$(2-7) \quad \frac{1}{2} |x^{n+1} - y|^2 + \frac{1}{2} |x^{n+1} - x^n|^2 \leq \frac{1}{2} |x^n - y|^2$$

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

(1)  $F(T)$  est fermé et convexe, BREZIS, loc. cit.

ce qui prouve que  $|x^{n+1}-x^n| \rightarrow 0$  lorsque  $n \rightarrow \infty$  et que  $\{x^n\}$  est bornée. On en conclut que  $x_n \rightharpoonup \xi \in F(T)$  à l'aide du lemme suivant qui sera réutilisé par la suite. #

Lemme 2.2 : Soit T une contraction de H telle que  $F(T) \neq \emptyset$  et que  $T^{n+1}x - T^n x \rightarrow 0$  lorsque  $n \rightarrow \infty$  alors  $T^n x \rightharpoonup \xi \in F(T)$ .

Démonstration

On pose  $x^n \equiv T^n x$ ,  $\{x^n\}$  étant bornée on en extrait une sous-suite faiblement convergente  $x^{n'} \rightharpoonup \eta$ . On a

$$|Tx^{n'} - T\eta|^2 \leq |x^{n'} - \eta|^2 \quad \text{soit}$$

$$(x^{n'+1} - x^{n'} + \eta - T\eta, x^{n'+1} + x^{n'} - T\eta - \eta) \leq 0$$

d'où à la limite

$$(\eta - T\eta, \eta + \eta - T\eta - \eta) \leq 0$$

qui montre que  $\eta \in F(T)$ . Comme  $|x^n - y|$  est une suite décroissante pour tout  $y \in F(T)$  on en conclut que  $x^n \rightharpoonup \xi \in F(T)$  à l'aide du lemme d'OPIAL suivant.

#

Lemme d'OPIAL :

Soit H un espace de Hilbert  $\{x_n\} \in H$  une suite bornée et  $F \subset H$  un ensemble non vide. On suppose que

(i) Toute suite  $x_{n'} \rightarrow x$  a sa limite  $x \in F$ ,

(ii)  $\forall f \in F, |x_n - f|$  converge vers un nombre réel, soit  $p(f) < +\infty$

Alors la suite  $x_n$  tout entière converge faiblement vers un élément de F.

Démonstration

Il suffit de montrer que

$$x_{n'} \rightarrow f_1, x_{m'} \rightarrow f_2 \quad \text{alors } f_1 = f_2.$$

On a  $\lim_{n \rightarrow \infty} |x_n - f_2|^2 = (p(f_2))^2$  et

$$|x_n - f_2|^2 = |x_n - f_1|^2 + 2(x_n - f_1, f_1 - f_2) + |f_1 - f_2|^2$$

en passant à la limite suivant  $n' : x_{n'} \rightarrow f_1$  d'où

$$(p(f_2))^2 = (p(f_1))^2 + |f_1 - f_2|^2 \quad ;$$

et en passant à la limite suivant  $m' : x_{m'} \rightarrow f_2$ , d'où

$$(p(f_2))^2 = (p(f_1))^2 - |f_1 - f_2|^2. \quad \#$$

### COMMENTAIRE

En dimension finie, la convergence est forte. On voit que la propriété d'être une contraction ferme exclut les rotations. Le fait d'avoir un point fixe exclut les translations. La conclusion du théorème est que si on exclut ces deux sources de contre-exemples, il y a bien convergence des itérés de la contraction vers un point fixe.  $\#$

### Généralisation

On dit que  $T$  est une contraction quasi ferme si  $T$  vérifie

$$(2-8) \quad |Tx - Ty|^2 \leq (1-\theta)(Tx - Ty, x - y) + \theta|x - y|^2 \quad \forall x, y \in C$$

pour  $0 < \theta < 1$ . Dans ce cas, le résultat du théorème 2.1 est encore applicable à  $T$ .

En effet dans ce cas, on aura

$$\begin{aligned} |x^{n+1} - y|^2 &\leq (1-\theta)(x^{n+1} - y, x^n - y) + \theta|x^n - y|^2 \\ &\leq \frac{1+\theta}{2} |x^n - y|^2 + \frac{1-\theta}{2} |x^{n+1} - y|^2 - \frac{1-\theta}{2} |x^{n+1} - x^n|^2 \\ \frac{1+\theta}{2} |x^{n+1} - y|^2 + \frac{1-\theta}{2} |x^{n+1} - x^n|^2 &\leq \frac{1+\theta}{2} |x^n - y|^2 \end{aligned}$$

à la place de (2-7) mais on en conclut toujours que

$$|x^{n+1} - x^n| \rightarrow 0 \quad \text{si } 0 < \theta < 1. \quad \#$$

Lemme 2.3 : 1°)- Si Q est une contraction ferme 2Q-I est une contraction  
 2°)- Si S est une contraction quelconque et A une contraction  
ferme alors T = θS + (1-θ)Q est quasi ferme pour 0 < θ < 1,  
et vérifie (2-8).

Démonstration

1°)- Soit  $y_i = 2Qx_i - x_i$  on a

$$\begin{aligned} |y_2 - y_1|^2 &= 4|Qx_2 - Qx_1|^2 - 4(Qx_2 - Qx_1, x_2 - x_1) + |x_2 - x_1|^2 \\ &\leq |x_2 - x_1|^2 \end{aligned}$$

puisque Q est une contraction ferme. #

2°)- Si  $T = \theta S + (1-\theta)Q$  on a

$$\begin{aligned} |Tx - Ty|^2 &= \theta^2 |Sx - Sy|^2 + 2\theta(1-\theta)(Sx - Sy, Qx - Qy) + (1-\theta)^2 (Qx - Qy)^2 \\ &\leq \theta^2 |Sx - Sy|^2 + \theta(1-\theta)(Sx - Sy, (2Q - I)x - (2Q - I)y) \\ &\quad + \theta(1-\theta)(Sx - Sy, x - y) + (1-\theta)^2 (Qx - Qy, x - y) \\ &= \theta(Sx - Sy, Rx - Ry) + (1-\theta)(Tx - Ty, x - y) \end{aligned}$$

ou l'on a posé  $R = \theta S + (1-\theta)(2Q - I)$  ;

R est une construction d'après la 1ère partie comme combinaison convexe entre 2 contractions. Donc

$$(Sx - Sy, Rx - Ry) \leq |Sx - Sy| |Rx - Ry| \leq |x - y|^2$$

d'où (2-8). #

Interprétation géométrique

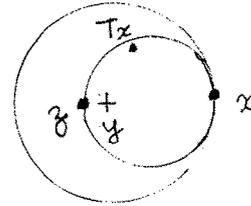
Soit  $y = Ty$  un point fixe de T vérifiant (2-8), alors x étant donné Tx appartient à la boule de diamètre x, z où  $z = y + \theta(y - x)$  ;

En effet (2-8) entraîne

$$(Tx - Ty + \theta(x - y), Ty - Tx + x - y) \geq 0$$

d'où  $(Tx - z, x - Tx) \geq 0$

dans ce cas où  $y = Ty$ . #



### 2.5 Application

1°)- Si le problème  $Au \ni 0$  a une solution et si  $A$  est maximal monotone l'algorithme  $u^{n+1} = J_A^\lambda u^n$  converge (pour tout  $\lambda > 0$ ) faiblement vers une solution du problème.

En effet, les points fixes de  $J_A$  sont visiblement les solutions de  $Au \ni 0$ . Cet algorithme est aussi connu sous le nom de "Proximal point algorithm", étudié par ROCKAFELLAR <sup>(1)</sup>. Malgré ses agréables propriétés de convergence, il n'est pas très utilisé, excepté dans le cas linéaire, car pour calculer  $u^{n+1}$  à partir de  $u^n$  il faut résoudre  $u^{n+1} + \lambda Au^{n+1} \ni u^n$  qui est en général presque aussi difficile à résoudre que le problème stationnaire.

Nous reviendrons là-dessus, à propos des problèmes d'évolution, dans la partie C.

2°)- Retour au schéma explicite. Si  $A$  est lipschitzien et  $A^{-1}$  coercif c'est-à-dire

$$(2-9) \quad (Au - Av, u - v) \geq \alpha |Au - Av|^2 \quad \forall u, v \in H$$

alors l'algorithme (2-4) converge pour  $0 < \lambda < 2\alpha$ .

En effet, soit  $T = I - \lambda A$ , on a

$$(2-10) \quad |Tu - Tv|^2 \leq |u - v|^2 + \left(\frac{\lambda^2}{\alpha} - 2\lambda\right) (Au - Av, u - v) \\ = (1 - \theta^2) |u - v|^2 + \theta^2 (Tu - Tv, u, v)$$

où  $\theta^2 = 2 - \frac{\lambda}{\alpha}$ . Pour  $\alpha < \lambda < 2\alpha$  on obtient donc (2-8).

(1) *SIAM J. Control Optimization* 14, (1976) 877-898.

Pour  $0 < \lambda \leq \alpha$ ,  $T$  est en fait une contraction ferme. En effet, dans ce cas  $\theta^2 > 1$  on écrit alors dans (2-10)

$$\theta^2(Tu-Tv, u-v) \leq (Tu-Tv, u-v) + (\theta^2-1)|u-v|^2. \quad \#$$

Remarque 2.2 : Cas  $V \subset H \subset V'$

Revenons à la situation de l'exemple 2.2 où  $A : V \rightarrow V'$  est un opérateur borné non linéaire. L'algorithme (2-4) est alors inapplicable. On peut alors utiliser un opérateur auxiliaire linéaire de préférence  $B : V \rightarrow V'$  et envisager le schéma itératif

$$(2-11) \quad Bu^{n+1} = Bu^n - \lambda Au^n$$

(lorsque  $A = \partial\phi$  cet algorithme est connu sous le nom d'algorithme de gradient avec opérateur auxiliaire).

Cet algorithme revient tout simplement à résoudre le problème  $B^{-1}Au = 0$  avec un schéma explicite, qui peut donner d'excellents résultats si  $B^{-1}A$  a de bonnes propriétés en particulier de lipschitz-continuité.

Numériquement cette remarque est très importante. Car bien que beaucoup d'opérateurs non bornés soient lipschitziens une fois discrétisés la constante de lipschitz est alors

inversement proportionnelle au pas de discrétisation <sup>(1)</sup> et nécessitent l'utilisation de très petites valeurs de  $\lambda$ , alors que l'introduction d'un opérateur auxiliaire  $B$  peut permettre de prendre de bien plus grandes valeurs de  $\lambda$ .

On retiendra la philosophie suivante :

- Si on peut démontrer la convergence d'un algorithme dans le cas continu, alors, si le problème est "bien" discrétisé, la vitesse de convergence de l'algorithme appliqué au problème discrétisé ne dépend pas (ou "pas trop") du pas de discrétisation. (cf. § 5).

<sup>(1)</sup> au carré du pas de discrétisation dans le cas d'un problème elliptique d'ordre 2.

### 3. SOMME D'OPERATEURS MAXIMAUX MONOTONES

#### 3.1. Difficulté

Soient A et B deux opérateurs maximaux monotones sur H, en général A+B n'est pas maximal monotone. En effet, on peut très bien avoir  $D(A) \cap D(B) = \emptyset$

De même si  $\phi$  et  $\psi$  sont deux fonctions convexes s.c.i. de H à valeurs dans  $\mathbb{R} \cup \{+\infty\}$ , on a

$$\partial\phi + \partial\psi \subset \partial(\phi+\psi)$$

mais l'inclusion est stricte en général, et donc  $\partial\phi + \partial\psi$  n'est pas en général maximal car  $\partial(\phi+\psi)$  est maximal monotone, au moins si  $D(\phi) \cap D(\psi) \neq \emptyset$ .

Un exemple concret est le cas où  $H = L^2(\mathbb{R}^n)$  et  $K = \{u \in H : u=0 \text{ sur } \Omega^c\}$  où  $\Omega^c$  est le complémentaire d'un ouvert  $\Omega \subset \mathbb{R}^n$ . Considérons

$$\phi(v) = \int_{\Omega} |\nabla v|^2 \, dx \quad \psi = \partial I_K$$

on a  $D(\phi+\psi) = H^1_0(\Omega)$

et  $D(\partial(\phi+\psi)) = H^2(\Omega) \cap H^1_0(\Omega)$

mais  $D(\partial\phi) \cap D(\partial\psi) = H^2_0(\Omega)$

donc  $\partial\phi + \partial\psi \neq \partial(\phi+\psi)$  .

On remarquera dans cet exemple l'importance du domaine d'un opérateur maximal monotone.

Il est, en général, plus difficile de montrer que  $0 \in \partial\phi(u) + \partial\psi(u)$  a une solution que de démontrer que  $\inf_{v \in H} [\phi(v) + \psi(v)]$  en a une, problème pour lequel on peut utiliser le théorème 1.1.

#### 3.2. Cas où B est lipschitzien

Théorème 3.1 : Si B est lipschitzien, alors A+B est maximal monotone

Démonstration : D'après la définition de la maximale monotonie, il est clair que  $\lambda(A+B)$  maximal monotone pour  $\lambda > 0 \Rightarrow A+B$  maximal monotone. Montrons que  $R(I+\lambda(A+B)) = H$  pour  $\lambda$  assez petit. Soit donc  $y \in H$  donné, résoudre  $x+\lambda(A(x)+B(x)) \ni y$  équivaut à résoudre

$$x = J_A^\lambda(y-\lambda Bx).$$

Or  $x \rightarrow J_A^\lambda(y-\lambda Bx)$  est une contraction stricte pour  $\lambda$  assez petit puisque  $B$  est lipschitzien.

Corollaire 3.1. : Si  $A$  est coercif et  $B$  lipschitzien le problème  $Au+Bu \ni 0$  admet une solution unique.

En effet  $A$  coercif équivaut à  $A-\alpha I$  monotone, et même maximal monotone pour  $\alpha > 0$  donné égal à la constante de coercivité de  $A$  ;  $A-\alpha I+B$  est donc maximal monotone d'après le théorème 3.1, et donc  $R(A+B) = H$ , d'où l'existence d'une solution. Supposons maintenant qu'il y ait deux solutions  $u_i, i=1,2$ , on a donc  $0 = a_i+Bu_i$  avec  $a_i \in A(u_i)$ . On a

$$\alpha |u_2-u_1|^2 \leq (a_2-a_1, u_2-u_1) \leq 0$$

d'après la monotonie de  $B$  d'où l'unicité.

### 3.3. Algorithme I

Dans le cas où  $B$  est lipschitzien, on peut utiliser l'algorithme (qui sera appelé Algorithme I) :

$$u^{n+1} = J_A^\lambda (I-\lambda B)u^n$$

pour résoudre  $Au + Bu \ni 0$ . En effet, un point fixe de  $J_A^\lambda(I-\lambda B)$  équivaut à

$$u + \lambda Au \ni u - \lambda Bu .$$

L'algorithme I peut être interprété comme un algorithme de décomposition obtenu en composant un pas explicite sur  $B$ , avec un pas implicite sur  $A$ .

Si B est coercif, alors  $I-\lambda B$  est une contraction stricte (cf § 2.3) pour  $\lambda$  assez petit et donc aussi  $J_A^\lambda(I-\lambda B)$ . L'algorithme I génère alors une suite  $u^n$  convergeant fortement vers la solution du problème.

Si  $B^{-1}$  est coercif, alors  $I-\lambda B$  est une contraction ferme pour  $\lambda$  assez petit (cf § 2.5) et  $J_A^\lambda(I-\lambda B)$  est donc le produit de deux contractions fermes. On en conclut que si le problème

$$Au + Bu \geq 0$$

admet des solutions, la suite  $u_n$  générée par l'algorithme I converge faiblement vers une de ces solutions, à l'aide du résultat suivant :

Théorème 3.2 : Soit Q et S deux contraction fermes, si  $T=QS$  admet un point fixe, alors  $T^n x \rightarrow \xi \in F(T)$  lorsque  $n \rightarrow +\infty$ .

Il suffit en effet d'après le lemme 2.2 de montrer que  $x^{n+1} - x^n \rightarrow 0$  où  $x^n \equiv Tx^n$ . Soit  $y \in F(T)$  on a

$$|x^{n+1}-y|^2 \leq (x^{n+1}-y, Sx^n-Sy) = \frac{1}{2} |x^{n+1}-y|^2 + \frac{1}{2} |Sx^n-Sy|^2 - \frac{1}{2} |x^{n+1}-y-Sx^n+Sy|^2$$

de même on a

$$|Sx^n-Sy|^2 \leq (Sx^n-Sy, x^n-y) = \frac{1}{2} |Sx^n-Sy|^2 + \frac{1}{2} |x^n-y|^2 - \frac{1}{2} |Sx^n-Sy-x^n+y|^2$$

d'où par addition

$$\frac{1}{2} |x^{n+1}-y|^2 + \frac{1}{2} |x^{n+1}-y-Sx^n+Sy|^2 + \frac{1}{2} |Sx^n-Sy-x^n+y|^2 \leq \frac{1}{2} |x^n-y|^2.$$

On en déduit par sommation de  $n=0$  à  $+\infty$  que

$$x^{n+1}-y-Sx^n+Sy \rightarrow 0$$

$$Sx^n-Sy-x^n+y \rightarrow 0$$

$$\Rightarrow (\text{par addition}) x^{n+1}-x^n \rightarrow 0. \quad \#$$

Remarque 3.1

1°)- La démonstration s'adapte facilement au cas où S vérifie seulement la relation (2-8) du § 2.4 (c'est-à-dire au cas où S est une combinaison convexe entre une contraction et une contraction ferme).  
 Dans le cas où  $B^{-1}$  est coercif :

$$(Bv-Bw, v-w) \geq \alpha |Bv-Bw|^2$$

cette remarque permet comme § 2.5 d'étendre la convergence (faible) de l'algorithme I à  $0 < \lambda < 2\alpha$ .

Remarquons aussi qu'il y a la convergence forte : de  $Sx^n - x^n$  vers  $Sy - y$ .  
 En revenant à B et  $u^n$  cela montre que  $Bu^n \rightarrow Bu$  si  $B^{-1}$  est coercif, ce qui peut d'ailleurs être démontré directement ainsi que l'unicité de Bu.

2°)- Vitesse de convergence : Dans le cas où B est coercif et lipschitzien :

$$|Bx_1 - Bx_2| \leq M |x_1 - x_2|, \quad \forall x_1, x_2 \in H,$$

$$(Bx_1 - Bx_2, x_1 - x_2) \geq \alpha |x_1 - x_2|^2, \quad \forall x_1, x_2 \in H,$$

Le calcul effectué lors de la démonstration du théorème 1.2 montre que

$$|u^{n+1} - u|^2 \leq (1 - 2\lambda\alpha + \lambda^2 M^2) |u^n - u|^2$$

le choix optimal est donc  $\lambda = \frac{\alpha}{M^2}$  pour lequel on a

$$|u^{n+1} - u| \leq \left(1 - \frac{\alpha^2}{M^2}\right)^{1/2} |u^n - u|$$

la vitesse de convergence est donc  $\left(1 - \frac{\alpha^2}{M^2}\right)^{1/2}$  c'est-à-dire que

$$|u^n - u| \leq \left(1 - \frac{\alpha^2}{M^2}\right)^{n/2} |u^0 - u|$$

3.4. Régularisation

Nous allons maintenant nous intéresser à l'algorithme II

$$(3-1) \quad u^{n+1} = J_A^\lambda J_B^\lambda u^n$$

qui à l'avantage d'être bien défini (et stable) pour A et B maximaux monotones quelconques.

On appelle régularisé Yosida de B l'opérateur (univoque)

$$B_\lambda \equiv \frac{1}{\lambda}(I - J_B^\lambda)$$

qui est visiblement lipschitzien de Rapport  $\frac{1}{\lambda}$ .

On remarque alors que les points fixes de  $J_A^\lambda J_B^\lambda$  sont solutions du problème "régularisé"

$$(3-2) \quad Au_\lambda + B_\lambda u_\lambda \ni 0$$

dont les solutions dépendent de  $\lambda$ .

(En effet (3-1) n'est autre que l'algorithme I appliqué au problème (3-2)).

On va donc étudier le problème régularisé et en particulier la convergence de  $u_\lambda$  lorsque  $\lambda \rightarrow 0$ .

On remarque en premier lieu qu'il y a une certaine équivalence entre le problème (3-2) où B a été régularisé et le problème

$$(3-3) \quad A_\lambda e_\lambda + B e_\lambda \ni 0$$

où c'est A qui a été régularisé.

En effet si on pose  $e_\lambda = J_B^\lambda u_\lambda$ , on aura  $\frac{1}{\lambda}(e_\lambda - u) + B e_\lambda \ni 0$ .  
D'autre part  $u_\lambda$  étant point fixe de  $J_A^\lambda J_B^\lambda$  on a  $u_\lambda = J_A^\lambda e_\lambda$  d'où (3-3).

Les problèmes (3-2) et (3-3) sont donc équivalents, en ce sens que si l'on connaît la solution de l'un on en déduit facilement la solution de l'autre.

Le théorème 3.2 montre la convergence faible de l'algorithme II :  
lorsque  $n \rightarrow \infty$  on a  $u_n \rightarrow u_\lambda$  solution de (3-2).

On remarquera que  $e^n = J_B^\lambda u_n$  peut être obtenu en itérant

$$e^{n+1} = J_B^\lambda J_A^\lambda e^n$$

et l'on a aussi  $e^n \rightarrow e_\lambda$  solution de (3-3); c'est-à-dire point fixe de  $J_B^\lambda J_A^\lambda$ .

On va maintenant étudier la convergence de  $u_\lambda$  vers  $u$  solution de  $Au + Bu \ni 0$  lorsque  $\lambda \rightarrow 0$ .

Théorème 3.3 : On suppose que  $A$  est coercif : il existe  $\alpha > 0$  et  $A_0$  maximal monotone tel que  $A = \alpha I + A_0$ . Les conditions suivantes sont équivalentes

(i)  $Au + Bu \ni 0$  admet une solution;

(ii)  $|B_\lambda u_\lambda| \leq C$  indépendamment de  $\lambda$ ,

de plus, on a  $|u - u_\lambda| \leq C \sqrt{\lambda}$ .

#

(i)  $\Rightarrow$  (ii).

Par hypothèse, il existe  $a \in Au$ ,  $b \in Bu$  tels que  $a + b = 0$ , et  $a_\lambda \in Au_\lambda$  tel que  $a_\lambda + b_\lambda = 0$  où  $b_\lambda \equiv B_\lambda u_\lambda$ .

Soit  $e_\lambda \equiv J_B^\lambda u_\lambda$ , on va montrer que  $b_\lambda \in Be_\lambda$ .

En effet on a  $b_\lambda = \frac{1}{\lambda}(u_\lambda - e_\lambda)$ , et  $u_\lambda = e_\lambda + \lambda \beta_\lambda$  où  $\beta_\lambda \in Be_\lambda$  d'où  $b_\lambda = \beta_\lambda \in Be_\lambda$ .

En appliquant alors la monotonie de  $B$  on obtient

$$(b_\lambda - b, e_\lambda - u) \geq 0$$

$$\Rightarrow (b_\lambda + a, e_\lambda - u) + (a - a_\lambda, u_\lambda - u) \geq 0$$

$$\Rightarrow (a - a_\lambda, u - u_\lambda) + \lambda |b_\lambda|^2 \leq -\lambda (a, b_\lambda)$$

d'où  $|b_\lambda| \leq |a|$ , en utilisant seulement la monotonie de  $A$ .

Si  $A$  est de plus coercif, on obtient

$$|u - u_\lambda|^2 \leq \frac{\lambda}{\alpha} |a|^2$$

ce qui fournit une évaluation de la constante  $C$  intervenant dans l'estimation d'erreur.

#

(ii)  $\Rightarrow$  (i)

On écrit (3-2) sous la forme  $\alpha u_\lambda + A_0 u_\lambda + B_\lambda u_\lambda \ni 0$

qui montre que  $\xi_\lambda \equiv -\alpha u_\lambda - b_\lambda \in A_0 u_\lambda$ . En appliquant la monotonie de  $A_0$  (c'est-à-dire la coercivité de A) on obtient

$$(3-4) \quad (\xi_\lambda - \xi_\mu, u_\lambda - u_\mu) = (\alpha u_\mu + b_\mu - \alpha u_\lambda - b_\lambda, u_\lambda - u_\mu) \geq 0$$

c'est-à-dire

$$\alpha |u_\lambda - u_\mu|^2 \leq (b_\mu - b_\lambda, u_\lambda - u_\mu) = (b_\mu - b_\lambda, u_\lambda - e_\lambda - u_\mu + e_\mu) + (b_\mu - b_\lambda, e_\lambda - e_\mu)$$

mais le dernier terme est négatif d'après la monotonie de B puisque  $b_\lambda \in B e_\lambda$  et  $b_\mu \in B e_\mu$ . On obtient

$$\alpha |u_\lambda - u_\mu|^2 \leq (b_\mu - b_\lambda, \lambda b_\lambda - \mu b_\mu)$$

ce qui montre que la suite  $u_\lambda$  est de Cauchy puisque  $|b_\lambda| \leq C$ .

On a donc  $u_\lambda \rightarrow u$  quand  $\lambda \rightarrow 0$ . Comme  $b_\lambda$  est borné, il existe une suite  $\lambda_n \rightarrow 0$  lorsque  $n \rightarrow \infty$  telle que  $b_{\lambda_n} \rightarrow b$ . Comme  $-b_\lambda \in A u_\lambda$ , on a

$$(-b_\lambda - z, u_\lambda - x) \geq 0, \quad \forall \{x, z\} \in A,$$

d'où à la limite

$$(-b - z, u - x) \geq 0, \quad \forall \{x, z\} \in A,$$

qui montre que  $-b \in A u$  par maximale monotonie de A.

De même  $b_\lambda \in B e_\lambda$  s'écrit

$$(b_\lambda - z, e_\lambda - x) \geq 0 \quad \forall \{x, z\} \in B$$

et donnera à la limite que  $b \in B u$ . En effet  $e_\lambda \rightarrow u$  lorsque  $\lambda \rightarrow 0$  puisque  $J_B^\lambda u_\lambda = J_B^\lambda u_\lambda - u_\lambda + u_\lambda$  et que  $|J_B^\lambda u_\lambda - u_\lambda| \leq \lambda |b_\lambda|$  tend vers zéro lorsque  $\lambda \rightarrow 0$ . Finalement  $b \in (-A) \cap B u \Rightarrow A u + B u \ni 0$ , c'est-à-dire (i). #

Remarque 3.2 : La coercivité de A entraîne l'existence d'une solution au problème régularisé (corollaire 3.1), propriété qui suffit pour montrer que (i)  $\Rightarrow$  (ii) comme on l'a vu au cours de la démonstration, mais la coercivité est nécessaire pour l'estimation d'erreur. #

Remarque 3.3 : La démonstration précédente s'adapte immédiatement au problème

$$Au + Bu \ni y \quad \text{où } y \in H$$

(en fait équivalent au problème  $Au + Bu \ni 0$  à une translation près sur l'un des deux opérateurs). Il est équivalent de résoudre ce problème ou de démontrer la maximale monotonie de  $A_0 + B$ .

De fait, le théorème 3.3 est l'outil principal pour démontrer la maximale monotonie de la somme de deux opérateurs.

Plaçons-nous dans le cas où  $A_0 = \partial\phi$ .

Supposons que l'on ait une majoration du type

$$(3-5) \quad \phi(J_B^\lambda x) \leq \phi(x) + C_\lambda, \quad \forall x \in D(\phi) \text{ et } \lambda > 0,$$

alors (ii) est vérifié (et en particulier  $B + \partial\phi$  est maximal monotone).

En effet si  $u_\lambda$  est la solution de

$$u_\lambda + \partial\phi(u_\lambda) + B_\lambda u_\lambda \ni y$$

on a  $y - u_\lambda - b_\lambda \in \partial\phi(u_\lambda)$  où  $b_\lambda \equiv B_\lambda u_\lambda$  et donc

$$(3-6) \quad \phi(v) - \phi(u_\lambda) \geq (y - u_\lambda - b_\lambda, v - u_\lambda) \quad \forall v \in H$$

d'où en choisissant  $v = J_B^\lambda u_\lambda$  et en appliquant (3-5)

$$C_\lambda \geq (y - u_\lambda - b_\lambda, J_B^\lambda u_\lambda - u_\lambda) = (y - u_\lambda - b_\lambda, -\lambda b_\lambda)$$

d'où  $|b_\lambda|^2 \leq |b_\lambda| |u_\lambda - y| + C$

$$(3-7) \quad |b_\lambda| \leq |y - u_\lambda| + \sqrt{C}$$

On choisit  $v \in D(B) \cap D(\phi)$  dans (3-6), de sorte qu'il existe  $z \in Bv$  et

$$(z - b_\lambda, v - e_\lambda) \geq 0$$

d'où  $(b_\lambda, v - u_\lambda) = (b_\lambda, v - e_\lambda) - \lambda |b_\lambda|^2 \leq (z, v - e_\lambda)$

$$\begin{aligned} (u_\lambda - y, u_\lambda - v) + \phi(u_\lambda) &\leq \phi(v) + (b_\lambda, v - u_\lambda) \\ &\leq \phi(v) + (z, v - e_\lambda) \leq \phi(v) + |z| |v - e_\lambda| \end{aligned}$$

Posons maintenant  $x_\lambda = v + \lambda z$  d'où  $v = J_B^\lambda x_\lambda$  et

$$|v - e_\lambda| \leq |x_\lambda - u_\lambda| \leq C(1 + |u_\lambda|) \quad \text{d'où} \quad |u_\lambda| \leq C$$

et  $|b_\lambda| \leq C$ , d'après (3-7). #

Remarque 3.4 : Vitesse de convergence de l'algorithme II

Dans le cas où  $B$  est coercif (cf Remarque 3.1),  $J_B^\lambda$  est une contraction stricte de rapport  $(1 + 2\lambda\alpha)^{-1/2}$ , comme on le voit en adaptant la démonstration du corollaire 2.1.

On a donc

$$|u^n - u_\lambda| \leq |u^0 - u_\lambda| (1 + 2\lambda\alpha)^{-n/2}$$

et la convergence de l'algorithme est donc d'autant plus lente que  $\lambda$  est petit, et le nombre d'itérations nécessaire pour obtenir  $|u^n - u_\lambda| < \varepsilon$  où  $\varepsilon$  est donné est en  $O(\frac{1}{\lambda})$ , autrement dit le nombre d'itérations pour avoir  $d$  chiffres exacts est en  $O(d \cdot \frac{1}{\lambda})$  au lieu de  $O(d)$  dans l'algorithme I.

L'algorithme II est donc a priori un mauvais algorithme pour résoudre le problème stationnaire  $Au + Bu \ni 0$  ; la méthode de régularisation sera donc relativement inefficace si on ne dispose pas d'algorithme pour résoudre (3-2) dont la vitesse de convergence soit indépendante de  $\lambda$  ; ceci est malheureusement vrai sauf dans le cas linéaire.

On en gardera la philosophie que la régularisation ne sert souvent qu'à masquer la difficulté. #

Remarque 3.5. :

1°)- Si  $A = \partial\phi$ ,  $B = \partial\psi$ , on peut obtenir directement le couple  $\{u_\lambda, e_\lambda\}$  comme solution du problème d'optimisation

$$\text{Inf}_{v,e} [\phi(v) + \psi(e) + \frac{1}{2\lambda} |v-e|^2]$$

qui correspond au problème

$$\text{Inf}_{v=e} [\phi(v) + \psi(e)]$$

où la contrainte  $v-e=0$  est pénalisée. Ce dernier problème est visiblement équivalent au problème  $Au+Bu \ni 0$  qui s'écrit aussi

$$\text{Inf}_{v \in H} [\phi(v) + \psi(v)]$$

2°)- Si de plus  $\psi = I_K$  fonction indicatrice de  $K \subset H$ , convexe fermé,  $u_\lambda$  est solution de

$$\text{Inf}_{v \in H} [\phi(v) + \frac{1}{2\lambda} |v-P_K v|^2]$$

où la contrainte  $v \in K$  a été pénalisée. #

### 3.5. Applications

#### 1°) Problème de l'obstacle

Soit  $H = L^2(\Omega)$ . On pose

$$\psi(v) = \begin{cases} \frac{1}{2} \int_{\Omega} |\nabla v|^2 dx & \text{si } v \in H_0^1(\Omega), \\ +\infty, & \text{sinon} \end{cases}$$

On choisit  $\phi = I_K$  fonction indicatrice du convexe  $K$  :

$$K = \{v \in L^2(\Omega) : v \leq 0 \text{ p.p. sur } \Omega\}$$

Le problème d'optimisation

$$(3-8) \quad \inf_{v \in H} [\psi(v) + \phi(v) - (f, v)]$$

a visiblement une solution unique  $u$ . En effet, l'inégalité de Poincaré s'écrit

$$\psi(v) \geq \alpha |v|^2 \quad \forall v \in D(\psi),$$

et montre la coercivité de  $B = \partial\psi$  (noter que  $B = -\Delta$  et  $D(B) = H^2 \cap H_0^1(\Omega)$ ).

Pour montrer la maximale monotonie de  $B + \partial\phi$ , on applique la relation (3-5) introduite à la remarque 3.3.

Soit  $u = J_B^\lambda v$ ,  $u$  vérifie

$$\begin{cases} -\Delta u + \lambda u = v & \text{sur } \Omega \\ u = 0 & \text{sur } \Gamma \end{cases}$$

et d'après le principe du maximum  $v \geq 0$  p.p.  $\Rightarrow u \geq 0$ , p.p. donc  $v \in K \Rightarrow J_B^\lambda v \in K$ , et (3-5) est vérifiée avec  $C = 0$ .

On en déduit que  $\partial(\phi + \psi) = \partial\phi + \partial\psi$ , et que  $u$  vérifie

$$Bu + \partial I_K(u) \ni f$$

On a ainsi démontré que  $-\Delta u \in L^2(\Omega)$  et donc que  $u \in H^2(\Omega)$ , ce qui est un résultat de régularité.

Pour ce qui est des algorithmes pour résoudre (3-8), on remarque que (2-4) est inaplicable puisque  $A = B + \partial I_K$  est multivoque et (2-5) est aussi difficile à résoudre (à chaque itération) que le problème que l'on veut résoudre.

Il est donc souhaitable d'utiliser la décomposition de  $A$  en deux opérateurs. Or

$$J_{\partial\phi}^\lambda = P_K$$

n'est autre que la  $L^2$ -projection sur  $K$  et se réduit donc à une simple troncature de la fonction à projeter.

L'algorithme I s'écrira

$$(3-9) \quad u^{n+1} = P_K((I-\lambda B)u^n)$$

et n'est pas directement applicable dans le cas continu, puisque B est non borné. On peut l'appliquer dans le cas discret, et il y aura une condition de stabilité sévère sur  $\lambda$  en fonction du pas de discrétisation, comme on le verra par la suite.

Noter que (3-9) est l'algorithme de gradient avec projection pour résoudre (3-8).

Enfin l'algorithme II est applicable dans le cas continu et à chaque pas de l'algorithme on aura à résoudre un problème linéaire pour calculer  $J_B^\lambda u^n$ , puis à effectuer une troncature pour calculer  $u^{n+1} = P_K J_B^\lambda u^n$ . Sur le plan numérique, on rencontrera l'inconvénient signalé à la remarque 3.4.

On verra par la suite que les algorithmes III et IV décrits plus loin donnent de bien meilleurs résultats que I et II.

## 2°) Problème de Mossolov

On va tout d'abord évaluer la sous différentielle de la fonction  $j(v) = \int_{\Omega} |\nabla v| dx$ .  
Soit  $V = H_0^1(\Omega)$  et  $Y = (L^2(\Omega))^n$ ,  $j$  est de la forme

$$j(v) = j_0(\Lambda v)$$

où  $\Lambda : V \rightarrow Y$  désigne l'opérateur gradient, et

$$j_0(p) = \int_{\Omega} |p| dx.$$

D'après EKELAND-TEMAM, loc. Cit., p. 27, on a

$$\partial j(u) = \Lambda^* \partial j_0(\Lambda u)$$

où  $\Lambda^*$  désigne l'opérateur  $Y \rightarrow V'$  transposé de  $\Lambda$ , c'est-à-dire

$$\partial j(u) = -\text{div } \partial j_0(\nabla u) .$$

Soit maintenant  $H = L^2(\Omega)$ , on choisit  $\psi$  et  $B = \partial\psi$  comme dans l'exemple précédent.

Le problème à résoudre est

$$(3-10) \quad \inf_{v \in H} [\psi(v) + j(v) - (f, v)]$$

Pour mettre ce problème sous la forme d'une somme d'opérateurs

$$(3-11) \quad Au + Bu \ni f$$

avec  $A = \partial j$ , il faut démontrer que  $B + \partial j$  est maximal monotone ou, ce qui est équivalent, montrer que la solution  $u$  du problème (3-10) vérifie le résultat de régularité  $-\Delta u \in L^2(\Omega)$ , c'est-à-dire  $u \in H^2(\Omega)$ .

Pour ce faire, on utilise encore la Remarque 3.3, mais la démonstration de la relation (3-5) est assez "technique", cf BREZIS<sup>(1)</sup>, p. 118.

Mais la décomposition du problème sous la forme (3-11) n'est pas intéressante sur le plan algorithmique car  $(I + \lambda \partial j)^{-1}v$  est difficile à évaluer.

Il vaut mieux choisir  $H = (L^2(\Omega))^n$  et  $R = \text{Im } \Lambda$  l'image de l'opérateur  $\Lambda^*$  qui est fermée d'après l'inégalité de Poincaré. Comme  $\Lambda$  est injectif,  $\Lambda^*$  est surjectif, et il existe  $\sigma_0 \in H$  tel que  $(f, v) = (\sigma_0, \Lambda v) \quad \forall v \in V$ .

Le problème (3-10) est donc équivalent au problème

$$\inf_{e \in R} \left[ \frac{1}{2} |e|^2 + j_0(e) - (\sigma_0, e) \right]$$

dont la solution optimale est  $\Lambda u$  où  $u$  est solution de (3-10).

On pose alors

$$\phi(e) = \frac{1}{2} |e|^2 + j_0(e) - (\sigma_0, e)$$

et  $\psi = I_R$ .

D'après EKELAND-TEMAM, loc. cit., p. 26,  $\phi$  étant continue, on a  $\partial(\phi + \psi) = \partial\phi + \partial\psi$  qui est donc maximal monotone.

L'algorithme II est donc utilisable avec  $A = \partial\phi$  et  $B = \partial\psi$  car évaluer  $e^n = J_B^\lambda u^n$  consiste à projeter sur  $R$ , c'est-à-dire à résoudre un problème linéaire

(1) "Monotonicity methods in Hilbert Spaces", *Contributions to non linear functional Analysis*, Zanantonello Ed., Academic Press, 1971, 101-156.

d'opérateur  $\Lambda^* \Lambda$ , et évaluer  $u^{n+1} = J_A^\lambda e^n$  consiste à résoudre un problème non linéaire posé presque partout dans  $(L^2(\Omega))^n$ , ce qui se ramène à résoudre presque partout un problème posé dans  $\mathbb{R}^n$ . On verra comment cette propriété est exploitée pour le problème discrétisé (§ 5).

On verra également d'autres algorithmes donnant de meilleurs résultats que l'algorithme I.

### 3.6. Algorithme III - Douglas-Rachford

Il s'agit à nouveau d'un algorithme du type  $v^{n+1} = F(\lambda)v^n$ , mais avec

$$(3-13) \quad F(\lambda) = J_A^\lambda(2J_B^\lambda - I) + (I - J_B^\lambda)$$

comme pour l'algorithme II, les points fixes de  $F(\lambda)$  dépendent de  $\lambda$ , mais ils en dépend d'une manière plus simple, puisque

$$v = F(\lambda)v \Leftrightarrow Au + Bu \ni 0 \quad \text{où } u = J_B^\lambda v.$$

En effet, si  $u = J_B^\lambda v$ , il existe  $b \in Bu$  tel que  $v = u + \lambda b$ , donc  $v$  point fixe de  $F(\lambda)$  équivaut à

$$v = J_A^\lambda(2u - v) + v - u \Leftrightarrow u = J_A^\lambda(u - \lambda b)$$

$$\Leftrightarrow u + \lambda Au \ni u - \lambda b \Leftrightarrow Au + Bu \ni 0. \quad \#$$

Lemme 3.1 : L'opérateur  $F(\lambda)$  défini en (3-12) est une contraction ferme, pour tout  $\lambda > 0$ .

En effet

$$\begin{aligned} |F(\lambda)x - F(\lambda)y|^2 &= |J_A^\lambda(2J_B^\lambda - I)x - J_A^\lambda(2J_B^\lambda - I)y|^2 + |(I - J_B^\lambda)x - (I - J_B^\lambda)y|^2 \\ &\quad + 2(J_A^\lambda(2J_B^\lambda - I)x - J_A^\lambda(2J_B^\lambda - I)y, (I - J_B^\lambda)x - (I - J_B^\lambda)y) \end{aligned}$$

Comme  $J_A^\lambda$  est une contraction ferme, en combinant les 1ers et 3èmes termes :

$$|F(\lambda)x - F(\lambda)y|^2 \leq (J_A^\lambda(2J_B^\lambda - I)x - J_A^\lambda(2J_B^\lambda - I)y, x-y) + |(I - J_B^\lambda)x - (I - J_B^\lambda)y|^2$$

$$(3-13) \quad |F(\lambda)x - F(\lambda)y|^2 \leq (F(\lambda)x - F(\lambda)y, x-y) - ((I - J_B^\lambda)x - (I - J_B^\lambda)y, J_B^\lambda x - J_B^\lambda y)$$

ce qui démontre le résultat puisque le 2ème terme est positif,  $J_B^\lambda$  étant une contraction ferme. #

Théorème 3.4 : La suite  $v^n$  générée par l'algorithme III converge faiblement vers un point fixe de  $F(\lambda)$ . La suite  $u^n \equiv J_B^\lambda v^n$  est bornée et vérifie

$$u^{n+1} - u^n \rightarrow 0.$$

Enfin  $b^n = \frac{v^n - u^n}{\lambda} \in Bu^n$  est bornée,  $b^{n+1} - b^n \rightarrow 0$  et  $(b^n - b, u^n - u) \rightarrow 0$  quand  $n \rightarrow \infty$ .

En effet  $F(\lambda)$  étant une contraction ferme, la première partie résulte du théorème 2.1 qui établit aussi que  $v^{n+1} - v^n \rightarrow 0$ , et donc aussi  $u^{n+1} - u^n$  puisque  $J_B^\lambda$  est une contraction.

La convergence de  $(b^n - b, u^n - u) \rightarrow 0$  résulte de l'inégalité (3-13) appliquée à  $x = v^n$  et  $y = F(\lambda)y = v$ , d'où  $J_B^\lambda x = u^n$  et  $J_B^\lambda y = u$  :

$$(3-14) \quad |v^{n+1} - v|^2 \leq (v^{n+1} - v, v^n - v) - ((v^n - u^n) - (v - u), u^n - u) \\ = |v^{n+1} - v|^2 + (v^{n+1} - v, v^n - v^{n+1}) - \lambda(b^n - b, u^n - u)$$

d'où

$$(b^n - b, u^n - u) \leq \frac{1}{\lambda} (v^{n+1} - v, v^n - v^{n+1})$$

d'où le résultat puisque  $v^{n+1}$  est borné  $|v^n - v^{n+1}| \rightarrow 0$ , et que  $(b^n - b, u^n - u) \geq 0$  par monotonie de B. #

Corollaire 3.2 : Si  $J_B^\lambda$  est faiblement fermé (en particulier si B est linéaire) alors  $u^n \rightharpoonup u = J_B^\lambda v$ , qui est donc une solution du problème  $Au + Bu \ni 0$ .

Si  $J_B^\lambda$  est compact (donc en particulier faiblement fermé) ou si B est coercif, la convergence de  $u^n$  vers u est forte. #

Remarque 3.6 : Vitesse de convergence

1°)- Dans le cas où B est à la fois coercif et lipschitzien (cf Remarque 3.1) on a  $|v^{n+1}-v| \leq (1 - \frac{2\lambda\alpha}{(1+\lambda M)^2})^{1/2} |v^n-v|$ .

En effet (3-14) donne

$$|v^{n+1}-v|^2 + 2\lambda(b^n-b, u^n-u) \leq |v^n-v|^2$$

d'où

$$|v^{n+1}-v|^2 + 2\lambda\alpha |u^n-u|^2 \leq |v^n-v|^2$$

Comme  $v^n-v = u^n-u + \lambda(b^n-b)$  on a aussi

$$|v^n-v| \leq (1 + \lambda M) |u^n-u|$$

puisque  $b^n = Bu^n$ ,  $b = Bu$  et que B est lipschitzien d'où la majoration annoncée. #

2°)- Il y a une valeur optimale de  $\lambda$  qui est  $\lambda = \frac{1}{M}$  pour la vitesse de convergence est donc  $(1 - \frac{\alpha}{2M})^{1/2}$ , c'est-à-dire ( $J_B^\lambda$  étant une contradiction

$$|u^n-u| \leq |v^n-v| \leq (1 - \frac{\alpha}{2M})^{n/2} |v^0-v|$$

ce qui démontre la supériorité de l'algorithme III sur l'algorithme I, dans cette situation.

Dans le cas du problème de l'obstacle (§ 3.5), il nécessite la résolution d'un système linéaire à chaque itération contrairement à l'algorithme I. Mais pour le problème discrétisé on a  $\frac{\alpha}{M} \sim h^2$  où h est le pas de discrétisation, qui est donc d'autant plus petit que la discrétisation est fine. Ceci explique que l'algorithme III puisse être tout de même plus économique. Sa supériorité sur l'algorithme II est encore plus évidente. #

3°)- Dans le cas où B est univoque, on a  $(I+\lambda B)J_B^\lambda = I$ , d'où

$$\lambda B J_B^\lambda = I - J_B^\lambda \quad \text{et} \quad (I - \lambda B) J_B^\lambda = 2J_B^\lambda - I. \quad \text{Ceci permet d'écrire}$$

l'algorithme III sous la forme

$$u^{n+1} = J_B^\lambda [J_A^\lambda (I - \lambda B) + \lambda B] u^n$$

qui a été introduit dans le cas linéaire par DOUGLAS et RACHFORD. #

### 3.7. Algorithme IV - Peaceman-Rachford

C'est encore un algorithme du type  $v^{n+1} = G(\lambda)v^n$  avec

$$(3-15) \quad G(\lambda) = (2J_A^\lambda - I)(2J_B^\lambda - I).$$

Il est immédiat de vérifier que  $G(\lambda)$  est une contraction, comme composé de deux contractions.

En effet,  $2J_A^\lambda - I$  est une contraction lorsque  $A$  est maximal monotone<sup>(1)</sup>. Soit  $v_i = (2J_A^\lambda - I)u_i$ ,  $i=1,2$ , on a

$$|v_2 - v_1|^2 = 4|J_A^\lambda u_2 - J_A^\lambda u_1|^2 - 4(J_A^\lambda u_2 - J_A^\lambda u_1, u_2 - u_1) + |u_2 - u_1|^2$$

d'où le résultat puisque  $J_A^\lambda$  est une contraction ferme.

On vérifie comme pour l'algorithme III que si  $v$  est point fixe de  $G(\lambda)$ , alors  $u = J_B^\lambda v$  est solution de  $Au + Bu \ni 0$ .

On pose encore  $b = \frac{v-u}{\lambda}$  d'où

$$\begin{aligned} v = G(\lambda)v &\Leftrightarrow v = (2J_A^\lambda - I)(2u - v) \\ &\Leftrightarrow v = 2J_A^\lambda(u - \lambda b) - 2u + v \Leftrightarrow u = J_A^\lambda(u - \lambda b) \Leftrightarrow Au + Bu \ni 0 \quad \# \end{aligned}$$

On renvoie à P.L. LIONS et B. MERCIER (SIAM J. Numr. Anal., à paraître) pour les démonstrations de convergence qui nécessitent des hypothèses un peu plus fortes que pour l'algorithme III.

La vitesse de convergence est ici  $(1 - \frac{\alpha}{M})^{1/2}$ , pour le même choix optimal de  $\lambda$ .

(1) Cf. Théorème 2.2.

#### 4. DUALITE

Nous donnons une présentation de la dualité différentielle (mais équivalente) à celle développée par Rockafellar et Ekeland-Témam).

##### 4.1. Fonctions convexes conjuguées

Soit  $\phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$  une fonction convexe s.c.i, on définit comme suit sa conjuguée  $\phi^*$  :

$$\phi^*(e) = \sup_{v \in H} [(e, v) - \phi(v)]$$

On démontre (EKELAND-TEMAM, loc. cit., p. 16) que  $\phi^*(e)$  est convexe s.c.i, et que  $\phi^{**} = \phi$ ; autrement dit  $\phi$  et  $\phi^*$  sont mutuellement conjuguées.

Lemme 4.1 : Soit  $A = \partial\phi$ , alors  $A^{-1} = \partial\phi^*$ .

En effet si  $e \in \partial\phi(u)$  on a

$$\phi^*(e) = (e, u) - \phi(u) \geq (e, v) - \phi(v), \quad \forall v \in V,$$

et donc  $(e, u) - \phi^*(e) = \phi(u) \geq (u, f) - \phi^*(f), \quad \forall f \in H$ .

puisque l'on a  $\phi(u) = \sup_{f \in H} [(u, f) - \phi^*(f)]$

d'où  $u \in \partial\phi^*(e)$ . #

Exemples : 1) La conjuguée  $J_K$ , fonction indicatrice d'un convexe fermé  $K \subset H$ ,

$$J_K(e) \equiv \sup_{\tau \in K} (\tau, e)$$

est appelée fonction d'appui du convexe  $K$ . Elle est positivement homogène de degré 1 et non différentiable.

2) La conjuguée de  $\phi(v) = \frac{1}{2} |v|^2 + J_K(v)$  est la fonction

$$\phi^*(e) = \frac{1}{2} |e - P_K e|^2 \quad \text{où } P_K \text{ est la projection sur } K.$$

$$\text{En effet } \sup_{e \in H} [(e, v) - \phi^*(e)] = \sup_{e \in H} \sup_{\tau \in K} [(e, v) - \frac{1}{2} |e - \tau|^2]$$

$$= \sup_{\tau \in K} \sup_{e \in H} [(e, v) - \frac{1}{2} |e - \tau|^2] = \sup_{\tau \in K} [(\tau, v) + \frac{1}{2} |\tau|^2]$$

Remarque 4.1 : Si  $A$  est coercif ( $A - \alpha I$  maximal monotone),  $A^{-1}$  est univoque et même lipschitzien ; en effet  $R(A) = H$  et

$$(y_2 - y_1, u_2 - u_1) \geq \alpha |u_2 - u_1|^2 \quad \forall \{u_i, y_i\} \in A$$

$$\Rightarrow |u_2 - u_1| \leq \frac{1}{\alpha} |y_2 - y_1| \quad \forall y_i \in H \text{ et } u_i \in A^{-1}(y_i)$$

En revenant à la fonction  $\phi$ , on voit que si  $\phi$  est coercive au sens suivant :

$$\frac{\phi(v)}{|v|} \rightarrow +\infty \text{ si } |v| \rightarrow +\infty$$

et strictement convexe, alors  $\phi^*$  est différentiable.

En effet dans ce cas,  $\forall e \in H$ , le problème d'optimisation

$$\inf_{v \in H} [\phi(v) - (e, v)] \Leftrightarrow \partial\phi(u) \ni e$$

admet une solution unique et donc  $(\partial\phi)^{-1}$  est non vide et réduit à un seul élément donc  $\phi^*$  est différentiable.

Pour illustrer cette remarque, on se reportera au 2ème exemple ci-dessous où l'on voit que  $\phi$  est coercif ;  $\phi^*$  est bien différentiable (ce n'est pas immédiat) et son gradient vaut  $e - P_K e$ . #

## 4.2. Dualité

On considère encore le problème

$$(4-1) \quad A\sigma + B\sigma \ni 0$$

où  $A$  et  $B$  sont maximaux monotones. Nous affirmons que ce problème est équivalent au problème

$$(4-2) \quad A^{-1}(e) - B^{-1}(-e) \ni 0.$$

En effet si  $u$  vérifie (4-1), il existe  $e \in A\sigma \cap (-B\sigma)$ , et donc  $\sigma \in A^{-1}(e)$  et  $\sigma \in B^{-1}(-e)$  d'où (4-2) et réciproquement.

Les problèmes (4-1) et (4-2) sont dits problèmes duaux l'un de l'autre.

On peut interpréter ces problèmes en termes d'optimisation lorsque

$A = \partial\phi$  et  $B = \partial\psi$ ;  $u$  sera solution de

$$(4-3) \quad \inf_{\tau \in H} [\phi(\tau) + \psi(\tau)]$$

et  $e$  sera solution de

$$(4-4) \quad \inf_{z \in H} [\phi^*(z) + \psi^*(-z)]$$

Les conditions pour que ces deux problèmes admettent des solutions sont plus faibles que pour les problèmes (4-1) et (4-2).

Il suffit donc que (4-1) ou (4-2) ait une solution pour que (4-3) et (4-4) aient une solution.

Pour cette raison, nous appellerons les deux premières relations, formulations fortes et les deux dernières formulations faibles.

Pour démontrer que (4-1) a une solution, on essaie en général de démontrer que  $\partial\phi + \partial\psi$ , est maximal monotone c'est-à-dire  $\partial(\phi + \psi) = \partial\phi + \partial\psi$ . Pour cela il suffit de trouver un point de  $H$  où  $\phi$  et  $\psi$  soient finies, et l'une d'elles continue. On appelle cette dernière condition la condition de qualification.

Les relations

$$\begin{cases} \sigma \in A^{-1}(e) \\ \sigma \in B^{-1}(-e) \end{cases} \Leftrightarrow \begin{cases} e \in A\sigma \\ -e \in B\sigma \end{cases}$$

sont appelées relations d'extrémalité. Trouver  $\sigma$  satisfaisant (4-1) ou  $e$  satisfaisant (4-2) est équivalent à trouver  $\sigma, e$  satisfaisant aux relations d'extrémalité.

### Cas particuliers

1°)- Soit  $\Lambda : V \rightarrow H$  un opérateur linéaire continu d'image fermée, et  $\phi : H \rightarrow \mathbb{R}$  une fonction convexe s.c.i.. Le problème

$$(4-5) \quad \inf_{v \in V} \phi(\Lambda v)$$

est visiblement équivalent au problème

$$(4-6) \quad \inf_{e \in R} \phi(e).$$

où  $R = \text{Im} \Lambda$  désigne l'image de  $V$  par  $\Lambda$ . La forme forte de ce problème d'optimisation est

$$\partial\phi(e) + \partial I_R(e) \ni 0, \quad \text{qui est bien un cas particulier de (4-1).}$$

On va expliciter le dual de (4-6), en utilisant le fait que la conjuguée de  $I_R$  est  $I_N$  où  $N = R^\perp = \text{Ker } \Lambda^*$  noyau de l'opérateur  $\Lambda^*$  adjoint de  $\Lambda$ . En effet

$$I_R^*(\tau) = \sup_{e \in R} (\tau, e) = \begin{cases} 0 & \text{si } \tau \in R^\perp \\ +\infty & \text{sinon} \end{cases}$$

Finalement le problème dual s'écrit

$$\inf_{\tau \in N} \phi^*(\tau).$$

Les relations d'extrémalité lient la solution  $\sigma$  du problème (4-7) à la solution  $u$  du problème (4-5) étant

$$\begin{cases} \sigma \in \partial\phi(\Lambda u) \\ \sigma \in N \end{cases}$$

Une condition suffisante pour que ce système ait une solution est donc que  $\phi$  soit continue en un point de  $R$  (qui entraîne que  $\partial(\phi + I_R) = \partial\phi + \partial I_R$  ou que  $\phi^*$  soit continue en un point de  $N$  ( $\partial(\phi^* + I_N) = \partial\phi^* + \partial I_N$ ). Cette condition est traditionnellement appelée condition de qualification.

2°)- On peut ramener le problème

$$(4-8) \quad \inf_{v \in V} [f(v) + g(\Lambda v)]$$

où  $f : V \rightarrow \mathbb{R} \cup \{+\infty\}$  est convexe s.c.i.

$g : Y \rightarrow \mathbb{R} \cup \{+\infty\}$  est convexe s.c.i.

$\Lambda : V \rightarrow Y$  est linéaire continu

au problème précédent en posant  $H = V \times Y$  et

$$\phi(e) = f(v) + g(y) \quad \text{pour } e = \{v, y\}$$

et  $R = \{e \in V \times Y : e = \{v, y\} \text{ tels que } \Lambda v - y = 0\}$ .

On a alors  $R^\perp = \{e \in H : \exists y \in Y \text{ tel que } e = \{\Lambda^* y, -y\}\}$

et  $\phi^*(e) = f^*(v) + g^*(y) \text{ pour } e = \{v, y\}$

d'où le problème dual

$$(4-9) \quad \inf_{g \in Y} [f^*(\Lambda^* y) + g^*(-y)]$$

La relation entre les solutions de (4-8) et (4-9) se décompose en une relation sur  $V$  et une relation sur  $Y$  :

$$\begin{aligned} -\Lambda^* z &\in \partial f(u) \\ z &\in \partial g(\Lambda u). \quad \# \end{aligned}$$

### 3°)- Généralisation

Considérons le problème suivant

$$(4-10) \quad \inf_{v \in V} \sum_{i=1}^N \phi_i(v)$$

où  $\phi_i : V \rightarrow \mathbb{R} \cup \{+\infty\}$  sont  $N$  fonctions convexes s.c.i..

Ce problème est en fait un problème du type

$$\sum_{i=1}^N A_i u \ni 0$$

où  $A_i = \partial \phi_i$  sont  $N$  opérateurs maximaux monotones. Soit  $H = \prod_{i=1}^N V$  et  $\phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$  défini par

$$\phi(e) = \sum_{i=1}^N \phi_i(v_i)$$

pour tout  $N$ -uplet  $e \in H$ ,  $e = \{v_1, \dots, v_N\}$ .

Soit  $D$  la diagonale de  $H$ , caractérisée comme l'ensemble des  $N$ -uples  $e = \{v, v, \dots, v\}$  où  $v \in V$ . Le problème (4-10) est visiblement équivalent au problème

$$\inf_{e \in D} \phi(e)$$

et se ramène donc à un problème avec deux opérateurs

$$\partial\phi(e) + \partial I_D(e) \ni 0.$$

On peut donc utiliser l'un des algorithmes I à IV pour résoudre ce problème qui n'est pas si compliqué qu'il n'en a l'air. En effet, la projection sur  $D$  revient à faire une moyenne des  $N$  composantes, et l'opérateur  $\partial\phi$  est en fait diagonal, c'est-à-dire que la minimisation sans contrainte équivaut à  $N$  minimisations sans contrainte sur les  $\phi_i$ . Ce procédé de décomposition a été étudié par PIERRA<sup>(1)</sup> en vue d'étudier la projection sur l'intersection de  $N$  convexes  $C_i$  (auquel cas

$$\phi_i(v) = \frac{1}{2N} |x-v|^2 + I_{C_i}(v).$$

#### 4.3. Résolution du primal ou du dual ?

Les deux cas particuliers (4-5) et (4-8) considérés se ramènent donc à un problème faisant intervenir la somme de deux opérateurs maximaux monotones.

On peut donc essayer d'utiliser l'un des algorithmes I, II, III ou IV. L'algorithme I nécessite que l'un des deux opérateurs soit univoque et si cette condition n'est pas vérifiée pour le problème primal, elle peut très bien l'être pour le dual (cf Remarque 4.1). Il se peut que les algorithmes I ou II soient plus rapides quand on les applique au dual.

Par contre, appliquer l'algorithme III au problème (4-1) ou à son dual (4-2) donne lieu à deux algorithmes équivalents en un sens qui est précisé dans le théorème 4.1.

---

(1) *Thèse, Université Scientifique et Médicale de Grenoble, 1976.*

Soit  $u^0, a^0 \in H$  et  $\lambda \geq 0$  donné, l'algorithme III appliqué au problème (4-1) s'écrit donc

$$(411) \quad v^{n+1} = [J_B^\lambda(2J_A^\lambda - I) + I - J_A^\lambda]v^n$$

$$\text{où } v^0 = u^0 + \lambda a^0.$$

Soit maintenant  $r \geq 0$  donné, l'algorithme III appliqué au problème dual s'écrit

$$(4-12) \quad e^{n+1} = [J_C^r(2J_{A^{-1}}^r - I) + I - J_{A^{-1}}^r]e^n$$

où  $C$  est l'opérateur défini par

$$Ce \equiv -B^{-1}(-e).$$

Théorème 4.1 : Pour  $e^0 = a^0 + ru^0$  et  $r = \frac{1}{\lambda}$  la suite  $e^n$  générée par l'algorithme (4-12) vérifie la relation

$$e^n = r v^n \quad \forall n \geq 0.$$

Lemme 4.1 : Pour  $r = \frac{1}{\lambda} > 0$  on a l'identité

$$J_C^r(-e) = \frac{1}{\lambda} (J_B^\lambda - I) \lambda e \quad \forall e \in H.$$

Démonstration du lemme :

$$\text{Soit } b = \frac{1}{\lambda} (J_B^\lambda - I) \lambda e$$

$$\text{on a } J_B^\lambda(\lambda e) = \lambda b + \lambda e.$$

$$\lambda e \in (I + \lambda B)(\lambda b + \lambda e)$$

$$\Rightarrow -b \in B(\lambda b + \lambda e)$$

$$\lambda b + \lambda e \in B^{-1}(-b)$$

$$-\lambda e \in \lambda b - B^{-1}(-b) \equiv \lambda b + Cb$$

$$-e \in b + r Cb \Rightarrow b = J_C^r(-e) \quad \#$$

Démonstration du théorème 4.1 :

Soit  $u^n = J_A v^n$  et  $a^n = \frac{1}{\lambda} (v^n - u^n) \in Au^n$ .

L'algorithme (4-11) peut s'écrire

$$u^{n+1} + \lambda a^{n+1} = J_B^\lambda (u^n - \lambda a^n) + \lambda a^n$$

c'est-à-dire

$$a^{n+1} + \lambda a^{n+1} = (J_B^\lambda - I)(u^n - \lambda a^n) + u^n$$

soit en divisant par  $\lambda$

$$\begin{aligned} a^{n+1} + ru^{n+1} &= \frac{1}{\lambda} (J_B^\lambda - I) \lambda (ru^n - a^n) + ru^n \\ &= J_C^r (a^n - ru^n) + ru^n \end{aligned}$$

où l'on a utilisé l'identité établie au Lemme 4.1.

Supposons par récurrence que  $e^n = a^n + ru^n$  ( ce qui est vrai pour  $n=0$  ).

On a

$$u^n \in A^{-1}(a^n) \Rightarrow e^n \in a^n + rA^{-1}(a^n) \Rightarrow a^n = J_{A^{-1}}^r(e^n)$$

d'où

$$\begin{aligned} a^{n+1} + ru^{n+1} &= J_C^r(2a^n - e^n) + e^n - a^n \\ &= (J_C^r(2J_{A^{-1}}^r - I) + I - J_{A^{-1}}^r)e^n \end{aligned}$$

d'où

$$e^{n+1} = a^{n+1} + ru^{n+1}. \quad \#$$

Remarque 4.2 : On peut démontrer un résultat analogue pour l'algorithme IV.

#

#### 4.4. Exemples

##### 1)- Problèmes de Mosolov

On a vu (§ 3.5) que le problème de Mosolov pouvait s'écrire

$$\inf_{e \in R} \left[ \frac{1}{2} \|e\|^2 + j_0(e) - (\sigma_0, e) \right]$$

avec  $H = (L^2(\Omega))^2$ ,  $R$  image de l'opérateur gradient de  $H_0^1(\Omega)$  dans  $H$ , et  $j_0(e) = \int_{\Omega} |e| dx$ , où  $|\cdot|$  désigne la norme euclidienne de  $\mathbb{R}^2$  et  $\|\cdot\|$  la norme de  $H$ .

On remarque alors que  $j_0(e) = j_K(e)$  où

$$j_K(e) = \sup_{\tau \in K} (\tau, e)$$

est la fonction d'appui du convexe  $K = \{e \in H : |e(x)| \leq 1, \text{ p.p. } x \in \Omega\}$

La conjuguée de  $\frac{1}{2} \|e\|^2 + j_0(e)$  est donc  $\frac{1}{2} \|\tau - P_K \tau\|^2$ , comme on l'a vu plus haut. On en déduit que la conjuguée de

$$\phi(e) \equiv \frac{1}{2} \|e\|^2 + j_0(e) - (\sigma_0, e) \quad \text{est}$$

$$\phi^*(\tau) \equiv \sup_{e \in H} [(\sigma_0 + \tau, e) - \frac{1}{2} \|e\|^2 - j_0(e)]$$

$$= \frac{1}{2} \|\sigma_0 + \tau - P_K(\sigma_0 + \tau)\|^2$$

Le problème dual (4-7) s'écrit donc

$$\inf_{\tau \in N} \frac{1}{2} \|\sigma_0 + \tau - P_K(\sigma_0 + \tau)\|^2$$

ou, en effectuant le changement de variable,  $\sigma = \sigma_0 + \tau$  et en posant  $E = N + \sigma_0$

$$\inf_{\sigma \in E} \frac{1}{2} \|\sigma - P_K \sigma\|^2$$

L'interprétation géométrique de ce problème est très simple :  $K$  est un convexe de  $H$  contenant l'origine,  $E$  une variété affine orthogonale à  $R$ , et le problème consiste donc à trouver la plus courte distance entre  $K$  et  $E$  (Fig. 4.1).

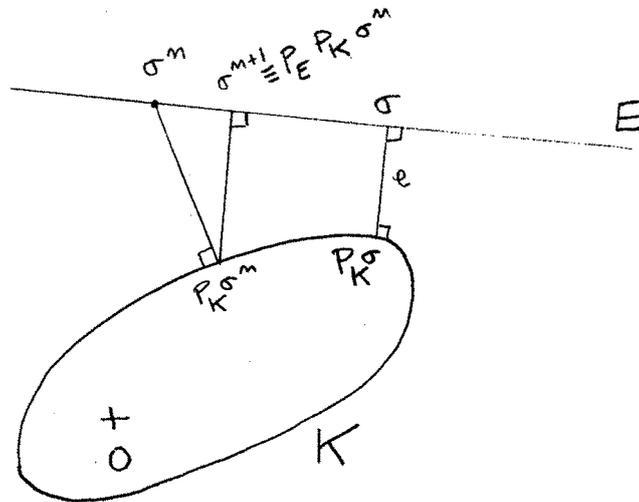


Figure 4.1 : Interprétation géométrique du problème dual (Problème de Mossolov)

La relation entre  $\tau$  et  $e$  est  $e \in \partial\phi^*(\tau)$ , c'est-à-dire  $e = \sigma_0 + \tau - P_K(\sigma_0 + \tau)$  d'où en revenant à la variable  $\sigma$  :  $e = \sigma - P_K \sigma$ .

La fonction  $\frac{1}{2} \|\sigma - P_K \sigma\|^2$  étant différentiable on peut donc appliquer l'algorithme I :

$$\sigma^{n+1} = P_E(\sigma^n - \lambda(\sigma^n - P_K \sigma^n))$$

qui a une interprétation géométrique très simple, dans le cas  $\lambda = 1$ , de projection alternée sur chacun des deux convexes. On notera que  $e^n = \sigma^n - P_K \sigma^n$  converge fortement vers la solution du problème primal (qui est unique puisque ce problème est coercif).

On reviendra plus loin sur l'interprétation mécanique de tout cet exemple : disons déjà que  $\sigma$  s'interprète comme un champ de contraintes et  $E$  comme l'ensemble des champs de contrainte en équilibre.

## 2)- Problème élasto-plastique modèle

C'est une variante du problème précédent qui consiste à chercher  $\sigma$  solution de

$$\inf_{\sigma \in K \cap E} \frac{1}{2} \|\sigma\|^2$$

où  $K$  et  $E$  sont définis comme plus haut (le champ de contraintes  $\sigma$  est cette fois-ci à la fois en équilibre et dans le convexe de "plasticité"  $K$ ).

C'est bien un problème du type

$$A\sigma + B\sigma \ni 0$$

avec  $A = I + \partial I_K$  et  $B = \partial I_E$ .

On évalue alors

$$I_E^*(e) = \sup_{\tau \in E} (\tau, e) = \begin{cases} (\sigma_0, e) & \text{si } e \in R \\ + \infty & \text{sinon} \end{cases} = (\sigma_0, e) + I_R(e)$$

où  $R$  et  $\sigma^0$  sont définis comme ci-dessus.

Le problème dual s'écrit donc, en remarquant que  $A^{-1} = P_K$

$$P_K e - \sigma_0 + \partial I_R(e) \ni 0$$

qui est la forme forte de

$$\inf_{e \in R} [\phi(e) - (\sigma_0, e)]$$

$$\text{où } \phi(e) = (e, P_K e) - \frac{1}{2} \|P_K e\|^2 = \frac{1}{2} \|e\|^2 - \frac{1}{2} \|e - P_K e\|^2$$

est la fonctionnelle convexe dont le gradient est précisément  $P_K$ .

On remarquera que cette fois-ci, aucun des deux problèmes ne vérifie la condition de qualification. Il y a d'ailleurs des contre-exemples montrant que le problème primal n'a pas de solution.

## 5. APPROXIMATIONS

Nous allons nous restreindre au cas particulier envisagé au paragraphe 4.2 où l'on cherche à résoudre

$$(5-1) \quad \inf_{v \in V} \phi(\Lambda v)$$

où  $\Lambda : V \rightarrow H$  est linéaire continue, injective et d'image fermée, de sorte que  $\|v\|_V \equiv |\Lambda v|$  est une norme sur  $V$ , et  $\phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$  est convexe s.c.i..

On va distinguer deux cas selon que c'est  $\phi$  ou  $\phi^*$  qui est coercive.

### 5.1. Cas où $\phi$ est coercive

On suppose que  $\phi = \phi_1 + \phi_2$ ,  $\phi_1$  différentiable et de gradient  $\phi_1'$  fortement monotone et lipschitzien

$$(5-2) \quad (\phi_1'(e) - \phi_1'(\tilde{e}), e - \tilde{e}) \geq \alpha |e - \tilde{e}|^2 \quad \forall e, \tilde{e} \in H$$

$$(5-3) \quad |\phi_1'(e) - \phi_1'(\tilde{e})| \leq M_1 |e - \tilde{e}|$$

et  $\phi_2$  convexe s.c.i. et lipschitzienne :

$$(5-4) \quad |\phi_2(e) - \phi_2(\tilde{e})| \leq M_2 |e - \tilde{e}| \quad \forall e, \tilde{e} \in H.$$

Avec ces notations u solution de (5-1) équivaut à  $e = \Lambda u$  solution de l'inéquation variationnelle

$$(5-5) \quad (\phi_1'(e), \tilde{e} - e) + \phi_2(\tilde{e}) - \phi_2(e) \geq 0 \quad \forall \tilde{e} \in R$$

et  $e \in R$ , où  $R \subset H$  est l'image de  $V$  par l'application  $\Lambda$ .

Dans ces conditions on se donne une famille  $(V_h)_{h>0}$  d'espaces de dimension finie  $V_h \subset V$ , un espace de Hilbert  $W$  dense dans  $V$ , avec injection continue et finalement une application  $r_h : W \rightarrow V_h$ , tels que

$$(5-6) \quad \|v - r_h v\|_V \leq \varepsilon(h) \|v\|_W, \quad \forall v \in W.$$

avec  $\varepsilon(h) \rightarrow 0$  si  $h \rightarrow 0$ .

L'espace  $W$  étant dense dans  $V$  cela montre que pour tout  $v \in V$  il existe  $v_h \in V_h$  tel que  $v_h \rightarrow v$  dans  $V$  lorsque  $h \rightarrow 0$ .

On choisit alors pour problème approché le problème

$$(5-7) \quad \inf_{v_h \in V_h} \phi(\Lambda v_h).$$

La coercivité de  $\phi$ , qui entraîne l'existence de  $u$ , entraîne aussi celle d'une solution  $u_h$  au problème approché (5-7). De même la stricte convexité de  $\phi$  entraîne l'unicité de  $u_h$ .

On pose alors  $e_h = \Lambda u_h$ .

Théorème 5.1 : Sous les hypothèses (5-2), (5-3), (5-4)  $u_h \rightarrow u$  lorsque  $h \rightarrow 0$ .  
Si de plus  $u \in W$ , on a l'estimation d'erreur

$$(5-8) \quad \|u - u_h\|_V^2 \leq C(\varepsilon(h) + (\varepsilon(h))^2),$$

où  $C$  est une constante dépendant de  $u$ .

Démonstration

Posons  $e_h = \Lambda u_h$  ;  $e_h$  vérifie l'inéquation

$$(5-9) \quad (\phi_1'(e_h), \tilde{e}_h - e_h) + \phi_2(\tilde{e}_h) - \phi_2(e_h) \geq 0 \quad \forall \tilde{e}_h \in R_h$$

où  $R_h = \Lambda(V_h)$  désigne l'image de l'espace  $V_h$  par l'application  $\Lambda$ .

Comme  $R_h \subset R$ , on peut choisir  $\tilde{e} = e_h$  dans (5-9) d'où en ajoutant avec (5-5), il vient

$$(\phi_1'(e), e_h - e) + (\phi_1'(e_h), \tilde{e}_h - e + e - e_h) + \phi_2(\tilde{e}_h) - \phi_2(e) \geq 0$$

c'est-à-dire

$$\begin{aligned} \alpha |e_h - e|^2 &\leq (\phi_1'(e) - \phi_1'(e_h), e - e_h) \leq (\phi_1'(e_h), \tilde{e}_h - e) + \phi_2(\tilde{e}_h) - \phi_2(e) \\ &\leq (\phi_1'(e), \tilde{e}_h - e) + (\phi_1'(e_h) - \phi_1'(e), \tilde{e}_h - e) + \phi_2(\tilde{e}_h) - \phi_2(e) \\ &\leq |\phi_1'(e)| |\tilde{e}_h - e| + \frac{\alpha}{2M_1^2} |\phi_1'(e_h) - \phi_1'(e)| + \frac{M_1^2}{2\alpha} |\tilde{e}_h - e|^2 + M_2 |\tilde{e}_h - e| \end{aligned}$$

$$\frac{\alpha}{2} |e_h - e|^2 \leq (M_2 + |\phi_1'(e)|) |\tilde{e}_h - e| + \frac{M_1^2}{2\alpha} |\tilde{e}_h - e|^2$$

d'où l'estimation (5-8) lorsque  $u \in W$ , en choisissant  $\tilde{e}_h = \Lambda(r_h u)$ .

Lorsque  $u \notin W$ , on raisonne par densité : on sait qu'il existe une suite  $v_h \in V_h$  vérifiant  $v_h \rightarrow u$  dans  $V$  fort d'où en choisissant  $\tilde{e}_h = \Lambda v_h$  la convergence de  $e_h$  vers  $e$  et donc de  $u_h$  vers  $u$ . #

### Généralisation

Lorsque  $\phi = \phi_1 + \phi_2 + I_K$  avec  $\phi_1, \phi_2$  vérifiant les hypothèses (5-2) à (5-4) et  $I_K$  est la fonction indicatrice d'un convexe fermé  $K \subset H$ , il faut restreindre  $e \in K \cap R$  dans l'inéquation (5-5) qui devra maintenant avoir lieu pour tout  $\tilde{e} \in K \cap R$ . Soit  $K_0 = \{v \in V : \Lambda v \in K\}$ , pour obtenir encore l'estimation d'erreur (5-8), il faudra exiger que  $r_h v \in K_0 \cap V_h$ , pour pouvoir encore choisir  $\tilde{e}_h = r_h v$  dans la démonstration précédente.

### Application au problème de Mosolov

Dans le cas du problème de Mosolov (cf. § 3.5), on a  $V = H_0^1(\Omega)$   
 $H = (L^2(\Omega))^2$ , et  $\Lambda$  est l'opérateur gradient. On choisira

$$\phi_1(e) = \frac{1}{2} |e|^2 - (\sigma_0, e)$$

et  $\phi_2(e) = j_0(e)$ , de sorte que les hypothèses (5-2) à (5-4) sont bien vérifiées. Dans ce cas, il n'y a pas besoin de mettre de fonction indicatrice dans  $\phi$  en plus de  $\phi_1$  et  $\phi_2$ . On verra plus loin un exemple où il faudra le faire, et les restrictions à imposer au sous-espace  $V_h$ .

Pour le problème de Mosolov, on choisira typiquement une triangulation  $T_h$  de  $\Omega$  et <sup>(1)</sup>.

$$V_h = \{v_h \in C^0(\bar{\Omega}) : v_h|_K \in \mathbb{P}_1, K \in T_h \text{ et } v_h = 0 \text{ sur } \partial\Omega\}$$

<sup>(1)</sup>  $\mathbb{P}_k$  désigne l'espace des polynômes de degré  $\leq k$  et à 2 variables.

La propriété (5-6) est alors vérifiée classiquement <sup>(1)</sup> avec  $W = H^2(\Omega)$  et  $r_h v =$  interpolé de  $v$  aux sommets de la triangulation, et enfin  $\varepsilon(h) : Ch$  où la constante  $C$  dépend du domaine  $\Omega$  ;  $V_h$  est un espace d'éléments finis de degré 1.

En utilisant des éléments finis de degré  $k > 1$ , on obtient  $\varepsilon(h) = Ch^k$  quitte à choisir  $W = H^{k+1}(\Omega)$ . Mais l'amélioration correspondante de la précision est souvent illusoire car en général  $u \notin H^{k+1}(\Omega)$  pour  $k > 1$ , et d'autre part il sera nécessaire d'utiliser une formule d'intégration numérique pour évaluer  $j_0(\nabla v) = \int_{\Omega} |\nabla v| dX$  (cf. GLOWINXKI-LIONS-TREMOLIERES, Analyse Numérique des Inéquations variationnelles, Dunod, 1976).

### 5.2 Cas où $\phi^*$ est coercive

On considère le problème dual de (5-1)

$$(5-10) \quad \inf_{\tau \in N} \phi^*(\tau)$$

(cf. 4-7)

On suppose que  $\phi^*$  est de la forme

$$(5-11) \quad \phi^* = J + I_K$$

où  $J$  a un gradient fortement monotone et lipschitzien (cf. (5-2) et (5-3), et  $I_K$  est la fonction indicatrice d'un convexe fermé. On suppose que  $K \cap N$  est non vide.

La solution  $\sigma \in K \cap N$  du problème (5-10) est donc unique et vérifie l'inéquation variationnelle

$$(5-12) \quad (J'(\sigma), \tau - \sigma) \geq 0 \quad \forall \tau \in K \cap N.$$

On introduit alors l'espace  $H_h$  destiné à approcher  $H$  et on demande que  $R_h \subset H_h$ .

(1) cf. P.G. CIARLET, The finite element method for elliptic problems, North Holland, 1978.

On pose alors

$$N_h = \{\tau_h \in H_h : (\tau_h, \Lambda v_h) = 0 \quad \forall v_h \in V_h\}$$

qui est l'orthogonal de  $R_h$  dans  $H_h$ .

Le problème approché est alors tout naturellement

$$(5-13) \quad \inf_{\tau_h \in N_h} \phi^*(\tau_h).$$

La coercivité de  $\phi^*$  qui entraîne l'existence de  $\sigma$  solution de (5-10) entraînera également celle de  $\sigma_h$  solution de (5-13).

On fait alors l'hypothèse qu'il existe  $\pi_h : H \rightarrow H_h$

$$(5-14a) \quad \pi_h \tau \rightarrow \tau \quad \text{lorsque } h \rightarrow 0$$

$$(5-14b) \quad \tau \in K \cap N \Rightarrow \pi_h \tau \in K \cap N_h \quad (1)$$

$$(5-14c) \quad |\pi_h \tau| \leq C |\tau| \quad \text{où } C \text{ est indépendant de } h.$$

La difficulté par rapport au § 5.1 est que  $N_h$  n'est pas inclus dans  $N$ . On va d'abord montrer la convergence de  $\sigma_h$  vers  $\sigma$ . Pour cela il faut d'abord montrer (cf. GLOWINSKI - LIONS - TREMOLIERES, loc. cit) que

$$(5-15) \quad \text{si } \tau_h \in N_h \text{ converge faiblement vers } \tau \text{ lorsque } h \rightarrow 0$$

alors  $\tau \in N$  ;

en effet, il suffit de montrer que  $(\tau, \Lambda v) = 0 \quad \forall v \in V$ , or d'après (5-6) il existe  $v_h \in V_h \rightarrow v$ , donc

$$(\tau_h, \Lambda v_h) \rightarrow (\tau, \Lambda v) \quad \text{d'où le résultat puisque } \tau_h \in N_h$$

(1) Si  $\pi_h$  est la projection de  $H$  sur  $H_h$  alors

$$\tau \in N \Rightarrow \pi_h \tau \in N_h.$$

En effet l'hypothèse  $R_h \subset H_h$  montre que

$$(\pi_h \tau, \Lambda v_h) = (\tau, \Lambda v_h) = 0 \quad \forall v_h \in V_h$$

Il reste à espérer que l'on a aussi  $\tau \in K \Rightarrow \pi_h \tau \in K$ .

La propriété (5-15) montre que  $N_h$  ne converge pas vers quelque chose de plus grand que  $N$ , et l'existence de  $\pi_h$ , montre que  $N_h$  ne converge pas vers quelque chose de plus petit que  $N$ .

Théorème 5.2 : Sous les hypothèses (5-6) et (5-14) la solution  $\sigma_h$  du problème approché converge fortement vers  $\sigma$ .

Démonstration :

On utilise la démarche classique : estimation a priori, convergence faible, convergence forte.

On a tout d'abord

$$J(\sigma_h) \leq J(\pi_h \sigma) \leq C$$

puisque  $\pi_h \sigma \in K \cap N_h$ , et en appliquant (5-14c).

Comme  $J$  est coercif, cela montre que  $\sigma_h$  est borné.

Soit  $\sigma_h \rightarrow \phi$  une sous suite faiblement convergente on a  $\phi \in K \cap N$  d'après (5-15), et d'autre part

$$(5-16) \quad J(\phi) \leq \liminf_{h \rightarrow 0} J(\sigma_h) \leq \lim_{h \rightarrow 0} J(\pi_h \tau) = J(\tau) \quad \forall \tau \in K \cap N$$

puisque  $J$  est faiblement s.c.i et fortement continue, ce qui montre que  $\phi = \sigma$ .

Pour démontrer la convergence de  $J(\sigma_h) \rightarrow J(\sigma)$  (qui résulte de (5-16) :

choisir  $\tau = \phi = \sigma$ ) et l'inégalité

$$J(\sigma) - J(\sigma_h) \geq (J'(\sigma_h), \sigma - \sigma_h) + \frac{\alpha}{2} \|\sigma_h - \sigma\|^2$$

qui résulte de la forte monotonie de  $J'$  et qui montre que  $\|\sigma_h - \sigma\| \rightarrow 0$ , en effet :

$$(J'(\sigma_h), \sigma - \sigma_h) = (J'(\sigma_h), \sigma - \pi_h \sigma) + (J'(\sigma_h), \pi_h \sigma - \sigma_h) \geq (J'(\sigma_h), \sigma - \pi_h \sigma)$$

quantité qui tend vers zéro lorsque  $h \rightarrow 0$  car  $J'(\sigma_h)$  est borné (puisque  $\sigma_h$  est borné). #

Estimation d'erreur

Pour obtenir une estimation d'erreur sur  $\|\sigma - \sigma_h\|$  on supposera qu'il existe un sous-espace dense  $\tilde{H} \subset H$  tel que

$$(5-17) \quad \|\tau - \pi_h \tau\| \leq \varepsilon(h) \|\tau\|_{\tilde{H}} \quad \forall \tau \in \tilde{H}$$

Mais celà ne suffit pas, il faudra également supposer que le problème

$$(5-18) \quad \partial \phi^*(\sigma) + \partial I_N(\sigma) \ni 0$$

qui peut être considéré comme une formulation forte du problème (5-10), admette une solution (Noter que  $\partial \phi^* + \partial I_N$  n'est pas forcément maximal monotone à cause de la présence de  $I_K$  dans  $\phi^*$ ).

Dans ce cas, en effet, il existera  $u \in V$  tel que

$$(5-19) \quad (J'(\sigma), \tau - \sigma) + (\Lambda u, \tau - \sigma) \geq 0, \quad \forall \tau \in K$$

puisque  $\partial I_N(\sigma) = N^{\perp} = R$ .

Ajoutant alors la condition d'optimalité pour  $\sigma_h$  :

$$(J'(\sigma_h), \tau_h - \sigma_h) \geq 0 \quad \forall \tau_h \in K \cap N_h$$

on obtiendra en choisissant  $\tau = \sigma_h$  dans (5-18) :

$$(J'(\sigma) - J'(\sigma_h), \sigma_h - \sigma) + (J'(\sigma_h), \tau_h - \sigma) + (\Lambda u, \sigma_h - \sigma) \geq 0$$

d'où, en utilisant l'égalité  $(\sigma_h - \sigma, \Lambda v_h) = 0$  vraie  $\forall v_h \in V_h$ ,

$$\begin{aligned} \alpha \|\sigma - \sigma_h\|^2 &\leq (J'(\sigma), \tau_h - \sigma) + (J'(\sigma_h) - J'(\sigma), \tau_h - \sigma) + (\Lambda(u - v_h), \sigma_h - \sigma) \\ &\leq C(\|\tau_h - \sigma\| + \|\tau_h - \sigma\|^2 + \|u - v_h\|^2) + \frac{\alpha}{2} \|\sigma_h - \sigma\|^2 \end{aligned}$$

d'où

$$\frac{\alpha}{2} \|\sigma - \sigma_h\|^2 \leq C(\varepsilon(h) + (\varepsilon(h))^2)$$

en supposant le résultat de régularité

$$\sigma \in \tilde{H}, \quad u \in W.$$

Exemple : Problème d'élasto-plasticité (cf. § 4.3)

On est exactement dans les conditions d'application de la présente théorie à un changement de variable près,  $J$  étant une forme quadratique définie positive. #

### 5.3 Dépendance de la vitesse de convergence des algorithmes en fonction de $h$ .

Dans les estimations de la vitesse de convergence des algorithmes I, III et IV, on a vu que seules les propriétés de coercivité et de lipschitz continuité d'un des deux opérateurs, (en l'occurrence l'opérateur B) intervenait dans la vitesse de convergence.

On voit que dans le cas du problème (5-1), écrit sous la forme (forte)

$$\partial\phi(e) + \partial I_R(e) \ni 0 ;$$

et discrétisé sous la forme

$$\partial\phi(e_h) + \partial I_{R_h}(e_h) \ni 0,$$

seul un des deux opérateurs est discrétisé. Si donc on choisit toujours  $B = \partial\phi$ , la vitesse de convergence des algorithmes I, III et IV sera indépendante de  $h$  <sup>(1)</sup>.

De même le problème dual (5-18) se discrétise sous la forme

$$\partial\phi^*(\sigma_h) + \partial I_{N_h}(\sigma_h) \ni 0$$

et en choisissant encore  $B = \partial\phi^*$ , on aura une vitesse de convergence indépendante de  $h$ .

En revanche le discrétisé naturel du problème de l'obstacle (cf. § 3.5) s'écrit

---

(1) Plus exactement les estimations donnent l'existence d'une vitesse minimale indépendante de  $h$ .

$$A_h u_h + \partial I_K(u_h) \ni f_h$$

où  $A_h : V_h \rightarrow V_h$  est une matrice vérifiant

$$(A_h u_h, v_h) = \int_{\Omega} \nabla u_h \nabla v_h$$

dont le conditionnement dépend fortement de  $h$ , et cela explique que les algorithmes I, III et IV donnent une vitesse de convergence qui dépend de  $h$  (cf. Remarque 3.6).

On a vu au § 4.2 que le problème (4-8) :

$$\begin{array}{ll} \text{Inf}_{v \in V} [f(v) + g(\Lambda v)] & \text{(où } f : V \rightarrow \mathbb{R} \text{ convexe} \\ & g : Y \rightarrow \mathbb{R} \text{ convexe} \\ & \Lambda : V \rightarrow Y \text{ linéaire} \end{array}$$

était un cas particulier du problème

$$\text{Inf}_{e \in R} \phi(e)$$

quitte à travailler sur l'espace produit  $H = V \times Y$ .

Dans le cas où f et g vérifient les hypothèses (5-2) à (5-4), on aura donc les mêmes conclusions, en ce qui concerne l'indépendance de  $h$  des vitesses de convergence des algorithmes I, III ou IV.

On remarquera que le problème de l'obstacle est un cas particulier du problème précédent avec  $f(v) = I_K(v)$  et donc ne vérifiant pas les conditions (5-2) à (5-4). On ne peut donc rien conclure en ce qui concerne la vitesse de convergence des algorithmes obtenus en utilisant cette nouvelle manière de poser le problème, mais il serait intéressant de faire des essais numériques en ce sens.

B - INEQUATIONS VARIATIONNELLES EN MECANIQUE DES MILIEUX CONTINUS

Plan

1. Equations du mouvement
2. Elasticité infinitésimale
3. Inégalité de Korn
4. Modèle de Hencky
5. Formulation variationnelle
6. Modèle de Prandtl-Reuss
7. Modèle élasto-visco-plastique
8. Modèle rigide-visco-plastique
9. Problèmes de frottement

## B - INEQUATIONS VARIATIONNELLES EN MECANIQUE DES MILIEUX CONTINUS

### 1. Equations du mouvement

Soit  $\Omega(t) \subset \mathbb{R}^3$  un milieu continu, et  $\Omega(0)$  sa position à l'instant initial. Le mouvement est décrit par la donnée d'une fonction  $\vec{F}$  donnant la position à l'instant  $t$  de la particule située initialement en un point  $\vec{a} \in \Omega(0)$

$$\vec{x} = \vec{F}(\vec{a}, t) \in \Omega(t),$$

où  $x_i = F_i(a_1, a_2, a_3, t)$ . Sa vitesse est donnée par  $\vec{v} = \frac{\partial \vec{F}}{\partial t}$ ,

La description des quantités physiques (densités, forces, etc.) peut se faire en coordonnées eulériennes en se donnant  $\rho$  fonction de  $\vec{x}$  et de  $t$  :  $\rho = \rho(x, t)$  ou lagrangiennes :  $\rho = \rho(a, t)$ .

Nous commençons par établir l'équation du mouvement en coordonnées eulériennes. Soit  $\omega(t)$  un élément de matière de forme (régulière) quelconque de  $\Omega(t)$ . On établit en mécanique <sup>(1)</sup> l'existence d'un tenseur de contraintes (Eulérien)  $\sigma(x, t)$  symétrique tel que, si  $\vec{n}(x, t)$  désigne le vecteur unitaire  $\vec{n}(x, t)$  en  $x$  normal à la surface  $\partial\omega(t)$  de  $\omega(t)$  (dirigé extérieurement), alors l'action du milieu continu sur  $\omega(t)$  se traduit par une force surfacique  $\vec{T}(x, t)$  de composantes

$$T_i = \sum_j \sigma_{ij} n_j$$

que l'on notera aussi

$$T_i = \sigma_{ij} n_j$$

en utilisant la convention de sommation ou  $\vec{T} = \sigma \cdot \vec{n}$  en notation vectorielle.

En écrivant alors que la variation de la quantité de mouvement est égale au travail des forces extérieures, on obtient que

---

(1) J. MANDEL, *Cours de Mécanique des Milieux Continus*, Dunod, Paris 1976. Tome I, p. 55-66.

$$(1-1) \quad \frac{d}{dt} \int_{\omega(t)} \rho \vec{v} \, dX = \int_{\partial\omega(t)} \sigma \cdot \vec{n} \, ds + \int_{\omega(t)} \vec{f} \, dX$$

où  $\vec{f}(x,t)$  désigne les forces massiques extérieures (gravité,...).

En remarquant que pour toute fonction  $g(x,t)$

$$\frac{d}{dt} \int_{\omega(t)} g \, dX = \int_{\omega(t)} \left( \frac{\partial g}{\partial t} + \operatorname{div} (g\vec{v}) \right) dX$$

et (formule de Green) que

$$\int_{\partial\omega(t)} \sigma \cdot \vec{n} \, ds = \int_{\omega(t)} \operatorname{div} \sigma \, dX$$

où  $\operatorname{div} \sigma$  est le vecteur de composantes  $\frac{\partial \sigma_{ij}}{\partial x_j}$ , on obtient, en notation tensorielle

$$(1-2) \quad \frac{\partial}{\partial t} (\rho v_i) + \frac{\partial}{\partial x_j} (\rho v_i v_j) = \frac{\partial \sigma_{ij}}{\partial x_j} + f_i \quad .$$

En utilisant la conservation de la masse :

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_j} (\rho v_j) = 0$$

on peut écrire (1-2) sous la forme

$$(1-3) \quad \rho \left( \frac{\partial v_i}{\partial t} + v_j \frac{\partial v_i}{\partial x_j} \right) = \frac{\partial \sigma_{ij}}{\partial x_j} + f_i \quad .$$

Soit  $g(x,t)$  une fonction définie en variables d'euler et  $G(a,t)$  la même fonction en variables de Lagrange, on a

$$G(a,t) = g(F(a,t),t)$$

$$\text{d'où} \quad \frac{\partial G}{\partial t} = \frac{\partial g}{\partial t} + \frac{\partial g}{\partial x_j} \frac{\partial x_j}{\partial t} = \frac{\partial g}{\partial t} + v_j \frac{\partial g}{\partial x_j} \equiv \frac{dg}{dt}$$

et la quantité  $\frac{dg}{dt}$ , appelée dérivée totale coïncide avec la dérivée lagrangienne. Avec cette notation, on peut écrire (1-3) sous la forme vectorielle

$$(1-4) \quad \rho \frac{d\vec{v}}{dt} = \text{div } \vec{\sigma} + \vec{f}$$

que l'on appelle forme eulérienne modifiée de l'équation du mouvement.

Pour obtenir la forme lagrangienne, il faudrait remplacer les dérivations en  $x$  de  $\text{div } \vec{\sigma}$  par des dérivations en  $a$ , à l'aide du Jacobien de la transformation  $a \rightarrow x$  :  $B \equiv \left( \left( \frac{\partial F_j}{\partial a_i} \right)_{i,j} \right)$ .

On a alors

$$\frac{\partial G}{\partial a_i} = \frac{\partial F_j}{\partial a_i} \frac{\partial g}{\partial x_j} \quad \text{ou} \quad \nabla G = B \cdot \nabla g$$

$$\text{c'est-à-dire en notant } B^{-1} = ((\beta_{ij})) \quad \frac{\partial g}{\partial x_j} = \beta_{jk} \frac{\partial G}{\partial a_k}$$

où les  $\beta_{ij} = \beta_{ij}(\vec{a})$  dépendent de  $\vec{a}$ .

Quand on passe aux coordonnées de Lagrange, on aura donc

$$(1-5) \quad \rho \frac{\partial v_i}{\partial t} = \beta_{jk} \frac{\partial \sigma_{ij}}{\partial a_k} + f_i$$

où les  $v_i, \sigma_{ij}, f_i$  sont maintenant fonctions de  $\vec{a}$  et de  $t$ .

Lorsque l'accélération  $\frac{d\vec{v}}{dt}$  est négligeable (pour des phénomènes lents dits quasi-statiques) ou nulle (pour des problèmes stationnaires, l'équation du mouvement (1-4) se réduit à l'équation d'équilibre

$$(1-6) \quad \text{div } \vec{\sigma} + \vec{f} = 0.$$

## 2. Elasticité infinitésimale

On suppose que  $\vec{x} = \vec{a} + \lambda \vec{w}(\vec{a})$  où  $\lambda$  est un petit paramètre.  
On a donc

$$\frac{\partial F_j}{\partial a_i} = \delta_{ij} + \lambda \frac{\partial w_j}{\partial a_i}$$

d'où  $\beta_{ij} = \delta_{ij} + 0(\lambda)$ .

L'équation du mouvement (1-5) en coordonnées de Lagrange se réduit donc au premier ordre près à

$$\rho \frac{\partial v_i}{\partial t} = \frac{\partial \sigma_{ij}}{\partial a_j} + f_i$$

Le vecteur  $\vec{u} = \lambda \vec{w}$  est le déplacement, et on s'aperçoit que pour cette approximation soit valable, il suffit que les  $\frac{\partial u_i}{\partial a_j}$  soient petits ("petites déformations").

On introduit ensuite le tenseur de déformation.

Soit  $a+da$  et  $a+\delta a$  deux points proches de  $a$ , et  $x+dx$  et  $x+\delta x$  leurs images par la transformation  $F$ .

On a

$$dx_i = \frac{\partial F_i}{\partial a_j} da_j, \quad \delta x_i = \frac{\partial F_i}{\partial a_k} \delta a_k$$

d'où

$$d\vec{x} \cdot \delta \vec{x} = \frac{\partial F_i}{\partial a_j} \frac{\partial F_i}{\partial a_k} da_j \delta a_k$$

$$d\vec{x} \cdot \delta \vec{x} - d\vec{a} \cdot \delta \vec{a} = \left( \frac{\partial F_i}{\partial a_j} \frac{\partial F_i}{\partial a_k} - \delta_{ij} \right) da_j \delta a_k \equiv 2 \Delta_{jk} da_j \delta a_k$$

où par définition  $\Delta \equiv ((\Delta_{ij}))$  est le tenseur des déformations.

En remarquant que

$$\frac{\partial F_j}{\partial a_i} = \delta_{ij} + \frac{\partial u_i}{\partial a_j} \quad \text{on voit que}$$

$$2 \Delta_{jk} = \left( \delta_{ji} + \frac{\partial u_i}{\partial a_j} \right) \left( \delta_{ki} + \frac{\partial u_i}{\partial a_k} \right) - \delta_{ij}$$

$$2 \Delta_{jk} = \frac{\partial u_k}{\partial a_j} + \frac{\partial u_j}{\partial a_k} + \frac{\partial u_i}{\partial a_j} \frac{\partial u_i}{\partial a_k}$$

on voit que lorsque  $\vec{u} = \lambda \vec{w}$ ,  $\Delta = \varepsilon(u) + O(\lambda)$  où  $\varepsilon(u)$  est le tenseur de déformations linéarisé, défini par

$$(2-1) \quad \varepsilon_{ij}(u) = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right).$$

qui est la partie symétrique du tenseur des déformations.

En élasticité linéaire, il y a une relation linéaire entre le tenseur des contraintes  $\sigma$  et le tenseur des déformations linéarisées  $\varepsilon(u)$  : dans le cas isotope on a

$$(2-2) \quad \sigma = \lambda(\operatorname{div} u)\delta + 2\mu\varepsilon(u) \equiv \Lambda\varepsilon(u)$$

où  $\lambda$  et  $\mu$  sont les coefficients de Lamé et  $\delta$  le tenseur de Kronecker.

L'inverse de cette relation s'écrit  $\varepsilon(u) = \Lambda^{-1}\sigma$  avec

$$(2-3) \quad \Lambda^{-1}\sigma \equiv \frac{1+\nu}{E}\sigma - \frac{\nu}{E}\operatorname{tr}(\sigma)\delta$$

où  $\operatorname{tr}(\sigma)$  désigne la trace du tenseur  $\sigma$ , avec les relations

$$\lambda = \frac{E}{(1+\nu)(1-2\nu)}, \quad \mu = \frac{E}{2(1+\nu)}$$

$E$  s'appelle module de Young, et  $\nu$  le coefficient de Poisson qui, physiquement, vérifie  $0 \leq \nu \leq \frac{1}{2}$  le cas  $\nu = \frac{1}{2}$  correspondant à l'incompressibilité.

Soit  $\Omega = \Omega(o)$ , en résumé, en élasticité linéaire, on recherche  $\sigma$  et  $u$  vérifiant sur  $\Omega$

$$(2-4) \quad \sigma - \Lambda\varepsilon(\vec{u}) = 0$$

$$(2-5) \quad \operatorname{div}\sigma + \vec{f} = 0$$

à quoi il faut rajouter les conditions aux limites :

$$(2-6) \quad \vec{u} = 0 \quad \text{sur } \Gamma_0$$

$$(2-7) \quad \sigma \cdot \vec{n} = \vec{F} \quad \text{sur } \Gamma_1$$

où  $\Gamma_0$  est la partie de la frontière de  $\Omega$  où le milieu continu est fixé et  $\Gamma_1$  celle où il est soumis à des forces pressantes. Nous supposons que  $\vec{f} \in (L^2(\Omega))^3$ ,  $\vec{F} \in (L^2(\Gamma_1))^3$  que  $\mu > 0$  et  $3\lambda + 2\mu > 0$

### 3. Inégalité de Korn

Soit  $V = \{\vec{v} \in (H^1(\Omega))^3 : \vec{v} = 0 \text{ sur } \Gamma_0\}$  l'espace des déplacements admissibles.

Lemme 1 (Théorème des travaux virtuels)

=====

Si  $\sigma \in (L^2(\Omega))^9$  et vérifie (2-5), (2-7), alors on a

$$(3-1) \quad \int_{\Omega} \sigma \cdot \varepsilon(\vec{v}) \, dX = \int_{\Omega} \vec{f} \cdot \vec{v} \, dX + \int_{\Gamma} \vec{F} \cdot \vec{v} \, ds \quad \forall \vec{v} \in V. \quad \#$$

La démonstration repose sur la formule de Green vraie pour tout  $v \in V$ .

$$(3-2) \quad \int_{\Omega} \sigma \cdot \varepsilon(\vec{v}) \, dX = \int_{\Gamma} (\sigma \cdot \vec{n}) \cdot \vec{v} \, ds - \int_{\Omega} (\operatorname{div} \sigma) \cdot \vec{v} \, dX \quad \#$$

Le déplacement  $\vec{u}$  est donc solution du problème variationnel linéaire

$$(3-3) \quad a(\vec{u}, \vec{v}) = L(\vec{v}) \quad \forall \vec{v} \in V$$

où  $a$  est la forme bilinéaire continue sur  $V \times V$  :

$$(3-4) \quad \begin{aligned} a(\vec{u}, \vec{v}) &\equiv \int_{\Omega} (\Lambda \varepsilon(u)) \cdot \varepsilon(v) \, dX \\ L(\vec{v}) &\equiv \int_{\Omega} \vec{f} \cdot \vec{v} \, dX + \int_{\Gamma_1} \vec{F} \vec{v} \, ds \end{aligned}$$

est une forme linéaire continue sur  $V$  d'après les hypothèses faites sur  $\vec{f}$  et  $\vec{F}$ .

L'existence et l'unicité résultent de la  $V$ -ellipticité de  $a$  :  
il existe  $\alpha > 0$  tel que

$$a(v, v) \geq \alpha \|v\|_1^2 \quad \forall v \in V$$

où  $\|\cdot\|_1$  désigne la norme de  $H^1(\Omega)$ .

Une étape essentielle de la démonstration est d'établir la célèbre inégalité de Korn.

Lemme 2 : Il existe une constante  $\alpha_0 > 0$  telle que

$$\int_{\Omega} (|\varepsilon(v)|^2 + |v|^2) \, dX \geq \alpha_0 \|v\|_1^2 \quad \forall v \in (H^1(\Omega))^3.$$

(La difficulté est que  $\varepsilon(v)$  ne contient pas toutes les dérivées  $\frac{\partial v_i}{\partial x_j}$  )

La démonstration s'appuie sur le lemme suivant <sup>(1)</sup>

Lemme 3 : Si  $\Omega$  est un ouvert borné de frontière assez régulière <sup>(2)</sup>, on a

$$\|p\|_{L^2(\Omega)}^2 \leq C (\|p\|_{H^{-1}(\Omega)}^2 + \|\nabla p\|_{(H^{-1}(\Omega))^3}^2). \quad \#$$

(1) Démontré dans DUVAUT-LIONS, Les Inéquations en mécanique et en physique Dunod 1972, pour le cas où la frontière de  $\Omega$  est  $C^\infty$

(2) Le résultat a été démontré par BOLLEY-CAMUS, C.R.Acad. Sc. Paris 1976, dans le cas où  $\Omega$  est polyédrique.

Démonstration du Lemme 2

On suppose donc que  $v_i$  et  $\varepsilon_{ij}(v) \in L^2(\Omega)$ , et on veut démontrer que toutes les dérivées  $\frac{\partial v_i}{\partial x_j}$  sont en fait dans  $L^2(\Omega)$ .

Pour cela il suffit de démontrer que  $\frac{\partial v_i}{\partial x_j} \in H^{-1}(\Omega)$  et que  $\frac{\partial^2 v_i}{\partial x_j \partial x_k} \in H^{-1}(\Omega)$  d'après le Lemme 3.

Or  $v_i \in L^2(\Omega) \Rightarrow \frac{\partial v_i}{\partial x_j} \in H^{-1}(\Omega)$  par continuité de la dérivation au sens des distributions, et  $\varepsilon_{ij}(v) \in L^2(\Omega) \Rightarrow \frac{\partial \varepsilon_{ij}(v)}{\partial x_k} = b_{ijk} \in H^{-1}(\Omega)$ , c'est-à-dire  $6 \times 3 = 18$  équations, pour 18 inconnues qui sont les  $\frac{\partial^2 v_i}{\partial x_j \partial x_k}$ . Ce système linéaire est inversible, en effet

$$\frac{\partial^2 v_i}{\partial x_j \partial x_k} = \frac{\partial}{\partial x_j} \varepsilon_{ik}(v) + \frac{\partial}{\partial x_k} \varepsilon_{ij}(v) - \frac{\partial}{\partial x_i} \varepsilon_{jk}(v) \in H^{-1}(\Omega)$$

d'où  $\frac{\partial v_i}{\partial x_j} \in L^2(\Omega)$ , d'où le résultat. #

Théorème 3.1 : Si  $\Gamma_0$  est de mesure non nulle, alors

$$\|\varepsilon(v)\|_{L^2(\Omega)} \geq C \|v\|_{L^2(\Omega)} \quad \forall v \in V.$$

Démonstration :

$\|\varepsilon(v)\|_{L^2(\Omega)}$  est une norme sur  $V$ , en effet  $\varepsilon(v) = 0$  entraîne que  $v$  est un mouvement rigide :  $\vec{v} = \vec{a} + \vec{b} \times \vec{r}$  où  $\vec{r} = \vec{0}_m$ . Comme ce mouvement rigide s'annule sur  $\Gamma_0$  dont la mesure est non nulle, il s'annule partout. On raisonne alors par l'absurde : il existerait donc une suite  $v_n$  avec  $\|v_n\|_{L^2(\Omega)} = 1$  et  $\|\varepsilon(v_n)\| \rightarrow 0$ . Cette suite étant bornée d'après le Lemme 2, elle admet une sous suite faiblement convergente  $v_{n'} \rightharpoonup v$  dans  $V$ , d'où

$$\liminf_{n' \rightarrow \infty} \|\varepsilon(v_{n'})\|^2 \geq \|\varepsilon(v)\|^2$$

$\Rightarrow \varepsilon(v) = 0 \Rightarrow v = 0$ . Or  $H^1(\Omega) \subset L^2(\Omega)$  étant compact  $v_{n'} \rightarrow 0$  dans  $L^2$  fort, ce qui contredit l'hypothèse  $\|v_{n'}\|_{L^2(\Omega)} = 1$ . #

Corollaire : Si  $\Gamma_0$  est de mesure non nulle et si  $E > 0$  et  $0 \leq \nu \leq \frac{1}{2}$ , le problème d'élasticité (3-3) admet une solution unique. #

Régularité : La régularité de la solution est analogue à celle obtenue pour le Laplacien. On consultera Necas <sup>(1)</sup> pour la régularité  $H^2$ , et Geymonat <sup>(2)</sup> pour la régularité  $W^{2,p}$ .

#### 4 - Modèle de Hencky

Une des caractéristiques du problème élastique étudié précédemment : est la linéarité.

Si l'on remplace les forces extérieures  $\vec{f}$  et  $\vec{F}$  par  $s\vec{f}$  et  $s\vec{F}$  où  $s$  est un coefficient réel, la solution  $u_s$  du problème élastique correspondant

$$(4-1) \quad a(u_s, v) = s L(v) \quad \forall v \in V$$

dépendra linéairement de  $s$  :  $u_s = s u$ .

Pour les métaux les choses se passent bien ainsi tant que le tenseur des contraintes  $\sigma$  ne dépasse pas localement un certain seuil appelé seuil de plasticité ; c'est-à-dire que  $\sigma(x) \in B$ , pour  $x \in \Omega$ , où  $B$  est un convexe donné de  $\mathbb{R}^9$  (contenant l'origine) appelé convexe de plasticité.

Soit  $\sigma^D \equiv \sigma - \frac{1}{3} \text{tr}(\sigma)\delta$  le déviateur de  $\sigma$ ,

c'est-à-dire ce qui reste lorsqu'on enlève à  $\sigma$  sa partie sphérique), le convexe  $B$  est très souvent de la forme

$$B = \{\sigma \in \mathbb{R}^9 : \mathcal{F}(\sigma^D) \leq 0\} \text{ où } \mathcal{F}: \mathbb{R}^9 \rightarrow \mathbb{R} \text{ est une fonction convexe donnée.}$$

En particulier le critère de Von Misès, correspond à

$$\mathcal{F}(\sigma) = |\sigma| - g$$

<sup>(1)</sup> NECAS : Méthodes directes dans la théorie des équations elliptiques, Masson, Paris, 196-222.

<sup>(2)</sup> GEYMONAT : Amali di Matematica Pura ed applicata 4, 69 (1965), 207-284.

où  $|\cdot|$  désigne la norme euclidienne de  $\mathbb{R}^9$ .

Le modèle de Hencky donne alors la relation suivante entre  $\varepsilon(u)$  et  $\sigma$

$$(4-2) \quad \varepsilon(u) = \Lambda^{-1} \sigma \in \partial I_p(\sigma) \quad \text{et} \quad \sigma \in P,$$

où  $P = \{\tau \in (L^2(\Omega))_S^9 : \tau(x) \in B, \text{ p.p. } x \in \Omega\}$

et  $(L^2(\Omega))_S^9$  désigne l'espace des tenseurs  $3 \times 3$  symétriques de carré intégrable.

L'interprétation locale (en un point  $x \in \Omega$ ) de cette relation est la suivante :

Si  $\sigma(x) \in \text{Int } B$ , alors  $\varepsilon(u) = \Lambda^{-1} \sigma$  (relation de type élastique)

sinon  $\varepsilon(u) = \Lambda^{-1} \sigma + \lambda$

où  $\lambda(\sigma - \phi) \geq 0, \quad \forall \phi \in B.$

La "déformation plastique"  $\lambda$  vérifie une propriété dite de "travail maximal".

### 5 - Formulation variationnelle

Soit  $(\cdot, \cdot)$  le produit scalaire de  $(L^2(\Omega))_S^9$  défini par

$$(\sigma, \tau) = \int_{\Omega} \sigma \cdot \tau \, dX$$

La relation (4-2) peut s'écrire

$$(5-1) \quad A(\sigma, \tau - \sigma) - (\varepsilon(u), \tau - \sigma) \geq 0 \quad \forall \tau \in P$$

où  $A(\sigma, \tau) \equiv (\Lambda^{-1} \sigma, \tau).$

Soit alors

$$M(s) = \{\tau \in (L^2(\Omega))_S^9 : (\tau, \varepsilon(v)) = sL(v) \quad \forall v \in V\}$$

L'ensemble des contraintes en équilibre pour les forces extérieures au

coefficient  $s$  près (cf. Lemme 1), le problème de Hencky est donc

(5-2) Trouver  $\sigma \in P \cap M(s)$  et  $u \in V$  satisfaisant (5-1)

En choisissant  $\tau \in P \cap M(s)$ , on voit que  $u$  disparaît, et que le problème se réduit à trouver  $\sigma \in P \cap M(s)$  solution de

$$(5-3) \quad \inf_{\tau \in P \cap M(s)} \frac{1}{2} A(\tau, \tau)$$

problème qui admet une solution unique à condition que  $P \cap M(s)$  soit non vide, ce qui est vrai pour  $s$  assez petit. Soit

$$\alpha = \sup \{s : P \cap M(s) \neq \emptyset\}$$

Le coefficient  $\alpha$  s'appelle le coefficient limite.

La situation est donc la suivante :

$0 \leq s \leq \alpha_e$       élasticité linéaire

$\alpha_e < s < \alpha$       élasto-plasticité

$s > \alpha$               pas de solution

où  $\alpha_e = \sup \{s : \Lambda \varepsilon(u_s) \in P\}$  et  $u_s$  est la solution du problème élastique (4-1).

En revanche le problème (5-2) n'a pas, en général, de solution même pour  $s < \alpha$ .

Cette propriété négative est encore une illustration du fait que la somme de deux opérateurs maximaux monotones n'est pas nécessairement maximale monotone ; en effet (4-2) peut aussi s'écrire

$$\Lambda^{-1} \sigma + \partial I_P(\sigma) + \partial I_M(\sigma) \ni 0$$

puisque  $\varepsilon(u) \in \partial I_M(\sigma)$  (voir l'analogie avec le problème élasto-plastique modèle introduit dans la partie A, § 4.3).

Physiquement le déplacement  $u$  peut avoir des discontinuités. (Il peut y avoir des lignes de glissement), et il n'est donc pas raisonnable d'espérer que le déplacement soit dans  $(H^1(\Omega))^3$ .

On montre <sup>(1)</sup> qu'il existe dans certains cas un déplacement dans  $(L^p(\Omega))^3$ , avec  $p > 1$ .

### 6 - Modèle de Prandtl-Reuss

Le modèle de Hencky est un modèle simplifié qui ne tient pas compte de la décharge. Supposons en effet que  $s$  croisse de 0 jusqu'à  $s_0 \in ]\alpha_e, \alpha[$  puis que  $s$  décroisse jusqu'à 0 ; dans un phénomène plastique réel, il y aura des déformations résiduelles, c'est-à-dire que le déplacement observé ne reviendra pas à 0.

Pour en tenir compte, on est obligé d'utiliser une loi incrémentale

$$\varepsilon(\dot{u}) - \Lambda^{-1} \dot{\sigma} \in \partial I_p(\sigma)$$

(où  $\dot{\sigma} \equiv \frac{\partial \sigma}{\partial t}$ ) dont l'interprétation locale montre qu'il y a une relation de type élastique ( $\varepsilon(\dot{u}) = \Lambda^{-1} \dot{\sigma}$ ) dans les deux cas suivants :

- $\sigma \in \text{Int } B$
- $\sigma \in \partial B$ , la frontière de  $B$ , et  $\Lambda^{-1} \dot{\sigma}$  est dirigé vers l'intérieur de  $B$  (décharge).

Le cas élasto-plastique survient lorsque  $\sigma \in \partial B$  et  $\Lambda^{-1} \dot{\sigma}$  est dirigé vers l'extérieur de  $B$ , on a alors

$$\varepsilon(\dot{u}) = \Lambda^{-1} \dot{\sigma} + \lambda \quad (\text{charge})$$

avec  $\lambda \cdot (\sigma - \tau) \geq 0, \forall \tau \in B$ .

---

(1) cf. R.S. FALK, B. MERCIER : *RAIRO Analyse Numérique* 11, (1977), 135-144.

Globalement, on a la formulation variationnelle

$$(6-1) \quad A(\dot{\sigma}, \tau - \sigma) - (\varepsilon(\dot{u}), \tau - \sigma) \geq 0 \quad \forall \tau \in P.$$

Cette relation est valable dans le cas dynamique, en rajoutant l'équation du mouvement (1-4) ou  $v = \dot{u}$ , et les conditions aux limites (2-6) et (2-7). Les actions extérieures  $\vec{f}$  et  $\vec{F}$  dépendent alors en général explicitement du temps.

Dans le cas quasi-statique, on néglige les termes d'accélération, et on peut éliminer  $u$ , d'où le problème aux contraintes : trouver  $\sigma \in K \cap M(t)$  tel que

$$(6-2) \quad \begin{aligned} A(\dot{\sigma}, \tau - \sigma) &\geq 0 \quad \forall \tau \in K \cap M(t) \\ \sigma(0) &= \sigma^0 \quad \sigma(t) \in K \cap M(t) \end{aligned}$$

où  $M(t)$  est le convexe des champs de contrainte en équilibre avec les actions extérieures massiques  $f(t)$  et surfaciques  $F(t)$ .

Là encore, on démontre <sup>(1)</sup> l'existence d'une solution au problème (6-2) à condition que  $K \cap M(t)$  soit un peu mieux que non vide pour tout  $t \geq 0$ , à condition que  $f$  et  $F$  soient assez réguliers en  $t$ , mais il n'existe pas en général de déplacement  $u$  vérifiant (6-1) autrement que dans un sens faible.

## 7 - Modèle élasto-visco-plastique

Soit  $\Pi$  la projection sur  $P$

et  $B_\mu \equiv \frac{1}{\mu} (\sigma - \Pi_\sigma)$  le régularisé Yosida de l'opérateur  $\partial I_p$ , la loi de comportement élasto-visco-plastique s'écrit

$$(7-1) \quad \varepsilon(\dot{u}) - \Lambda^{-1} \dot{\sigma} = B_\mu \sigma$$

qui peut aussi s'écrire, après élimination de  $\dot{u}$  :

<sup>(1)</sup> C. JOHNSON : Existence theorems for plasticity problems. *J. Math. pures et appliquées*, t. 55, (1976), 431-444.

Trouver  $\sigma(t) \in P$  satisfaisant

$$(7-2) \quad A(\dot{\sigma}, \tau - \sigma) + (B_{\mu} \sigma; \tau - \sigma) = 0 \quad \forall \tau \in M(t)$$

Le coefficient  $\mu$  s'interprète alors comme une viscosité, et le terme  $B$  comme un terme de rappel vers le convexe plasticité  $P$ , d'autant plus fort que  $\mu$  est petit.

En notant  $\sigma_{\mu}(t)$  la solution du problème élasto-visco-plastique (7-2) la méthode employée par Johnson pour démontrer l'existence d'une solution à (6-2) consiste à obtenir des estimations a priori sur  $\sigma_{\mu}$  et à passer à la limite lorsque  $\mu \rightarrow 0$ .

Ce modèle étudié par Duvaut et Lions est très riche et redonne, comme on va le voir, d'autres modèles par passage à la limite.

### 8 - Modèle rigide-visco-plastique

Supposons que le coefficient d'élasticité  $E$  tende vers l'infini (cf (2-3)) alors le terme en  $\Lambda^{-1} \dot{\sigma}$  deviendra négligeable dans (7-1) qui donnera à la limite la relation de comportement

$$(8-1) \quad \varepsilon(\dot{u}) = B_{\mu} \sigma \equiv \frac{1}{\mu} (\sigma - \Pi \sigma)$$

logiquement appelée rigide-visco-plastique. En effet, localement si  $\sigma(x) \in B$ , alors  $\varepsilon(\dot{u}) = 0$ , et le milieu continu se comporte localement de façon rigide.

La solution du problème obtenu en rajoutant la condition d'équilibre  $\sigma \in M(t)$  et la condition aux limites (2-6) montre que pour chaque  $t$ ,  $\sigma(t)$  et  $v(t) = \dot{u}(t)$  sont solution d'un problème stationnaire. Le milieu continu "oublie" donc complètement la configuration initiale, et c'est là où se situe la limite entre un solide et un fluide (on peut alors travailler en coordonnées purement Eulériennes).

Le champ de contraintes  $\sigma(t)$  est solution du problème stationnaire (dont la solution apparait comme indépendante de  $\mu$ )

$$(8-2) \quad \inf_{\tau \in M(t)} \frac{1}{2\mu} \|\tau - \Pi\tau\|^2$$

après quoi, la relation (8-1) permet de calculer  $v = \dot{u}$ .

On notera que si (8-2) a une solution, alors

$$(8-3) \quad P \cap M(t) = \emptyset \Leftrightarrow \frac{1}{\mu} (\sigma - \Pi\sigma) = \varepsilon(v) = 0 \Leftrightarrow v = 0$$

et le calcul des solutions de (8-2) est un moyen de déterminer le coefficient limite  $\alpha$  introduit au § 5. <sup>(1)</sup>

On peut facilement écrire le problème dual de (8-2) en remarquant que

$$\frac{\mu}{2} \|e\|^2 + j_p(e) \text{ est la conjuguée de } \frac{1}{2\mu} \|\tau - \Pi\tau\|^2$$

(cf Partie A du § 4), et le champ de vitesse  $v = \dot{u}$  est solution de

$$(8-4) \quad \inf_{w \in V} \left[ \frac{\mu}{2} \|\varepsilon(w)\|^2 + j_p(\varepsilon(w)) - L(w) \right]$$

qui redonne le problème de Mosolov dans le cas particulier d'un écoulement dans une conduite cylindrique (Partie A, § 1).

La fonction support de  $P : j_p(e) \equiv \sup_{\tau \in P} (\tau, e)$  n'est pas bornée si  $P$  n'est pas borné ; dans le cas du critère de Von Misès, on a

$$j_p(e) < +\infty \Leftrightarrow \text{tr}(e) = 0,$$

et il y a donc implicitement une contrainte  $\text{div } w = 0$  dans le problème (8-4).

Dans le cas dynamique, les équations du mouvement s'écrivent sous forme variationnelle

$$p\left(\frac{dv}{dt}, w\right) + (\sigma, \varepsilon(w)) = L(w) \quad \forall w \in V$$

(1) cf. B. MERCIER : *J. Mécanique* 16 (1977), 467-491.

On transforme alors la relation de comportement

$$\frac{1}{\mu} (\sigma - \Pi \sigma) = \varepsilon(v)$$

en 
$$\sigma - \mu \varepsilon(v) \in \partial j_p(\varepsilon(v))$$

c'est-à-dire

(8-5) 
$$j_p(e) - j_p(\varepsilon(v)) \geq (\sigma - \mu \varepsilon(v), e - \varepsilon(v))$$

d'où

(8-6) 
$$\rho \left( \frac{dv}{dt}, w-v \right) + \mu (\varepsilon(v), \varepsilon(w-v)) + j_p(\varepsilon(w)) - j_p(\varepsilon(v)) \geq L(w-v) \quad \forall w \in V.$$

Cette inéquation redonne les équations de Navier-Stokes dans le cas limite où le seuil de plasticité tend vers zéro, c'est-à-dire  $B = \{\tau \in \mathbb{R}^9 : |\tau^D| = 0\}$ . Dans ce cas en effet  $j_p(\varepsilon(v))$  se réduit à la fonction indicatrice de l'ensemble des champs de vitesse à divergence nulle.

Les composantes de la dérivée totale  $\frac{dv}{dt}$  s'écrivent

$$\frac{dv_i}{dt} = \frac{\partial v_i}{\partial t} + v_j \frac{\partial v_i}{\partial x_j}$$

et l'on peut souvent négliger les termes de convection, car les vitesses sont raisonnablement petites par rapport aux longueurs.

Le problème (8-6) peut alors s'écrire dans le cas où les actions extérieures surfaciques sont nulles ( $\vec{F} = 0$ )

(8-7) 
$$\begin{cases} \rho \frac{\partial v}{\partial t} + \mu A v + \partial j(v) \ni f(t) \\ v(0) = v^0 \end{cases}$$

où  $A$  est la restriction de  $H = (L^2(\Omega))^3$  de l'opérateur de  $V \rightarrow V'$  associé

à la forme bilinéaire  $(\varepsilon(u), \varepsilon(v))$ , et  $\partial j$  le sous différentiel de la fonction  $j(v) = j_p(\varepsilon(v))$ .

Dans le cas particulier d'un écoulement dans une conduite cylindrique, on a vu (Partie A) que l'opérateur  $A + \partial j$  est maximal monotone. Le problème est ouvert dans le cas général.

Le problème (8-7) est un problème d'évolution faisant intervenir la somme de deux opérateurs monotones et servira d'exemple à la troisième partie.

Remarque : Pour les métaux à haute température on adopte une loi de comportement dite pseudo-plastique, du type suivant

$$\sigma^D = C |\varepsilon(v)|^{p-2} \varepsilon(v)$$

et  $\operatorname{div} v = 0$  qui donne lieu à un problème bien posé dans  $(W^{1,p}(\Omega))^3$  pour les vitesses (avec en général  $p$  proche de 1). #

## 9 - Problèmes de frottement

Nous voulons maintenant envisager des conditions aux limites de type frottement. Pour simplifier nous nous replaçons dans le cadre du § 8 du matériau visco-plastique.

Soit  $\partial\Omega = \Gamma_0 \cup \Gamma_1 \cup \Gamma_2$ , nous prenons toujours les mêmes conditions aux limites sur  $\Gamma_0$  et  $\Gamma_1$ . Sur  $\Gamma_2$ , nous supposons qu'il y a contact unilatéral avec frottement entre le milieu continu et un obstacle donné.

On décompose  $\sigma \cdot \vec{n}$  et  $\vec{v}$  en partie normale et partie tangentielle (à  $\Gamma_2$ ) :

$$\sigma \cdot \vec{n} = \sigma_N \vec{n} + \vec{\sigma}_T, \quad \vec{v} = v_N \vec{n} + \vec{v}_T$$

et on impose les conditions

- $\sigma_N \leq 0$  (l'obstacle ne peut pas tirer le milieu continu)
- $|\vec{\sigma}_T| \leq g$  (l'obstacle n'exerce pas sur le milieu continu d'effort tangentiel supérieur à un certain seuil)

- $\vec{v}_T = \begin{cases} -\lambda \vec{\sigma}_T & \text{avec } \lambda \geq 0 \text{ si } |\sigma_T| = g \\ 0 & \text{si } |\sigma_T| < g \end{cases}$
- $v_N \leq 0$  si  $\sigma_N = 0$  (décollement)
- $v_N = 0$  si  $\sigma_N < 0$  (contact maintenu)

L'interprétation de la condition  $\vec{v}_T = -\lambda \vec{\sigma}_T$  si  $|\sigma_T| = g$ , est que si le milieu continu bouge par rapport à l'obstacle, l'obstacle réagit sur lui de manière à s'opposer au mouvement (donc dans la direction opposée au mouvement) mais avec une intensité inconnue, d'où le paramètre  $\lambda$ .

Soit alors  $C \subset \mathbb{R}^3$  le convexe

$$C = \{ \vec{\tau} = \tau_N \vec{n} + \vec{\tau}_T : \tau_N \leq 0, |\tau_T| \leq g \}$$

On vérifie que

$$(9-1) \quad -\vec{v} \in \partial I_C(\sigma \cdot \vec{n}) \quad \text{sur } \Gamma_2$$

est équivalent à l'ensemble des conditions aux limites énoncées plus haut.

En résumé, nous avons les conditions aux limites

$$\vec{v} = 0 \text{ sur } \Gamma_0, \quad \sigma \cdot \vec{n} = \vec{F} \text{ sur } \Gamma_1, \quad (9-1) \text{ sur } \Gamma_2$$

l'équation d'équilibre  $\text{div } \sigma + \vec{f} = 0$  sur  $\Omega$

et la relation du comportement :  $\varepsilon(v) = B_\mu \sigma$  sur  $\Omega$

Pour donner une formulation variationnelle à ce problème nous remarquons que (9-1) s'écrit aussi

$$\sigma \cdot \vec{n} \in \partial j_C(-\vec{v}) \text{ sur } \Gamma_2$$

(où  $j_C(v) \equiv \int_{\Gamma_2} g |\vec{v}_T| ds$  si  $v_N \leq 0$  sur  $\Gamma_2$  et  $+\infty$  sinon ; est l'intégrale de la fonction support de  $C$ ).

C'est-à-dire  $j_C(-\vec{w}) - j_C(-\vec{v}) \geq \int_{\Gamma_2} (\sigma \cdot \vec{n}) \cdot (-\vec{w} + \vec{v}) ds$

or d'après la formule de Green

$$\begin{aligned} \int_{\Gamma_2} -(\sigma \cdot \vec{n}) \cdot (-\vec{w} + \vec{v}) &= \int_{\Omega} \sigma \cdot \varepsilon(\vec{v} - \vec{w}) - \int_{\Omega} \vec{f} \cdot (\vec{v} - \vec{w}) - \int_{\Gamma_1} \vec{F} \cdot (\vec{v} - \vec{w}) ds \\ &= (\sigma, \varepsilon(v-w)) - L(v-w) \end{aligned}$$

on a vu (cf (8-5)) que

$$(\sigma, \varepsilon(w-v)) \leq j_p(\varepsilon(w)) - j_p(\varepsilon(v)) + \mu(\varepsilon(v), \varepsilon(w-v))$$

On en tire

$$j_C(-\vec{w}) - j_C(-\vec{v}) \geq j_p(\varepsilon(v)) - j_p(\varepsilon(w)) - L(v-w) - \mu(\varepsilon(v), \varepsilon(w-v))$$

et finalement

$$\mu(\varepsilon(v), \varepsilon(w-v)) + j_p(\varepsilon(v)) - j_p(\varepsilon(w)) + j_C(-\vec{w}) - j_C(-\vec{v}) - L(w-v) \geq 0 \quad \forall w \in V$$

qui équivaut au problème d'optimisation

$$\inf_{w \in V} \left[ \frac{\mu}{2} \|\varepsilon(w)\|^2 + j_p(\varepsilon(w)) + j_C(-w) - L(w) \right] .$$

L'effet du frottement est donc de rajouter à la fonctionnelle d'énergie un terme de dissipation.

C - ANALYSE NUMERIQUE DES INEQUATIONS D'EVOLUTION

Plan

1. Rappel de résultats d'existence
2. Le résultat de Brézis-Pazy
3. Application à la somme de deux opérateurs
4. Application aux algorithmes III et IV
5. Approximation

C - ANALYSE NUMERIQUE DES INEQUATIONS D'EVOLUTION

1. Rappel de résultats d'existence

On adopte les mêmes notations que dans la partie A, § 2 :  
H est un espace de Hilbert muni du produit scalaire  $(\cdot, \cdot)$  et de la norme  $|\cdot|$ .

Soit A un opérateur maximal monotone, on appelle  $A^0$  sa section principale, c'est-à-dire l'opérateur univoque tel que  $A^0x$  soit la projection de x sur l'ensemble (convexe fermé)  $A(x)$ .

On rappelle quelques propriétés de l'opérateur  $A^0$ , démontrées dans BREZIS <sup>(1)</sup>.

Lemme 1 : Soit  $A_\lambda = \frac{1}{\lambda} (I - J_A^\lambda)$  le régularisé Yosida de A, alors  
=====

$$A_\lambda x \rightarrow A^0 x \text{ lorsque } \lambda \rightarrow 0$$

Lemme 2 : Si  $\{x, y\} \in \overline{D(A)} \times H$  vérifie  
=====

$$(y - A^0 \xi, x - \xi) \geq 0 \quad \forall \xi \in D(A)$$

alors  $y \in Ax$ . #

On cherche à résoudre le problème d'évolution :

$$(1-1) \quad \begin{aligned} \frac{du}{dt} + Au &\ni 0 \\ u(0) &= u_0 \end{aligned}$$

On rappelle <sup>(1)</sup> que pour tout  $u_0 \in D(A)$  ce problème admet une solution unique  $u(t) = S(t) u_0$  dépendant non linéairement de la donnée initiale  $u_0$  par l'intermédiaire d'une famille de contractions  $S(t)$  appelée semi-groupe de contraction.

---

(1) H. BREZIS : Opérateurs maximaux monotones, North Holland, Lecture Notes, 1973, pp. 28-29.

La monotonie de  $A$  entraîne en effet que, si  $v(t) = S(t) v_0$

$$\left(-\frac{du}{dt} + \frac{dv}{dt}, u-v\right) \geq 0 \quad \forall t,$$

d'où

$$\frac{1}{2} \frac{d}{dt} |u-v|^2 \leq 0 \Rightarrow |u(t) - v(t)|^2 \leq |u(0) - v(0)|^2$$

c'est-à-dire

$$|S(t) v_0 - S(t) u_0| \leq |v_0 - u_0|.$$

Le caractère de semi-groupe est dû à la relation immédiate

$$S(t_1+t_2) = S(t_1) S(t_2).$$

Pour établir l'existence de  $u(t)$ , la méthode consiste à obtenir des estimations et à passer à la limite sur l'équation

$$\frac{du_\lambda}{dt} + A_\lambda u_\lambda = 0 \quad u_\lambda(0) = u_0,$$

qui admet une solution d'après le théorème de Cauchy-Lipschitz car l'opérateur  $A_\lambda$  est Lipschitzien.

Au cours de la démonstration (BREZIS, p. 56) on établit l'estimation d'erreur :

$$(1-2) \quad |u_\lambda(t) - u(t)| \leq \sqrt{\lambda t} |A^0 u_0|.$$

## 2. Le résultat de BREZIS et PAZY

Pour résoudre l'équation d'évolution (1-1) on peut utiliser le "schéma" implicite suivant :

$$(2-1) \quad \frac{u^{n+1} - u^n}{\Delta t} + A(u^{n+1}) \ni 0$$

où  $u^0 = u_0$ ,  $\Delta t$  est le pas de temps et  $u^n$  représente une "approximation" de la solution de (1-1) en  $t = n\Delta t$ .

Plus généralement, on considère des "schémas", c'est-à-dire des algorithmes du type suivant :

$$(2-2) \quad u^{n+1} = F(\lambda)u^n$$

où  $\lambda = \frac{t}{n}$ , et  $F(\lambda)$  est un opérateur donné (en général une contraction, au moins pour  $\lambda$  assez petit).

Il y aura convergence si

$$(2-3) \quad u^n = F\left(\frac{t}{n}\right)^n u_0 \rightarrow u(t) = S(t) u_0$$

lorsque  $n \rightarrow +\infty$ .

Nous allons établir un résultat dû à Brézis-Pazy qui donne une condition suffisante de convergence.

Théorème 1 : Soit  $C = \overline{D(A)}$  et  $F(\lambda) : C \rightarrow C$  une famille de contractions vérifiant la propriété de consistance

$$(2-4) \quad \lim_{\lambda \rightarrow 0} \frac{x - F(\lambda)x}{\lambda} = A^0 x \quad \forall x \in D(A)$$

alors si  $A$  est maximal monotone, il y a convergence, c'est-à-dire que (2-3) a lieu uniformément pour  $t \in [0, T]$  et  $T$  fini.

Dans le cas du schéma implicite (2-1),  $F(\lambda) \equiv J_A^\lambda$ , qui est bien une contraction. La propriété de consistance (2-4) résulte alors immédiatement du Lemme 1.

La démonstration du théorème 1 repose sur les deux résultats préliminaires suivants.

Lemme 3 : Soit  $A^n$  une famille d'opérateurs maximaux monotones vérifiant, pour  $\rho > 0$

$$(I + \rho A^n)^{-1} x \rightarrow (I + \rho A)^{-1} x \quad \forall x \in D(A)$$

et  $u^n$  la solution de

$$(2-5) \quad \frac{du^n}{dt} + A^n u^n \ni 0$$

$$u^n(0) = u_{0n}$$

où  $u_{0n} \rightarrow u_0$  lorsque  $n \rightarrow \infty$ ; alors  $u^n(t) \rightarrow u(t)$  uniformément sur  $[0, T]$  pour  $T$  fini.

Ce résultat est démontré dans BREZIS, loc. cit. p. 102. C'est un cas particulier du résultat établi au § 4 ci-dessous.

Lemme 4 : (estimation de Chernoff)

===== Soit  $J$  une contraction et  $u$  vérifiant

$$\dot{u} + (u - Ju) = 0 \quad (\text{où } \dot{u} \equiv \frac{du}{dt})$$

Alors  $|u(t) - J^n u(0)| \leq [(n-t)^2 + t]^{1/2} |u(0) - Ju(0)|$

Démonstration : On pose  $u_0 = u(0)$

$$1^\circ) - \text{Soit } \phi_n(t) = |u(t) - J^n u_0| |u_0 - Ju_0|^{-1}$$

$$\text{Montrons que } \phi_n(t) \leq n e^{-t} + \int_0^t e^{s-t} \phi_{n-1}(s) ds$$

Par intégration par parties on a

$$\int_0^t \dot{u}(s) e^s ds = |u(s)e^s|_0^t - \int_0^t u(s)e^s ds$$

$$\text{d'où } \int_0^t e^s Ju(s) ds = u(t)e^t - u(0)$$

$$\Rightarrow u(t) = e^{-t}u(0) + \int_0^t e^{s-t} Ju(s) ds$$

De même (prendre  $u \equiv 1$ )

$$1 = e^{-t} + \int_0^t e^{s-t} ds$$

d'où

$$u(t) - J^n u_0 = (u_0 - J^n u_0) e^{-t} + \int_0^t e^{s-t} (Ju(s) - J^n u_0) ds$$

$$|u(t) - J^n u_0| \leq e^{-t} |u_0 - J^n u_0| + \int_0^t e^{s-t} |u(s) - J^{n-1} u_0| ds$$

puisque  $J$  est une contraction. On a aussi

$$|u_0 - J^n u_0| \leq \sum_{i=1}^n |J^{i-1} u_0 - J^i u_0| \leq n |u_0 - Ju_0|$$

d'où le résultat.

2°)- Montrons que  $\phi_0(t) = |u(t) - u^0| |u_0 - Ju_0|^{-1} \leq t$  ;

$J$  étant une contraction  $I-J$  est monotone, et  $u(0) \rightarrow u(t)$  est donc un semi groupe de contraction, d'où

$$|u(t+h) - u(t)| \leq |u(h) - u(0)|$$

En dérivant par  $h$  et en faisant  $h \rightarrow 0$ , on obtient

$$|\dot{u}(t)| \leq |u(0)| = |u(0) - Ju(0)| \equiv k$$

on a donc  $|u(t) - u(0)| \leq t \sup_s |\dot{u}(s)| \leq kt$

d'où le résultat.

3°)- Démontrons par récurrence que

$$\phi_n(t) \leq ((n-t)^2 + t)^{1/2}$$

ce qui établira le résultat.

Si donc  $\phi_{n-1}(s) \leq ((n-1-s)^2+s)^{1/2}$ , il s'agit de vérifier que l'on a bien

$$ne^{-t} + \int_0^t e^{s-t} ((n-1-s)^2+s)^{1/2} ds \leq ((n-t)^2+t)^{1/2}$$

c'est-à-dire

$$f(t) \leq g(t)$$

où  $f(t) = n + \int_0^t e^s ((n-1-s)^2+s)^{1/2}$  et  $g(t) = e^t ((n-t)^2+t)^{1/2}$

Or  $f(0) = g(0)$ , il suffit donc de vérifier que  $f'(t) \leq g'(t)$ , pour tout  $t \geq 0$ . C'est-à-dire

$$e^t ((n-1-t)^2+t)^{1/2} \leq e^t ((n-t)^2+t)^{1/2} + e^t ((n-t)^2+t)^{-1/2} \frac{1}{2} (2(t-n)+1)$$

ou encore, en divisant par  $e^t$ , en multipliant par  $((n-t)^2+t)^{1/2}$

$$(2-6) \quad ((n-1-t)^2+t)^{1/2} ((n-t)^2+t)^{1/2} \leq (n-t)^2+t+t-n + \frac{1}{2}$$

On pose alors  $p = n-t-\frac{1}{2}$

$$d'où \quad (n-t)^2 = (p+\frac{1}{2})^2 = p^2 + p + \frac{1}{4}$$

$$(n-t-1)^2 = (p-\frac{1}{2})^2 = p^2 - p + \frac{1}{4}$$

L'inégalité (2-6) s'écrit donc

$$(p^2-p+\frac{1}{4}+t)^{1/2} (p^2+p+\frac{1}{4}+t)^{1/2} \leq (p^2+p+\frac{1}{4})+t-p = p^2+\frac{1}{4}+t$$

soit en posant  $q \equiv p^2 + t + \frac{1}{4}$

$$(q-p)^{1/2}(q+p)^{1/2} \leq q$$

ce qui est vrai puisque  $q$  est positif.

#

Corollaire 1 : Si u vérifie

$$\dot{u} + \frac{1}{\lambda} (u - Ju) = 0$$

alors on a

$$|u(t) - J^n u(0)| \leq \left[ \left(n - \frac{t}{\lambda}\right)^2 + \frac{t}{\lambda} \right]^{1/2} |u(0) - Ju(0)| \quad \#$$

(Il suffit en effet d'effectuer le changement de variable  $\frac{t}{\lambda} \rightarrow t$ ).

Démonstration du théorème 1 :

Nous nous plaçons pour simplifier dans le cas où  $C \equiv \overline{D(A)} = H$ .  
(le cas général est traité dans BREZIS, loc.cit.).

On pose  $C_\lambda = \frac{1}{\lambda} (I - F(\lambda))$ .

1°)- Démontrons que, pour  $\rho > 0$

$$(2-7) \quad u_\lambda \equiv (I + \rho C_\lambda)^{-1} x \rightarrow (I + \rho A)^{-1} x$$

lorsque  $\lambda \rightarrow 0$ , pour tout  $x \in \overline{D(A)}$ . En remarquant que

$$u_\lambda + \rho C_\lambda u_\lambda = x \quad \text{et donc} \quad \frac{1}{\rho} (x - u_\lambda) \in C_\lambda u_\lambda$$

et que pour  $\xi \in D(A)$ ,  $\eta_\lambda \equiv C_\lambda \xi \rightarrow A^0 \xi$  par hypothèse, on a, en utilisant la monotonie de  $C_\lambda$  :

$$(2-8) \quad \left( \frac{1}{\rho} (x - u_\lambda) - \eta_\lambda, u_\lambda - \xi \right) \geq 0$$

Donc  $|u_\lambda|$  est borné. En extrayant une sous-suite  $u_{\lambda'}$ ,  $\rightarrow u$  lorsque  $\lambda' \rightarrow 0$ , on obtient

$$(2-9) \quad \left( \frac{1}{\rho} (x - u) - A^0 \xi, u - \xi \right) \geq 0 \quad \forall \xi \in D(A)$$

d'où  $\frac{x-u}{\rho} \in Au$  d'après le Lemme 2.

Finalement  $u = J_A^0 x$ , unique, d'où la convergence faible. Pour

établir la convergence forte, on choisit  $\xi = u$   
d'où

$$\limsup_{\lambda \rightarrow 0} |u_\lambda|^2 \leq \lim_{\lambda \rightarrow 0} (x - \rho n_\lambda, u_\lambda - \xi) + (\xi, u_\lambda) = (\xi, u) = |u|^2. \quad \#$$

2°)- D'après l'estimation de Chernoff (Corollaire 1), on a

$$(2-10) \quad |S_\lambda(t)x - F(\lambda)^n x| \leq \left( \left(n - \frac{t}{\lambda}\right)^2 + \frac{t}{\lambda} \right)^{1/2} |x - F(\lambda)x| = (\lambda t)^{1/2} \left| \frac{x - F(\lambda)x}{\lambda} \right|$$

si l'on choisit  $\lambda = \frac{t}{n}$ ;  $u_\lambda(t) \equiv S_\lambda(t)x$  est la solution de

$$\frac{du_\lambda}{dt} + C_\lambda u_\lambda = 0, \quad u_\lambda(0) = x,$$

c'est-à-dire que  $S_\lambda$  est le semi-groupe associé à  $C_\lambda$ .

Finalement, on a avec  $\rho = \frac{t}{n}$

$$\left| F\left(\frac{t}{n}\right)^n x - S(t)x \right| \leq \left| F(\lambda)^n x - S_\lambda(t)x \right| + |S(t)x - S(t)x|$$

où le 1er terme tend vers zéro lorsque  $n \rightarrow \infty$  et donc  $\rho \rightarrow 0$

d'après (2-10) et l'hypothèse (2-4); d'autre part, le lemme 1 montre que le 2è terme converge vers zéro, d'où le résultat. #

### COMMENTAIRE

Si  $F(\lambda)$  admet un point fixe  $\bar{u}$ ,  $F(\lambda)$  étant une contraction on aura  $|u^n - \bar{u}| \leq |u^0 - \bar{u}|$ ,  $\forall n$ , ce qui démontre la stabilité. Le théorème 1 est donc une illustration du théorème de Lax :

$$\underline{\text{Stabilité}} + \underline{\text{consistance}} \Rightarrow \underline{\text{convergence}}$$

Dans de nombreux cas <sup>(1)</sup> (par exemple  $F(\lambda) = J_A^\lambda$ ) les points fixes de  $F(\lambda)$  seront aussi solutions du problème stationnaire

$$A(\bar{u}) = 0.$$

(1) Pour tous les algorithmes étudiés à la partie A sauf l'algorithme II.

La bornitude de  $u^n$  est donc liée à l'existence d'une solution pour ce problème. Noter que des résultats récents <sup>(1)</sup> ont montré que la non-existence de  $\bar{u}$  entraîne  $u(t) \rightarrow +\infty$  si  $t \rightarrow \infty$ . Dans ce cas, il n'est évidemment pas raisonnable d'espérer que  $u^n$  reste borné pour tout  $n$ .

Remarque 1 : On peut remplacer (2-4) par la condition :

$\forall \xi \in D(A)$  il existe  $\xi_\lambda \rightarrow \xi$  tel que

$$(2-11) \quad \lim_{\lambda \rightarrow 0} \frac{\xi_\lambda - F(\lambda)\xi_\lambda}{\lambda} = A^0 \xi$$

La démonstration s'adapte de la façon suivante. On remplace  $\xi$  par  $\xi_\lambda$  dans (2-8) en choisissant  $\eta_\lambda = C_\lambda \xi_\lambda$ , et on obtient toujours (2-9) à la limite.

Dans la 2<sup>ème</sup> partie, on écrit, avec  $\lambda = \frac{t}{n}$  :

$$\begin{aligned} |F(\frac{t}{n})^n x - S(t)x| &\leq |F(\lambda)^n x - F(\lambda)^n x_\lambda| + |F(\lambda)^n x_\lambda - S_\lambda(t)x_\lambda| \\ &\quad + |S_\lambda(t)x_\lambda - S_\lambda(t)x| + |S_\lambda(t)x - S(t)x| \end{aligned}$$

Seuls le 1<sup>er</sup> et 3<sup>è</sup> terme sont nouveaux. Tous les deux sont majorés par  $|x_\lambda - x|$  qui tend vers zéro quand  $\lambda \rightarrow 0$ .

Remarque 2 : La condition (2-7) entraîne (2-11).

En effet si l'on a (2.7), on a

$$y_\lambda = (I + \rho \frac{1}{\lambda} (I - F(\lambda)))^{-1} x \rightarrow y = (I + \rho A)^{-1} x$$

alors

$$x = y_\lambda + \rho \frac{y_\lambda - F(\lambda)y_\lambda}{\lambda} = y + \rho a \quad \text{où} \quad a \in Ay$$

<sup>(1)</sup> PAZY, *J. Funct. Anal.* 27 (1978), 292-307.

$$\Rightarrow \frac{y_\lambda - F(\lambda)y_\lambda}{\lambda} = a + \frac{1}{\rho} (y - y_\lambda)$$

d'où

$$\Rightarrow \lim \frac{y_\lambda - F(\lambda)y_\lambda}{\lambda} = a \in Ay \quad \text{et} \quad y_\lambda \rightarrow y$$

d'où (2-11).

#

Applications : En plus du schéma implicite signalé plus haut qui correspond au choix  $F(\lambda) = J_A^\lambda$ , on peut choisir

$$(2-12) \quad F(\lambda) = 2(I + \frac{\lambda}{2} A)^{-1} - I \equiv 2J_A^{\lambda/2} - I$$

qui est une contraction  $\forall \lambda > 0$  <sup>(1)</sup> et vérifie (2-4). En effet

$$\frac{1}{\lambda} (x - F(\lambda)x) = \frac{x - J_A^{\lambda/2} x}{\lambda/2} = A_{\lambda/2} x \rightarrow A^0 x \quad \lambda \rightarrow 0$$

Dans le cas linéaire on retrouve ainsi le schéma de Crank-Nicholson, qui a une précision en  $O(\lambda^2)$  alors que le schéma implicite est en  $O(\lambda)$  seulement.

Dans le cas où  $A$  est lipschitzien on peut choisir

$$F(\lambda) = I - \lambda A$$

qui correspond à un schéma explicite, et vérifie trivialement la propriété de consistance. On a vu que dans certains cas,  $F(\lambda)$  est alors une contraction pour  $\lambda$  assez petit (cf Partie A, § 2).

On peut aussi choisir pour  $0 < \theta < 1$

$$(2-13) \quad F(\lambda) = J_A^{\lambda\theta} (I - (1-\theta)\lambda A)$$

qui donne un algorithme intermédiaire entre explicite ( $\theta = 0$ ) et implicite ( $\theta = 1$ ). Le cas  $\theta = \frac{1}{2}$  correspond au schéma de Crank-Nicholson. Cet algorithme n'est que conditionnellement stable sauf dans le cas  $\theta = 1$ .

---

(1) cf. Théorème 2.2, partie A.

Montrons qu'il est consistant : on pose  $\mu = \lambda\theta$  et  $x_\mu = x - (1-\theta)\lambda Ax$

$$\frac{x - F(\lambda)x}{\lambda} = \theta \left( \frac{x - x_\mu}{\mu} + \frac{x_\mu - J_A^\mu x_\mu}{\mu} \right)$$

où  $\frac{x - x_\mu}{\mu} \equiv \frac{1-\theta}{\theta} Ax$

et  $\frac{x_\mu - J_A^\mu x_\mu}{\mu} \rightarrow Ax$  lorsque  $\mu \rightarrow 0$ , puisque  $A$  est univoque. (Appliquer le lemme 5 ci-dessous).

Finalement  $\frac{x - F(\lambda)x}{\lambda} \rightarrow ((1-\theta)Ax + \theta Ax) = Ax$ , ce qui démontre la consistance de (2-13). #

### 3. Application à la somme de deux opérateurs

On considère l'équation d'évolution

$$(3-1) \quad \begin{aligned} \frac{\partial u}{\partial t} + Au + Bu &\ni 0 \\ u(0) &= u_0 \end{aligned}$$

où  $A, B$  et  $A+B$  sont maximaux monotones.

On va donner plusieurs exemples d'algorithmes du type (2.2) convergent vers la solution du problème.

Tout d'abord un résultat assez général, dans le cas où  $B$  est univoque.

Théorème 2 : Soit  $G(\lambda)$  une famille de contractions consistante avec  $B$  supposé univoque

$$(3-2) \quad \lim_{\lambda \rightarrow 0} \frac{x - G(\lambda)x}{\lambda} = Bx \quad \forall x \in D(B)$$

alors  $F(\lambda) = J_A^\lambda \circ G(\lambda)$  est consistant avec  $A+B$  :

$$\lim_{\lambda \rightarrow 0} \frac{x - F(\lambda)x}{\lambda} = (A+B)^0 x \quad \forall x \in D(A+B)$$

On établit le résultat préliminaire :

Lemme 5 : Soit C un opérateur maximal monotone,  $u \in D(C)$  et  $u_\lambda$  une suite  
telle que  $\lim_{\lambda \rightarrow 0} \frac{u_\lambda - u}{\lambda} = y$ , alors

$$\lim_{\lambda \rightarrow 0} \frac{u_\lambda - J_C^\lambda u_\lambda}{\lambda} = P_{Cu}(y)$$

où  $P_{Cu}(\cdot)$  désigne la projection sur l'ensemble  $Cu$ .

Démonstration : Soit E l'opérateur défini par  $Eu = Cu - y$ . On a pour  $u \in D(A)$

$$J_E^\lambda(u_\lambda - y) = J_C^\lambda u_\lambda$$

Soit  $v_\lambda = u_\lambda - \lambda y$ , on a

$$\frac{1}{\lambda} (v_\lambda - J_E^\lambda v_\lambda) = \frac{1}{\lambda} (v_\lambda - u) + \frac{1}{\lambda} (u - J_E^\lambda u) + \frac{1}{\lambda} (J_E^\lambda u - J_E^\lambda v_\lambda)$$

Or

$$\lim_{\lambda \rightarrow 0} \frac{1}{\lambda} (v_\lambda - u) = 0 \quad \text{par hypothèse.}$$

(cf Lemme 1)

$$E_\lambda u \equiv \frac{1}{\lambda} (u - J_E^\lambda u) \rightarrow E^0 u \equiv P_{Eu}(0), \text{ et } J_E^\lambda \text{ étant une contraction}$$

$$\frac{1}{\lambda} |J_E^\lambda u - J_E^\lambda v_\lambda| \leq \frac{1}{\lambda} |u - v_\lambda| \rightarrow 0 \text{ quand } \lambda \rightarrow 0.$$

Finalement

$$\lim_{\lambda \rightarrow 0} \frac{1}{\lambda} (v_\lambda - J_E^\lambda v_\lambda) = P_{Eu}(0)$$

d'où le résultat.

#

Démonstration du théorème 2 :

On pose  $u_\lambda = G(\lambda)x$  d'où  $\frac{x-u_\lambda}{\lambda} \rightarrow b \equiv Bx$  par hypothèse. On a donc

$$\frac{x-F(\lambda)x}{\lambda} = \frac{x-u_\lambda}{\lambda} + \frac{u_\lambda - J_A u_\lambda}{\lambda} \quad B + P_{Ax}(-b)$$

en appliquant le Lemme 5. On conclut en remarquant que

$$z = b + P_{Ax}(-b) \text{ vérifie}$$

$$(-b - (z-b), (z-b)-a) \geq 0 \quad \forall a \in Ax$$

et donc  $z = P_{(A+B)x}(0) \equiv (A+B)^0 x.$  #

Applications

Le résultat précédent s'applique immédiatement à l'algorithme I qui correspond au choix  $G(\lambda) = I - \lambda B$  (et vérifie trivialement (3-2)).

On peut aussi choisir  $G(\lambda) = 2J_B^{\lambda/2} - I$  ou  $G(\lambda) = J_B^{\lambda\theta} (I - (1-\theta)\lambda B)$  comme en (2-12) ou en (2-13).

On peut aussi choisir  $F(\lambda) = J_A^\lambda J_B^\lambda$  comme dans l'algorithme II. Dans ce dernier cas, on peut démontrer la convergence de  $F(\frac{t}{n})^n u_0$  vers  $u(t)$  solution de (3-1), même dans le cas où  $B$  est multivoque, en établissant directement (2-7) qui suffit, comme on l'a signalé à la remarque 2 (cf BREZIS, loc. cit. Prop. 4-3).

Signalons encore la possibilité de choisir  $G(\lambda) = S_B(\lambda)$  ou  $S_B$  est le semi-groupe associé à  $B$ .

Application aux algorithmes III et IV

L'algorithme III correspond au choix

$$F(\lambda) = J_A^\lambda (2J_B^\lambda - I) + I - J_B^\lambda$$

$$v^{n+1} = F(\lambda) v^n$$

avec  $v^0 = u^0 + \lambda b^0$  et  $b^0 \in Bu^0$  donné.

Nous allons montrer dans un premier temps que

$$(3-3) \quad F\left(\frac{t}{n}\right)^n u^0 \rightarrow u(t) \quad n \rightarrow \infty$$

On en déduira alors que  $v^n \rightarrow u(t)$  puisque  $F(\lambda)$  est une contraction

$$|v^n - F\left(\frac{t}{n}\right)^n u^0| \leq \frac{t}{n} |b^0|$$

et aussi que  $u^n \equiv J_B^{t/n} v^n \rightarrow u(t)$ .

Pour montrer (3-3) il suffit de montrer la consistance (2-11) de  $F(\lambda)$ .

Soit alors  $\xi \in D(A+B)$ , on pose  $(A+B)^0 \xi = a+b$  ou  $a \in A\xi$ ,  $b \in B\xi$ .  
De l'inéquation

$$(-(a+b), a+b-c) \geq 0 \quad \forall c \in (A+B)(\xi)$$

on déduit, en choisissant  $c = a'+b$  avec  $a' \in Ax$

$$((-b)-a, a-a') \geq 0, \quad \forall a' \in A\xi,$$

et donc  $a = P_{A\xi}(-b)$ .

On choisit  $\xi_\lambda = \xi + \lambda b$  (d'où  $\xi = J_B^\lambda \xi_\lambda$ ) et

$$\begin{aligned} \frac{\xi_\lambda - F(\lambda)\xi_\lambda}{\lambda} &= \frac{1}{\lambda} (\xi - J_A^\lambda(2\xi - \xi_\lambda)) \\ &= \frac{1}{\lambda} (\xi - J_A^\lambda(\xi - \lambda b)) \\ &= b + \frac{1}{\lambda} (u_\lambda - J_A^\lambda u_\lambda), \end{aligned}$$

où l'on a posé  $u_\lambda = \xi - \lambda b$ . D'après le Lemme 5, on en déduit que

$$\lim_{\lambda \rightarrow 0} \frac{\xi_\lambda - F(\lambda)\xi_\lambda}{\lambda} = b + P_{A\xi}(-b) = (A+B)^0(\xi)$$

ce qui établit la consistance (2-11).

Pour l'algorithme IV, on démontrera de la même façon que

$$F(\lambda) = (2J_A^{\lambda/2} - I) (2J_B^{\lambda/2} - I)$$

vérifie (2-11). Cet algorithme a l'avantage important d'être précis en  $O(\lambda^2)$  dans le cas linéaire.

En conclusion on retiendra qu'il y a beaucoup plus de souplesse dans le choix d'un algorithme convergeant vers la solution du problème d'évolution ( $u^n \rightarrow u(t)$  lorsque  $n \rightarrow \infty$  et  $\lambda = \frac{t}{n} \rightarrow 0$ ) que vers celle du problème stationnaire ( $u^n \rightarrow \bar{u}$  lorsque  $n \rightarrow \infty$  et  $\lambda$  fixé).

Les algorithmes I, III et IV ont l'avantage de posséder les deux propriétés ; III et IV étant en plus inconditionnellement stable et acceptant A et B multivoque. L'algorithme IV est en plus d'ordre 2 dans le cas linéaire. Certains auteurs l'appelle "algorithme universel".

#### 4. Approximation des problèmes paraboliques

On se donne une famille d'espaces de dimension finie  $V_n$  et des opérateurs maximaux monotones  $A^n : V_n \rightarrow V_n$ . On considère le problème "approché" :

$$(4-1) \quad \begin{aligned} \frac{du_n}{dt} + A^n u_n &\ni 0 \\ u_n(0) &= u_{0n} \end{aligned}$$

où  $u_{0n} = P_n u_0$  et  $P_n : H \rightarrow V_n$  désigne la projection sur  $V_n$ .

Théorème 3 : On fait les hypothèses suivantes :  $\forall z \in D(A)$

$$(4-2) \quad \begin{cases} P_n z \rightarrow z, \text{ lorsque } n \rightarrow +\infty \\ (I + \lambda A^n)^{-1} P_n z \rightarrow (I + \lambda A)^{-1} z, \text{ lorsque } n \rightarrow +\infty \end{cases}$$

et pour tout  $\lambda > 0$ . Alors  $u_n(t)$  solution de (4-1) converge uniformément vers  $u(t)$  solution de (1-1) pour  $t \in [0, T]$  pour tout  $T$  fini.

Démonstration :

Posons  $y_n = (I + \sqrt{\lambda} A^n)^{-1} P_n u_0$  et  $y = (I + \sqrt{\lambda} A)^{-1} u_0$

on introduit  $v_n(t)$ ,  $v_{n,\lambda}(t)$  solutions dans  $V_n$  de

$$\frac{dv_n}{dt} + A^n v_n = 0, \quad v_n(0) = y_n$$

$$\frac{dv_{n,\lambda}}{dt} + A_\lambda^n v_{n,\lambda} = 0, \quad v_{n,\lambda}(0) = y_n,$$

et  $v(t)$ ,  $v_\lambda(t) \in H$ , solutions de

$$\frac{dv}{dt} + Av = 0, \quad v(0) = y,$$

$$\frac{dv_\lambda}{dt} + A_\lambda v_\lambda = 0, \quad v_\lambda(0) = y.$$

On va écrire

$$(4-3) \quad |u_n(t) - u(t)| = |u_n(t) - v_n(t)| + |v_n(t) - v_{n,\lambda}(t)| + |v_{n,\lambda}(t) - v_\lambda(t)| \\ + |v_\lambda(t) - v(t)| + |v(t) - u(t)|$$

et montrer que chacun de ces 5 termes tend vers zéro.

1°)- Par monotonie de  $A^n$  et de  $A$ , on a

$$|u_n(t) - v_n(t)| \leq |u_{0n} - y_n|$$

$$|v(t) - u(t)| \leq |y - u_0|$$

2°)- D'après l'estimation (1-2), on a, pour  $t \in [0, T]$

$$|v_n(t) - v_{n,\lambda}(t)| \leq \sqrt{\lambda T} |(A^n)^0 y_n|$$

$$|v_\lambda(t) - v(t)| \leq \sqrt{\lambda T} |A^0 y|$$

$$\text{D'autre part } y_n + \sqrt{\lambda} A^n y_n \ni P_n u_0 \Rightarrow \sqrt{\lambda} |(A^n)^0 y_n| \leq |y_n - P_n u_0|$$

$$y + \sqrt{\lambda} A y \ni u_0 \Rightarrow \sqrt{\lambda} |A^0 y| \leq |y - u_0|,$$

puisque  $(A^n)^0 y^n$  et  $A^0 y$  sont les éléments de norme minimale de  $A^n y_n$  et  $A y$ , d'où

$$|v_n(t) - v_{n,\lambda}(t)| \leq \sqrt{T} |y_n - P_n u_0|$$

$$|v_\lambda(t) - v(t)| \leq \sqrt{T} |y - u_0|$$

3°)- Pour estimer le 3è terme on écrit

$$v_{n,\lambda}(t) - v_\lambda(t) = y_n - y + \int_0^t \left( \frac{d}{ds} v_{n,\lambda} - \frac{d}{ds} v_\lambda \right) ds$$

d'où

$$\begin{aligned} |v_{n,\lambda}(t) - v_\lambda(t)| &\leq |y_n - y| + \int_0^T |A_\lambda^n v_{n,\lambda} - A_\lambda v_\lambda| ds \\ &\leq a_n + b \int_0^T |v_{n,\lambda}(s) - v_\lambda(s)| ds \end{aligned}$$

$$\text{où } a_n = |y_n - y| + \int_0^T |A_\lambda^n P_n v_\lambda - A_\lambda v_\lambda| ds \quad \text{et } b = \frac{1}{\lambda}$$

(Remarquer que  $|A_\lambda^n v_{n,\lambda} - A_\lambda v_\lambda| \leq |A_\lambda^n v_{n,\lambda} - A_\lambda^n P_n v_\lambda| + |A_\lambda^n P_n v_\lambda - A_\lambda v_\lambda|$  et le 1er terme est majoré  $\frac{1}{\lambda} |v_{n,\lambda} - v_\lambda|$  puisque  $v_{n,\lambda} = P_n v_{n,\lambda}$  et que  $A_\lambda^n$  est lipschitzien de rapport  $\frac{1}{\lambda}$ ).

Le Lemme de Gronwall montre alors que (pour  $0 \leq t \leq T$ )

$$|v_{n,\lambda}(t) - v_\lambda(t)| \leq a_n e^{bT}.$$

Montrons que  $a_n \rightarrow 0$ , lorsque  $n \rightarrow \infty$  et ceci  $\forall \lambda > 0$ .

Soit  $g_n(S) = A_\lambda^n P_n v_\lambda(S)$ . On a

$$g_n(S) = \frac{1}{\lambda} (P_n v_\lambda(S) - (I + \lambda A^n)^{-1} P_n v_\lambda(S)) \rightarrow A_\lambda v_\lambda(S)$$

lorsque  $n \rightarrow \infty$  et ceci pour tout  $S \in [0, T]$ .

D'autre part, soit  $z \in D(A)$  donné

$$\begin{aligned} |g_n(S)| &\leq |A_\lambda^n P_n v_\lambda(S) - A_\lambda^n P_n z| + |A_\lambda^n P_n z| \\ &\leq \frac{1}{\lambda} |v_\lambda(S) - z| + 2|A_\lambda z| \end{aligned}$$

pour  $n \geq n_0$  donné puisque  $A_\lambda^n P_n z \rightarrow A_\lambda z$  lorsque  $n \rightarrow +\infty$ . La fonction  $\frac{1}{\lambda} |v_\lambda(S) - z|$  étant intégrable dans  $L^1(0, T)$ , on en déduit, d'après le théorème de convergence dominée, que

$$\int_0^T |g_n(S) - A_\lambda v_\lambda(S)| dS \rightarrow 0,$$

d'où  $a_n \rightarrow 0$ .

Finalement en passant à la limite pour  $n \rightarrow \infty$  dans (4-3), on trouve

$$\limsup_{n \rightarrow \infty} \|u_n - u\|_{L^\infty(0, T; H)} \leq (2 + 2\sqrt{T}) |y - u_0|$$

qui est aussi petit que l'on veut pourvu que  $\lambda$  soit assez petit. #

Remarque : On peut supposer seulement que  $u_{0n} \rightarrow u_0$  dans le résultat précédent.

## APPLICATIONS

### 1) Méthodes de Galerkin

Si  $V_n \subset D(A)$  on choisit tout simplement  $A_n = P_n A$ . L'hypothèse (4-2) revient alors à établir un résultat de convergence pour le problème stationnaire.

Dans le cas où  $V_n \subset V \subset H \subset V'$  et où  $A$  est la restriction d'un opérateur univoque monotone de  $V$  dans  $V'$ , le problème approché revient à

$$\left(\frac{du_n}{dt}, v_n\right) + \langle Au_n, v_n \rangle = 0 \quad \forall v_n \in V_n$$

et il suffit de montrer que la solution  $y_n$  de

$$(y_n, v_n) + \lambda \langle Ay_n, v_n \rangle = (z, v_n)$$

converge vers celle de  $y + \lambda Ay = z$  lorsque  $n \rightarrow \infty$ .

On n'a jamais vraiment besoin de définir l'opérateur  $A^n$  et ceci s'adapte à de nombreux cas multivoques. Par exemple pour le problème de Mosolov évolutif :

$$\left(\frac{du}{dt}, v-u\right) + a(u, v-u) + j(v) - j(u) - (f, v-u) \geq 0 \quad \forall v \in V$$

que l'on approche par

$$\left(\frac{du_n}{dt}, v_n - u_n\right) + a(u_n, v_n - u_n) + j(v_n) - j(u_n) - (f, v_n - u_n) \geq 0 \quad \forall v_n \in V_n$$

$y_n$  est la solution de

$$(y_n, v_n - u_n) + \lambda a(y_n, v_n - u_n) + \lambda j(v_n) - \lambda j(u_n) - \lambda (f, v_n - u_n) \geq 0 \quad \forall v_n \in V_n$$

## 2) Méthodes de différences finies

$V_n$  est alors un espace de fonctions étagées sur un réseau de différences finies,  $A_n$  peut être dans le cas linéaire la matrice du schéma à 5 points si  $A$  est l'opposé du laplacien ; là encore il suffit de montrer la convergence de la solution du problème stationnaire correspondant et on peut reprendre tous les exemples de la partie A section 5.

N° d'Impression 402

1er trimestre 1980