Lectures on
Numerical Methods For Non-Linear
Variational Problems

By

R. Glowinski

Tata Institute of Fundamental Research
Bombay
1980

Lectures on
Numerical Methods For Non-Linear
Variational Problems

By

R. Glowinski

Notes by

G. Vijayasundaram

Adimurthi

Published for the
Tata Institute of Fundamental Research, Bombay
Springer-Verlag
Berlin Heidelberg New York
1980

Author

R. Glowinski
Université Pierre et Marie Curie
Laboratoire d'Analyse Numérique 189
Tour 55–65 5ém étage
4, Place Jussieu
75230 PARIS CEDEX 05
FRANCE

ISBN 3-540-08774-5 Springer-Verlag, Berlin Heidelberg, New York
ISBN 0-387-08774-5 Springer-Verlag, New York. Heidelberg,. Berlin

Printed by N.S. Ray at The Book Centre Limited Sion East, Bombay 400 022 and Published by H. Goetze Springer-Verlag, Heidelberg, West Germany
Printed in India

To the memory of ***G. Stampacchia***

Preface

These notes correspond to a course of about fifteen lectures given at the Tata Institute of Fundamental Research Centre, Indian Institute of Science, Bangalore in January and February 1977.

The main goal of this course and of the corresponding notes is to provide an introduction to the study of Nonlinear Variational Problems; they do not have pretention to cover all the aspects of this very important subject, since for example the Navier–Stokes equations for newtonian incompressible viscous flows have not been considered here (we refer for this last problem to, e.g., TEMAM [1] and GIRAULT-RAVIART [1]).

Some questions pertinent to the main subject of these notes have not been treated here since they have been considered in the T.I.F.R. Lecture Notes of P.G. CIARLET [1] and J.CEA [2].

Chapters 1 and 2 are concerned with *Elliptic Variational Inequalitites* (E.V.I.) more precisely with their approximation (mostly by finite element methods) and also their iterative solution. Several examples, coming from Mechanics illustrate the methods which are described in these two chapters.

The following Chapter 3 is only an introduction to the approximation of Parabolic Variational Inequalities (P.V.I.); we have however studied with some details a particular P.V.I. related to the unsteady flow of some viscous plastic media (Bingham fluids) in a cylindrical pipe.

In Chapter 4 we show how Variational Inequalities concepts and methods may be useful to study some Nonlinear Variational equations.

In Chapter 5 we discuss the iterative solution of some Variational

Problems with a very specific structure allowing their solution by decomposition - coordination methods via augmented lagrangians; several iterative methods are described and illustrated by examples, mostly from Mechanics.

In Chapter 6, which unlike the previous chapters is largely heuristical, we show how some of the tools of the Chapters I–IV may be used to solve numerically a difficult and important nonlinear problem of Fluid Dynamics: namely the steady transonic potential flow of an inviscid compressible fluid. This last chapter is obviously just an introduction to this very important and difficult subject.

I would like to thank all the people who make my stay in India a most enjoyable experience and more particularly Professors K.G. RAMANATHAN, K. BALAGANGADHARAN and M.K.V. MURTHY.

These Notes were taken by M. ADIMURTHI and M.G. VIJAYASUNDARAM; I would like to thank them for their devoted efforts.

I would like to thank also S. KESAVAN and L. REINHART for their careful reading of the proofs and the various improvements they have suggested. Eventually I would like to express all my acknowledgements to Mrs. F. WEBER for her beautiful typing of these Notes and to Mr. M. Bazot who did all the artwork.

R. Glowinski
Rocquencourt, France
November, 1979

Contents

Chapter 1

Generalities On Elliptic Variational Inequalities And On Their Approximation

1 Introduction

An important and very useful class of non-linear problems arising from mechanics, physics etc. consists of the so-called Variational Inequalities. We mainly consider the following two types of variational inequalities, namely **1**

1. Elliptic Variational Inequalities (EVI),
2. Parabolic Variational Inequalities (PVI).

In this chapter (following LIONS-STAMPACCHIA 1) we shall restrict our attention to the study of the existence, uniqueness and approximation of the solutions of EVI.

2 Functional Context

In this section we consider two classes of EVI, namely EVI of the first kind and EVI of the second kind.

2.1 Notations

- V : real Hilbert space with scalar product $(\cdot,\cdot)$ and associated norm $\|\cdot\|$.
- V^* : the dual space of V.
- $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ is a bilinear, continuous and V *-elliptic* form on $V \times V$.

A bilinear form $a(\cdot,\cdot)$ is said to be V *-elliptic* if there exists a positive constant α such that $a(v,v) \geq \alpha \parallel v \parallel^2 \forall v \in V$.

In general we do not assume $a(\cdot,\cdot)$ to be symmetric, since in some applications non-symmetric bilinear forms may occur naturally (see for instance COMINCIOLI [1]).

- $L : V \to \mathbb{R}$ continuous, linear functional.
- K is a closed, convex, non-empty subset of V.
- $j(\cdot) : V \to \bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ is a convex, lower semi -continuous (l.s.c.) and proper functional

($j(\cdot)$ is proper if $j(v) > -\infty \, \forall v \in V$ and $j \not\equiv \infty$).

2.2 EVI of first kind

2 *To find $u \in V$ such that u is a solution of the problem*

$$(P_1)\begin{cases} a(u, v-u) \geq L(v-u), \forall v \in K, \\ u \in K. \end{cases}$$

2.3 EVI of second kind

To find $u \in V$ such that u is a solution of the problem

$$(P_2)\begin{cases} a(u, v-u) + j(v) - j(u) \geq L(v-u) \forall v \in V, \\ u \in V. \end{cases}$$

2.4 Remarks:

REMARK 2.1. *The cases above considered are the simplest and most important. LIONS and BENSOUSSAN [1] considered some generalization of problem* (P_1) *called* Quasi Variational Inequalities *(QVI) which arises for instance from Decision Sciences. A typical problem of QVI is :*

To find $u \in V$ such that

$$\begin{cases} a(u, v-u) \geq L(v-u) \; \forall v \in K(u), \\ u \in K(u) \end{cases}$$

where $v \to K(v)$ *is a family of closed, convex non-empty subsets of* V.

REMARK 2.2. *If* $K = V$ *and* $j \equiv 0$ *then the problems* (P_1) *and* (P_2) *reduce to the classical variational equation*

$$\begin{cases} a(u, v) = L(v) \; \forall v \in V, \\ \\ u \in V. \end{cases}$$

REMARK 2.3. *The distinction between* (P_1) *and* (P_2) *is artificial, for* (P_1) *can be considered as a particular case of* (P_2) *by replacing* $j(\cdot)$ *in* (P_2) *by the indicator function* I_K *of* K *defined by* 3

$$I_K(v) = \begin{cases} 0 \text{ if } v \in K \\ +\infty \text{ if } v \notin K. \end{cases}$$

Even though (P_1) is a particular case of (P_2) it is worthwhile considering (P_1) separately because it arises in a natural way and we will get geometrical insight into the problem.

Exercise 2.1. *Prove that* I_K *is a convex, l.s.c. and proper functional.*

Exercise 2.2. *Show that* (P_1) *is equivalent to the problem of finding* $u \in V$ *such that* $a(u, v-u) + I_K(v) - I_K(u) \geq L(v-u) \; \forall v \in V$.

3 Existence And Uniqueness Results For EVI of First Kind

3.1 A Theorem of existence and uniqueness

THEOREM 3.1. *(LIONS-STAMPACCHIA 1).* The problem (P_1) has one and only one solution

Proof. (I)Uniqueness:

Let u_1 and u_2 be solutions of (P_1). We have then

$$a(u_1, v - u_1) \geq L(v - u_1) \ \forall v \in K, u_1 \in K, \tag{3.1}$$

$$a(u_2, v - u_2) \geq L(v - u_2) \ \forall v \in K, u_2 \in K. \tag{3.2}$$

□

Putting u_2 for v in (3.1) and u_1 for v in (3.2) and adding we get, by using the V -ellipticity of $a(\cdot, \cdot)$,

$$\alpha \parallel u_2 - u_1 \parallel^2 \leq a(u_2 - u_1, u_2 - u_1) \leq 0$$

which proves $u_1 = u_2$ since $\alpha > 0$.

(2) Existence

4 We use a generalization of the proof used by CLARLET [1] for proving the Lax-Milgram Lemma, i. e. we will reduce the problem (P_1) to a *fixed point* problem.

By the Riesz representation theorem for Hilbert space there exist $A \in \mathscr{L}(V, V)(A = A^t$ if $a(\cdot, \cdot)$ is symmetric) and $\ell \in V$ such that

$$(Au, v) = a(u, v) \ \ \forall u, v \in V \text{ and } L(v) = (\ell, v) \ \ \forall v \in V. \tag{3.3}$$

Then the problem (P_1) is equivalent to finding $u \in V$ such that

$$\begin{cases} (u - \rho(Au - \ell) - u, v - u) \leq 0 \ \ \forall v \in K, \\ \\ u \in K, \rho > 0. \end{cases} \tag{3.4}$$

This is equivalent to finding u such that

$$u = P_K(u - \rho(Au - \ell)), \text{ for some } \rho > 0, \tag{3.5}$$

where P_K denotes the projection operator from V to K in the $\| \cdot \|$ norm. Consider the map $W_\rho : V \to V$ defined by

$$W_\rho(v) = P_K(v - \rho(Av - \ell)). \tag{3.6}$$

Let $v_1, v_2 \in V$. Then since P_K is a contraction we have

$$\begin{cases} \| W_\rho(v_1) - W_\rho(v_2) \|^2 \leq \| v_2 - v_1 \|^2 + \rho^2 \| A(v_2 - v_1 \|^2 \\ \\ -2\rho a(v_2 - v_1, v_2 - v_1). \end{cases}$$

Hence we have

$$\| W_\rho(v_1) - W_\rho(v_2) \|^2 \leq (1 - 2\rho\alpha + \rho^2 \| A \|^2) \| v_2 - v_1 \|^2 . \tag{3.7}$$

Thus W_ρ is a strict contraction mapping if $0 < \rho < \dfrac{2\alpha}{\| A \|^2}$. By taking ρ in this range we have a unique solution for the fixed point problem which implies the existence of a solution for (P_1).

3.2 Remarks

REMARK 3.1. *If $K = V$, Theorem 3.1 reduces to Lax-Milgram Lemma (see CIARLET [1]).* 5

REMARK 3.2. *If $a(\cdot, \cdot)$ is symmetric then Theorem 3.1 can be proved using optimization methods (see CEA [1]).*

Let $J : V \to \mathbb{R}$ be defined by

$$J(v) = \frac{1}{2} a(v, v) - L(v). \tag{3.8}$$

Then

(i) $\lim_{\|v\|\to+\infty} J(v) = +\infty$

since $J(v) = \frac{1}{2}a(v,v) - L(v) \geq \frac{\alpha}{2} \| v \|^2 - \| L \| \| v \|$.

(ii) *J is strictly convex.*

Since L is linear, to prove the strict convexity of J it suffices to prove that

$$v \to a(v,v)$$

is strictly convex. Let $0 < t < 1$ and $u, v \in V$ with $u \neq v$; $0 < a(v-u, v-u) = a(u,u) + a(v,v) - 2a(u,v)$. Hence we have

$$2a(u,v) < a(u,u) + a(v,v). \tag{3.9}$$

Using (3.9) we have

$$\begin{cases} a(tu + (1-t)v, tu + (1-t)v) = \\ \qquad t^2 a(u,u) + 2t(1-t)a(u,v) + (1-t)^2 a(v,v) < \\ \\ < ta(u,u) + (1-t)a(v,v). \end{cases} \tag{3.10}$$

Therefore $a(v,v)$ is strictly convex.

6 (iii) Since $a(\cdot,\cdot)$ and L are continuous, J is continuous.

From these properties of J and standard results of Optimization Theory (cf. CEA [1]) it follows that the minimization problem of finding u such that

$$(\pi) \begin{cases} J(u) \leq J(v) \quad \forall v \in K, \\ \\ u \in K \end{cases}$$

has one and only one solution. Therefore (π) is equivalent to the problem of finding u such that

$$\begin{cases} (J'(u), v-u) \geq 0 \quad \forall v \in K, \\ \\ u \in K, \end{cases} \tag{3.11}$$

where $J'(u)$ is the *Gateaux derivative* of J at u. Since $(J'(u), v) = a(u, v) - L(v)$ we see that (P_1) and (π) are equivalent if $a(\cdot, \cdot)$ is symmetric.

Exercise 3.1. *Prove that* $(J'(u), v) = a(u, v) - L(v) \ \forall u, v \in V$ *and hence deduce that* $J'(u) = Au \sim \ell \ \forall u \in V$.

REMARK 3.3. *The proof of Theorem 3.1 given a natural a natural algorithm for solving* (P_1) *since* $v \to P_K(v - \rho(Av - \ell))$ *is a* contraction *mapping for* $0 < \rho < \dfrac{2\alpha}{\| A \|^2}$. *Hence we can use the following algorithm to find* u*:*

$$\text{Let } u^0 \in V, \tag{3.12}$$

$$u^{n+1} = P_K(u^n - \rho(Au^n - \ell)). \tag{3.13}$$

Then $u^n \to u$ strongly in V where u is the solution of (P_1). In practice it is not easy to calculate ℓ and A unless $V = V^*$. To project over K may be as difficult as solving (P_1). In general this method cannot be used for computing the solution of (P_1) if $K \neq V$ (at least not so directly).

We observe that if $a(\cdot, \cdot)$ is symmetric then $J'(u) = Au - \ell$ and hence 7
(3.13) becomes

$$u^{n+1} = P_K(u^n - \rho(J'(u^n)). \tag{3.13'}$$

This method is know as the *Gradient -Projection* method.

4 Existence And Uniqueness Results For EVI of Second Kind

THEOREM 4.1. *(LIONS-STAMPACCHIA [1]) Problem* (P_2) *has one only one solution.*

Proof. As in Theorem 3.1 we shall first prove uniqueness and then existence

(1) Uniqueness. Let u_1 and u_2 be two solutions of (P_2). Then we have

$$a(u_1, v - u_1) + j(v) - j(u_1) \geq L(v - u_1) \ \forall v \in V, u_1 \in V, \tag{4.1}$$

$$a(u_2, v - u_2) + j(v) - j(u_2) \geq L(v - u_2) \ \forall v \in V, u_2 \in V, \tag{4.2}$$

□

Since $j(\cdot)$ is a proper map there exists $v_0 \in V$ such that $-\infty < j(v_0) < \infty$. Hence for $i = 1, 2$

$$-\infty < j(u_i) \leq j(v_0) - L(v_0 - u_i) + a(u_i, v_0 - u_i). \tag{4.3}$$

This shows that $j(u_i)$ is finite for $i = 1, 2$. Hence by substituting u_2 for v in (4.1) and u_1 for v in (4.2) and adding we obtain

$$\alpha \parallel u_1 - u_2 \parallel^2 \leq a(u_1 - u_2, u_1 - u_2) \leq 0. \tag{4.4}$$

Hence $u_1 = u_2$.

(2) Existence. For each $u \in V$ and $\rho > 0$ we associate a problem (π_ρ^u) of type (P_2) defined as below :

To find $w \in V$ such that

$$(\pi_\rho^u) \begin{cases} (w, v - w) + \rho j(v) - \rho j(w) \\ \qquad \geq (u, v - w) + \rho L(v - w) - \rho a(u, v - w) \ \ \forall v \in V, \\ w \in V. \end{cases} \tag{4.5}$$

8 The advantage of considering this problem over the problem (P_2) is that the bilinear form associated with (π_ρ^u) is the inner product of V which is symmetric.

Let us first assume that (π_ρ^u) has a unique solution for all $u \in V$ and $\rho > 0$. For each ρ define the map $f_\rho : V \to V$ by $f_\rho(u) = w$ where w is the unique solution of (π_ρ^u).

We shall show that f_ρ is a uniformly strict contraction mapping for suitable chosen ρ.

Let $u_1, u_2 \in V$ and $w_i = f_\rho(u_i)$, $i = 1, 2$. Since $j(\cdot)$ is proper we have $j(u_i)$ finite which can be proved as in (4.3). Therefore we have

$$\begin{aligned}(w_1, w_2 - w_1) + \rho j(w_2) - \rho j(w_1) \\ \geq (u_1, w_2 - w_1) + \rho L(w_2 - w_1) - \rho a(u_1, w_2 - w_1),\end{aligned} \tag{4.6}$$

$$(w_2, w_1 - w_2) + \rho j(w_1) - \rho j(w_2) \geq (u_2, w_1 - w_2) + \rho L(w_1 - w_2) - \rho a(u_2, w_1 - w_2). \quad (4.7)$$

Adding these inequalities we obtain

$$\begin{cases} \| f_\rho(u_1) - f_\rho(u_2) \|^2 = \| w_2 - w_1 \|^2 \\ \qquad = \leq ((I - \rho A)(u_2 - u_1), w_2 - w_1) \\ \qquad = \leq \| I - \rho A \| \ \| u_2 - u_1 \| \ \| w_2 - w_1 \| . \end{cases} \quad (4.8)$$

Hence

$$\| f_\rho(u_1) - f_\rho(u_2) \| \leq \| I - \rho A \| \ \| u_2 - u_1 \|$$

It is easy to show that $\| I - \rho A \| < 1$ when $0 < \rho < \dfrac{2\alpha}{\| A \|^2}$. This proves that f_ρ is uniformly a strict contracting mapping and hence has a unique fixed point u. This u turns out to be the solution of (P_2) since $f_\rho(u) = u$ implies $(u, v - u) + \rho j(v) - \rho j(u) \geq (u, v - u) + \rho L(v - u) - \rho a(u, v - u)$ $\forall v \in V$. Therefore

$$a(u, v - u) + j(v) - j(u) \geq L(v - u) \ \ \forall v \in V. \quad (4.9)$$

Hence (P_2) has a unique solution.

The existence and uniqueness of the problem (π^u_ρ) follows from the following

Lemma 4.1. *Let $b : V \times V \to \mathbb{R}$ be a symmetric continuous, bilinear, V* **9**
-elliptic form with V -elliptic constant β. Let $L \in V^*$ and $j : V \to \overline{\mathbb{R}}$ be a convex, l.s.c. proper functional. Let $J(v) = \frac{1}{2} b(v, v) + j(v) - L(v)$. Then the minimization problem (π):

To find u such that

$$(\pi) \begin{cases} J(u) \leq J(v) \ \forall v \in V, \\ u \in V \end{cases}$$

has a unique solution which is characterised by

$$\begin{cases} b(u, v-u) + j(v) - j(u) \geq L(v-u) \quad \forall v \in V, \\ \\ u \in V. \end{cases} \tag{4.10}$$

Proof. (i) Existence and uniqueness of u

Since $b(\cdot,\cdot)$ is strictly convex, j is convex and L is linear, we have J strictly convex. J is l.s.c. because $b(\cdot,\cdot)$ and L are continuous and j is l.s.c. □

Since j is convex, l.s.c. and proper, there exists $\lambda \in V^*$ and $\mu \in \mathbb{R}$ such that

$$j(v) \geq \lambda(v) + \mu \text{ (cf. EKLAND - TEMAM [1]),}$$

therefore

$$\begin{cases} J(v) \quad \geq \frac{\beta}{2} \| v \|^2 - \| \lambda \| \| v \| - \| L \| \quad \| v \| + \mu \\ \\ \qquad = \left(\sqrt{\frac{\beta}{2}} \| v \| - \frac{(\|\lambda\| + \|L\|)}{2} \sqrt{\frac{2}{\beta}} \right)^2 + \mu - \frac{(\|\lambda\| + \|L\|)^2}{2\beta}. \end{cases} \tag{4.11}$$

Hence

$$J(v) \to +\infty \text{ as } \| v \| \to +\infty. \tag{4.12}$$

10 Hence (cf. CEA [1]) there exists a unique solution for the optimization problem (π).

Characterisation of u : We show that the problem (π) is equivalent to (4.10) and thus get a characterisation of u.

(2) Necessity of (4.10) : Let $0 < t \leq 1$. Let u be the solution of (π). Then for all $v \in V$ we have

$$J(u) \leq J(u + t(v-u)). \tag{4.13}$$

Set $J_0(V) = \frac{1}{2} b(v, v) - L(v)$, then (4.13) becomes

$$\begin{cases} 0 \quad \leq J_0(u + t(v-u)) - J_0(u) + j(u + t(v-u)) - j(u) \\ \\ \qquad \leq J_0(u + t(v-u)) - J_0(u) + t[j(v) - j(u)] \quad \forall v \in V \end{cases} \tag{4.14}$$

got by using convexity of j. Dividing by t in (4.14) and taking the limit as $t \to 0$ we get

$$0 \leq (J_0'(u), v-u) + j(v) - j(u) \quad \forall v \in V. \tag{4.15}$$

Since $b(\cdot,\cdot)$ is symmetric we have

$$(J_0'(v), w) = b(v,w) - L(w) \quad \forall v, w \in V. \tag{4.16}$$

From (4.15) and 4.16 we obtain

$$b(u, v-u) + j(v) - j(u) \geq L(v-u) \quad \forall v \in V.$$

This proves the necessity.

(3) Sufficiency of (4.10). Let u be a solution of (4.10) ; for $v \in V$

$$J(v) - J(u) = \frac{1}{2}[b(v,v) - b(u,u)] + j(v) - j(u) - L(v-u). \tag{4.17}$$

But

$$b(v,v) = b(u+v-u, u+v-u)$$

$$= b(u,u) + 2b(u, v-u) + b(u-v, u-v).$$

Therefore **11**

$$J(v) - J(u) = b(u, v-u) + j(v) - j(u) - L(v-u) + \frac{1}{2}b(v-u, v-u). \tag{4.18}$$

Since u is a solution of (4.10) and $b(v-u, v-u) \geq 0$ we get

$$J(v) - J(u) \geq 0. \tag{4.19}$$

Hence u is a solution of (π).

By taking $b(\cdot,\cdot)$ to be the inner product in V and replacing $j(v)$ and $L(v)$ in Lemma 4.1 by $\rho j(v)$ and $(u,v) + \rho L(v) - \rho a(u,v)$, respectively, we get the solution for (π_ρ^u).

REMARK 4.1. *From the proof of Theorem 4.1 we get an algorithm for solving* (P_2)*. This algorithm is given by*

$$\begin{cases} (1) & u^0 \in V, 0 < \rho < \frac{2\alpha}{\|A\|}^2, \\ (2) & (u^{n+1}, v - u^{n+1}) + \rho j(v) - \rho j(u^{n+1}) \geq (u^n, v - u^{n+1}) \\ & \qquad +\rho L(v - u^{n+1}) - \rho a(u^n, v - u^{n+1}) \ \ \forall v \in V, \\ (3) & u^{n+1} \in V. \end{cases} \tag{4.20}$$

Then one can easily see that $u_n \to u$ *strongly* in V and u will be the solution of (P_2). Difficulties may arise in using this scheme when $j(\cdot)$ is not assumed to be differentiable. At each iteration the problem we have to solve is also a problem of the same order of difficulty as that of the original problem (actually conditioning can be better provided ρ has been conveniently chosen). If $a(\cdot,\cdot)$ is not symmetric the fact that $(\cdot,\cdot)$ is symmetric can also give some simplification.

5 -Internal Approximation of EVI of First Kind

5.1 Introduction

In this chapter we shall study the approximation of EVI of the first kind from an abstract, axiomatic point of view.

5.2 The continuous problem

12 The assumptions on V, K, L and $a(\cdot,\cdot)$ are as in section 2. We are interested in the approximation of

$$(P_1) \begin{cases} a(u, v - u) \geq L(v - u) \ \ \forall v \in K, \\ \\ u \in K, \end{cases}$$

which has one and only solution by Theorem 3.1.

5.3 The approximate problem

5.3.1 The approximation of V and K

We are given a parameter h converging to 0 and a family $(V_h)_h$ of closed subspaces of V. (In practice V_h are finite dimensional and the parameter h varies over a sequence). We are also given a family $(K_h)_h$ of closed, convex, non-empty subsets of V with $K_h \subset V_h \ \ \forall h$ (in general we do not assume $K_h \subset K$) such that $(K_h)_h$ satisfies the following two conditions :

(i) If $(v_h)_h$ is such that $v_h \in K_h \ \ \forall h$ and $(v_h)_H$ is bounded in V then the *weak* cluster points of $(v_h)_h$ belong to K.

(ii) Assume there exist $\chi \subset V$, $\overline{\chi} = K$ and $r_h : \chi \to K_h$ such that $\lim_{h\to 0} r_h v = v$ *strongly in* V, $\forall v \in \chi$.

REMARK 5.1. *If $K_h \subset K \ \forall h$ then (i) is trivially satisfied because K is weakly closed.*

REMARK 5.2. $\cap_h K_h \subset K$.

REMARK 5.3. *A useful variant of condition (ii) for r_h is (ii)' Assume there exists a subset $\chi \subset V$ such that $\overline{\chi} = K$ and $r_h : \chi \to V_h$ with the property that for each $v \in \chi$, there exists $h_0 = h_0(v)$ with $r_h v \in K_h$ for all $h \le h_0(v)$ and $\lim_{h\to 0} r_h v = v$* strongly *in V.*

5.3.2 Approximation of (P_1):

The problem (P_1) is approximated by

$$(P_{1h}) \begin{cases} a(u_h, v_h - u_h) \ge L(v_h - u_h) \ \ \forall v_h \in K_h, \\ \\ u_h \in K_h. \end{cases}$$

THEOREM 5.1. (P_{1h}) has a unique solution. **13**

Proof. In Theorem 3.1 taking V to be V_h and K to be K_h we have the result. □

REMARK 5.4. *In most of the cases it will be necessary to replace $a(\cdot,\cdot)$ and L by $a_h(.,.)$ and L_h (usually defined - in practical cases - from $a(\cdot,\cdot)$ and L by a Numerical Integration procedure). Since there is nothing very new on that matter compared to the classical linear case, we shall say nothing about this problem for which we refer to CIARLET [1, Chap. 8].*

5.4 Convergence results

THEOREM 5.2. *With the above assumptions on K and $(K_h)_h$ we have* $\lim_{h\to 0} \| u_h - u \|_V = 0$ *with u_h the solution of (P_{1h}) and u the solution of (P_1).*

Proof. In this kind of convergence we usually divide the proof into three parts. First we obtain a priori estimates for $(u_h)_h$, then weak convergence of $(u_h)_h$ and finally with the help of weak convergence, we will prove strong convergence. □

(1) Estimation for u_h.

We will now show that there exist constants C_1 and C_2 independent of h such that

$$\| u_h \|^2 \leq C_1 \| u_h \| + C_2, \forall h. \tag{5.1}$$

Since u_h is the solution of (P_{1h}) we have

$$a(u_h, v_h - u_h) \geq L(v_h - u_h) \forall v_h \in K_h \tag{5.2}$$

i.e.

$$a(u_h, u_h) \leq a(u_h, v_h) - L(v_h - u_h).$$

By V -ellipticity we get

$$\alpha \| u_h \|^2 \leq \| A \| \cdot \| u_h \| \cdot \| v_h \| + \| L \| (\| v_h \| + \| u - h \|) \ \forall v_h \in K_h. \tag{5.3}$$

Let $v_0 \in \chi$ and $v_h = r_h v_0 \in K_h$. By condition (ii) on K_h we have $r_h v_0 \to v_0$ strongly in V and hence $\| v_h \|$ is uniformly bounded by a constant m. Hence (5.3) can be written as

$$\| u_h \|^2 \leq \frac{1}{\alpha}\{(m \| A \| + \| L \|) \| u_h \| + \| L \| m\} = C_1 \| u_h \| + C_2,$$

where $C_1 = \frac{1}{\alpha}(m \parallel A \parallel + \parallel L \parallel)$ and $C_2 = \frac{m}{\alpha} \parallel L \parallel$; then (5.1) implies **14**
$\parallel u_h \parallel \leq C \ \forall h.$

(2) Weak convergence of $(u_h)_h$ **:** Relation (5.1) gives u_h is uniformly bounded. Hence there exists a subsequence say $\{u_{h_i}\}$ such that u_{h_i} converges to u^* weakly in V. By condition (i) on $(K_h)_h$ we have $u^* \in K$. We will prove that u^* is a solution for (P_1). We have

$$a(u_{h_i}, u_{h_i}) \leq a(u_{h_i}, v_{h_i}) - L(v_{h_i}, u_{h_i}) \ \ \forall v_{h_i} \in K_{h_i}. \tag{5.4}$$

Let $v \in \chi$ and $v_{h_i} = r_{h_i} v$. Then (5.4) becomes

$$a(u_{h_i}, u_{h_i}) \leq a(u_{h_i}, r_{h_i} v) - L(r_{h_i} v - u_{h_i}). \tag{5.5}$$

Since $r_{h_i} v$ converges strongly to v and u_{h_i} converges to u^* weakly as $h_i \to 0$ taking the limit in (5.5) we get

$$\liminf_{h_i \to 0} a(u_{h_i}, u_{h_i}) \leq a(u^*, v) - L(v - u^*) \ \ \forall v \in \chi. \tag{5.6}$$

Also we have

$$0 \leq a(u_{h_i} - u^*, u_{h_i} - u^*) \leq a(u_{h_i}, u_{h_i}) - a(u_{h_i}, u^*) - a(u^*, u_{h_i}) + a(u^*, u^*)$$

i. e.

$$a(u_{h_i}, u^*) + a(u^*, u_{h_i}) - a(u^*, u^*) \leq a(u_{h_i}, u_{h_i}).$$

By taking the limit we obtain

$$a(u^*, u^*) \leq \liminf_{h_i \to 0} a(u_{h_i}, u_{h_i}). \tag{5.7}$$

From (5.6) and (5.7) we get

$$a(u^*, u^*) \leq \liminf_{h_i \to 0} a(u_{h_i}, u_{h_i}) \leq a(u^*, v) - L(v - u^*) \ \forall v \in \chi.$$

Therefore we have, **15**

$$\begin{cases} a(u^*, v - u^*) \geq L(v - u^*) \ \ \forall v \in \chi, \\ \\ u^* \in K. \end{cases} \tag{5.8}$$

Since χ is dense in K and $a(\cdot,\cdot)$, L are continuous, we get from (5.8)

$$\begin{cases} a(u^*, v-u^*) \geq L(v-u^*) \ \ \forall v \in K, \\ u^* \in K. \end{cases} \tag{5.9}$$

Hence u^* is a solution of (P_1). By Theorem 3.1, the solution for (P_1) is unique and hence $u^* = u$ is the unique solution. Hence u is the only cluster point of $\{u_h\}_h$ in the weak topology of V. Hence the whole $\{u_h\}_h$ converges to u weakly.

(3) Strong convergence: We have by $V-$ellipticity of $a(\cdot,\cdot)$

$$0 \leq \alpha \parallel u_h - u \parallel^2 \leq a(u_h - u, u_h - u) = a(u_h, u_h) - a(u_h, u) - a(u, u_h) + a(u, u) \tag{5.10}$$

where u_h is the solution of (P_{1h}) and u is the solution of (P_1). Since u_h is the solution of (P_{1h}) and $r_h v \in K_h$ for any $v \in \chi$, we get by (P_{1h})

$$a(u_h, u_h) \leq a(u_h, r_h v) - L(r_h v - u_h) \quad \forall v \in \chi. \tag{5.11}$$

Since $\lim_{h\to 0} u_h = u$ *weakly* in V and $\lim_{h\to 0} r_h v = v$ *strongly* in V (by condition (ii)) we obtain, from (5.10), (5.11) and after taking the lim, that $\forall v \in \chi$ we have:

$$0 \leq \alpha \liminf \parallel u_h - u \parallel^2 \leq \alpha \limsup \parallel u_h - u \parallel^2 \leq a(u, v-u) - L(v-u). \tag{5.12}$$

By *density* and *continuity*, (5.12) also holds $\forall v \in K$; then taking $v = u$ in (5.12) we obtain that

$$\lim_{h\to 0} \parallel u_h - u \parallel^2 = 0$$

i.e. the strong convergence.

16 **REMARK 5.5.** *Error estimates for the EVI of the first kind can be found in FALK [1], [2], [3], STRANG-MOSCO [1], STRANG [1], GLOWINSKI-LIONS-TREMOLIERES (G.L.T.) [1], [2], CIARLET [1], BREZZI [1], FALK-MERCIER [1], GLOWINSKI [1]. But like in many nonlinear problems the methods used to obtain these estimates are specific to the*

particular problem under consideration (as we shall see in the following sections).

This remark still holds for the approximation of EVI of the second kind which is the subject of Section 6.

REMARK 5.6. *If for a given problem, several approximations are available, and if computations are needed, the choice of the approximations to be used is not obvious. We have to take into account not only the convergence properties of the method, but also the computation involved in that method. Some iterative methods are best suited only for some problems. Some methods are easier to program than others.*

6 Internal Approximation of EVI of Second Kind

6.1 The Continuous Problem

The assumptions on V, $a(\cdot,\cdot)$, L $j(\cdot)$ being as in Section 2.1, we shall consider the approximation of

$$(P_2)\begin{cases} a(u, v-u) + j(v) - j(u) \geq L(v-u) \quad \forall v \in V, \\ \\ u \in V \end{cases}$$

which has one and only one solution by Theorem 4.1.

6.2 Definition of the approximate problem

Preliminary remark: We assume in the sequel that $j : V \to \mathbb{R}$ is continuous. We can prove the same sort of results as in this section under less restrictions (see Chapter 4, Section 2).

6.2.1 Approximation of V

Given a real parameter h converging to 0 and a family $(V_h)_h$ of closed subspaces of V (in practice we will take V_h to be finite dimensional and h to vary over a sequence), we assume that $(V_h)_h$ satisfies

(i) there exists $U \subset V$ such that $\overline{U} = V$ and for each h, a map $r_h : U \to V_h$ such that $\lim_{h\to 0} r_h v = v$ strongly in V, $\forall v \in U$.

6.2.2 Approximation of $j(\cdot)$

We approximate the functional $j(\cdot)$ by $(j_h)_h$ where for each h, j_h satisfies

$$\begin{cases} j_h : V_h \to \overline{\mathbb{R}} \\ \\ j_h \textit{is convex, l.s.c. and uniformly proper in } h. \end{cases} \tag{6.1}$$

The family $(j_h)_h$ is said to be *uniformly proper in h* if there exist $\lambda \in V^*$ and $\mu \in \mathbb{R}$ such that

$$h(v_h) \geq \lambda(v_h) + \mu \quad \forall v_h \in V_h, \forall h. \tag{6.2}$$

Furthermore we assume that $(j_h)_h$ satisfies

(ii) if $v_h \to v$ *weakly* in V then

$$\liminf_{h\to 0} j_h(v_h) \geq j(v)$$

(iii) $\lim_{h\to 0} j_h(r_h v) = j(v) \quad \forall v \in U$.

REMARK 6.1. *In all the applications we know, if $j(\cdot)$ is a continuous functional then it is always possible to construct continuous j_h satisfying (ii) and (iii).*

REMARK 6.2. *In some cases we are fortunate enough to have $j_h(v_h) = j(v_h) \forall v_h$, $\forall h$, and then (ii) and (iii) are trivially satisfied.*

6.2.3 Approximation of (P_2)

18 We approximate (P_2) by

$$(P_{2h}) \begin{cases} a(u_h, v_h - u_h) + j_h(v_h) - j_h(u_h) \geq L(v_h - u_h) \quad \forall v_h \in V_h, \\ \\ u_h \in V_h. \end{cases}$$

THEOREM 6.1. Problem (P_{2h}) has one only one solution.

Proof. In Theorem 4.1 taking V to be V_h, $j(\cdot)$ to be $j_h(\cdot)$ we get the result. □

REMARK 6.3. *Remark 5.4 of Section 5 still holds for* (P_2) *and* (P_{2h}).

6.3 Convergence results

THEOREM 6.2. *Under the above assumptions on* $(V_h)_h$ *and* $(j_h)_h$ *we have*

$$\begin{cases} \lim_{h\to 0} \| u_h - u \| = 0, \\ \lim_{h\to 0} j_h(u_h) = j(u). \end{cases} \tag{6.3}$$

Proof. As in the proof of Theorem 5.2 we divide the proof into three parts. □

(1) Estimation for u_h. We will show that there exist positive constants C_1 and C_2 independent of h such that

$$\| u_h \|^2 \leq C_1 \| u_h \| + C_2 \quad \forall h. \tag{6.4}$$

Since u_h is the solution of (P_{2h}) we have

$$a(u_h, u_h) + j_h(u_h) \leq a(u_h, v_h) + j_h(v_h) - L(v_h - u_h) \ \ \forall v_h \in V_h. \tag{6.5}$$

By using relation (6.2) we get

$$\alpha \| u_h \|^2 \leq \| \lambda \| \ \| u_h \| + |\mu| + \| A \| \ \| u_h \| \ \| v_h \| + |j_h(v_h)| + \| L \| (\| v_h \| + \| u_h \|). \tag{6.6}$$

Let $v_0 \in U$ and $v_h = r_h v_0$. By using condition (i) and (iii) there exists a constant m, independent of h such that $\| v_h \| \leq m$ and $|j_h(v_h)| \leq m$. Therefore (6.6) can be written as

$$\begin{cases} \| u_h \|^2 & \leq \frac{1}{\alpha}(\| \lambda \| + \| A \| \cdot m + \| L \|) \| u_h \| + \frac{m}{\alpha}(1 + \| L \|) + \frac{|\mu|}{\alpha} \\ & = C_1 \| u_h \| + C_2 \end{cases}$$

where 19

$$C_1 = \frac{1}{\alpha}(\| \lambda \| + \| A \| \cdot m + \| L \|) \text{ and } C_2 = \frac{m}{\alpha}(1 + \| L \|) + \frac{|\mu|}{\alpha}$$

and (6.4) implies

$$\| u_h \| \leq C \ \forall h \text{ where } C \text{ is a constant.}$$

(2) Weak convergence of u_h: Relation (6.4) gives that u_h is uniformly bounded. Therefore there exists a subsequence $(u_{h_i})_{h_i}$ such that $u_{h_i} \to u_h$ weakly in V.

Since u_h is the solution of (P_{1h}) and $r_h v \in V_h \ \forall h$ and $\forall v \in U$ we get

$$a(u_{h_i}, u_{h_i}) + j_{h_i}(u_{h_i}) \leq a(u_{h_i}, r_{h_i} v) + j_{h_i}(r_h v_i) - L(r_{h_i} v - u_{h_i}). \quad (6.7)$$

By condition (iii) and weak convergence of $\{u_{h_i}\}$ we get

$$\liminf_{h \to 0} [a(u_{h_i}, u_{h_i}) + j_{h_i}(u_{h_i})] \leq a(u^*, v) + j(v) - L(v - u^*) \ \forall v \in U. \quad (6.8)$$

As in (5.7) and using condition (ii), we get

$$a(u^*, u^*) + j(u^*) \leq \liminf_{h \to 0} [a(u_{h_i}, u_{h_i}) + j_{h_i}(u_{h_i})]. \quad (6.9)$$

From (6.8), (6.9) and using the density of U we have

$$\begin{cases} a(u^*, v - u^*) + j(v) - j(u^*) \geq L(v - u^*) \ \forall v \in V, \\ u^* \in V. \end{cases}$$

This implies u^* is a solution of (P_2). Hence $u^* = u$ is the unique solution of (P_2) and this shows that (u_h) converges to u *weakly*.

20 **(3) Strong convergence of $(u_h)_h$:** We have by V -ellipticity of $a(\cdot, \cdot)$ and by (P_{2h})

$$\begin{cases} \alpha \| u_h - u \|^2 & + j_h(u_h) \leq a(u_h - u, u_h - u) + j_h(u_h) = \\ & = a(u_h, u_h) - a(u, u_h) - a(u_h, u) + a(u, u) + j_h(u_h) \leq \\ & \leq a(u_h, r - hv) + j_h(r_h v) - L(r_h v - u_h) - a(u, u_h) \\ & -a(u_h, u) + a(u, u) \ \forall v \in U. \end{cases} \quad (6.10)$$

The right hand side of inequality (6.10) tends to $a(u, v-u) + j(v) - L(v-u)$ as $h \to 0 \forall v \in U$. Therefore we have

$$\begin{cases} \liminf_{h\to 0} j_h(u_h) & \leq \liminf_{h\to 0}[\alpha \parallel u_h - u \parallel^2 + j_h(u_h)] \leq \\ & \leq \limsup_{h\to 0}[\alpha \parallel u_h - u \parallel^2 + j_h(u_h)] \leq \\ & \leq a(u, v-u) + j(v) - L(v-u) \, \forall v \in U. \end{cases} \tag{6.11}$$

By density of U, (6.11) holds $\forall v \in V$. Replacing v by u in (6.11) and using condition (ii) we obtain

$$j(u) \leq \liminf_{h\to 0} j_h(u_h) \leq \limsup_{h\to 0}[\alpha \parallel u - u_h \parallel^2 + j_h(u_h)] \leq j(u).$$

This implies that

$$\lim_{h\to 0} j_h(u_h) = j(u) \quad \textit{and}$$
$$\lim_{h\to 0} \parallel u_h - u \parallel = 0.$$

This proves the theorem.

7 References

For generalities on variational inequalities from a theoretical point of view see Lions-Stampacchia [1], Lions [1], Ekeland-Temam [1]. **21**

For generalities on the approximation of variational inequalities from the numerical point of view see Falk [1], Glowinski-Lions-Tremolieres [1], [2], Strang [1], Brezzi-Hager-Raviart [1].

Chapter 2

Application of The Finite Element Method To The Approximation of Some Second Order EVI

1 Introduction

In this chapter we consider some examples of EVI of the first and second kinds. These EVI are related to second order partial differential operators (for fourth order problems see GLOWINSKI [2]). The Physical Interpretation and some properties of the solution are given. Finite element approximations of these EVI are considered and convergence results are proved. In some particular cases we prove error estimates. 22

Some of the results in this chapter may be found in G.L.T. [1],[2]. For the approximation of the EVI of first kind by the finite element methods we refer also to FALK [1], STRANG [1], MOSCO-STRANG [1], CIARLET [1], BREZZI-HAGER-RAVIART [1].

We also deal with iterative methods for solving the corresponding approximate problems (cf. CEA [1], G.L.T. [1], [2]).

2 An Example of EVI of The First Kind: The Obstacle Problem

Notations 1. *All the properties of Sobolev spaces used in this chapter are proved in LIONS [2], NECAS [1]. Usually we shall have*

- Ω*: a bounded domain in* $\mathbb{R}^2$
- $\Gamma : \partial\Omega$
- $x = \{x_1, x_2\}$ *a generic point of* Ω
- $\nabla = \{\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}\}$
- $C^m(\bar{\Omega})$*: space of m -times continuously differentiable real valued functions for which all the derivatives up to order m are continuous in* $\bar{\Omega}$*,*
- $C_0^m(\bar{\Omega}) = \{v \in C^m(\bar{\Omega}) : \mathrm{Supp}(v)$ *is a compact subset of* $\Omega\}$.

23

- $\| v \|_{m,p,\Omega} = \sum_{|\alpha| \le m} \| D^\alpha v \|_{L^p(\Omega)}$ *for* $v \in C^m(\overline{\Omega})$ *where* $\alpha = (\alpha_1, \alpha_2)$; α_1, α_2 *non-negative integers,* $|\alpha| = \alpha_1 + \alpha_2$ *and* $D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}}$.
- $W^{m,p}(\Omega)$: *completion of* $C^m(\overline{\Omega})$ *in the norm defined above.*
- $W^{m,p}(\Omega)$*: completion of* $C_0^m(\Omega)$ *is the above norm*
- $H^m(\Omega) = W^{m,2}(\Omega)$,
- $H_0^m(\Omega) = W^{m,2}(\Omega)$,

$\mathscr{D}(\Omega) = C_0^\infty(\Omega)$.

2.1 The continuous problem:

Let

$$V = H_0^1(\Omega) = \{v \in H^1(\Omega) : v|\Gamma = \text{ trace de v sur } \Gamma = 0\}$$

(cf. LIONS [2], NECAS[1]).

$$a(u, v) = \int_\Omega \nabla u.\nabla v dx$$

where

$$\nabla u.\nabla v = \frac{\partial u}{\partial x_1}\frac{\partial v}{\partial x_1} + \frac{\partial u}{\partial x_2}\frac{\partial v}{\partial x_2}.,$$

$L(v) = \langle f, v\rangle$ for $f \in V^* = H^{-1}(\Omega)$ and $v \in V$. Let $\Psi \in H^1(\Omega) \cap C^0(\overline{\Omega})$ and $\Psi|_\Gamma \leq 0$. Define $K = \{v \in H_0^1(\Omega) : v \geq \Psi a.e. \text{ on } \Omega\}$.

Then the obstacle problem is a (P_1) problem defined by :

Find u such that

$$\begin{cases} a(u, v-u) \geq L(v-u)\, \forall v \in K, \\ u \in K. \end{cases} \tag{2.1}$$

The physical interpretation of this problem is the following: let an elastic membrane occupy a region Ω in the x_1, x_2 plane; this membrane is fixed along the boundary Γ of Ω. When there is no obstacle, from the theory of elasticity the vertical displacement u, obtained by applying a vertical force F, is given by the Dirichlet problem

$$\begin{cases} -\Delta u = f \ \ in \ \ \Omega, \\ u|_\Gamma = 0 \end{cases} \tag{2.2}$$

where $f = F/t$, t being the tension. **24**

When there is an obstacle, we have a free boundary problem and the displacement u satisfies the variational inequality (2.1) with Ψ being the height of the obstacle. Similar EVI also occur, sometimes with non-symmetric bilinear forms, in mathematical models for the following problems:

- Lubrication phenomena (cf. CRYER [1]).

- Filtration of liquids in porous media (cf. BAIOCCHI [1], COMINCIOLI [1]),
- Two dimensional, irrotational flows of perfect fluids (cf. BREZIS STAMPACCHIA [1], BREZIS [1], CLAVLDINI-TOURNEMINE [1]).
- Wake problems (cf. BOURGAT- DUVAUT [1]).

2.2 Existence and uniqueness results:

For proving the existence and uniqueness of the problem (2.1), we need the following lemmas stated below without proof (for proof of the lemmas, see for instance LIONS [1], NECAS [1], STAMPACCHIA [1]).

Lemma 2.1. *Let Ω be a bounded domain in $\mathbb{R}^N$. Then the semi-norm on $H^1(\Omega)$*

$$v \to \left(\int_\Omega |\nabla_v|^2 dx \right)^{1/2}$$

is a norm on $H_0^1(\Omega)$ and it is equivalent to the norm on $H_0^1(\Omega)$ induced from $H^1(\Omega)$.

The above Lemma 2.1 is known as Poincare-Friedrichs lemma.

Lemma 2.2 (STAMPACCHIA [1])**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be uniformly Lipschitz continuous (i.e. $\exists k > 0$ such that $|f(t) - f(t')| \leq k|t - t'| \, \forall t, t' \in R$) and such f' has a finite number of points of discontinuity. Then the induced map f^* on $H^1(\Omega)$ defined by $u \to f(u)$ is a continuous map into $H^1(\Omega)$. Similar results holds for $H_0^1(\Omega)$ when ever $f(0) = 0$.*

25 **COROLLARY 2.1.** *If v^+ and v^- denote the positive and the negative parts of v for $v \in H^1(\Omega)$ (respectively $H_0^1(\Omega)$) then the map $v \to \{v^+, v^-\}$ is continuous from $H^1(\Omega) \to H^1(\Omega) \times H^1(\Omega)$ (respectively $H_0^1(\Omega) \to H_0^1(\Omega) \times H_0^1(\Omega)$). Also $v \to |v|$ is continuous.*

THEOREM 2.1. Problem (2.1) has a unique solution.

Proof. In order to apply Theorem 3.1 of Chapter 1 we have to prove that $(.,.)$ is V-elliptic and that K is a closed, convex, non-empty set.

The V-ellipticity of $a(.,.)$ follows from Lemma 2.1 and the convexity of K is trivial; then

(1) *K is non-empty.* We have

$$\Psi \in H^1(\Omega) \cap C^0(\overline{\Omega}) \text{ with } \Psi \leq 0 \text{ on } \Gamma.$$

Hence, by Corollary 2.1, $\Psi^+ \in H^1(\Omega)$. Since $\Psi|_\Gamma \leq 0$ we have $\Psi^+|_\Gamma = 0$. This implies $\Psi^+ \in H_0^1(\Omega)$; then

$$\Psi^+ = \max\{\Psi, 0\} \geq \Psi$$

Thus $\Psi^+ \in K$. Hence K is non-empty.

(2) *K is closed.* Let $v_n \to v$ strongly in $H_0^1(\Omega)$ where $v_n \in K$ and $v \in H_0^1(\Omega)$. Hence $v_n \to v$ strongly in $L^2(\Omega)$. Therefore we can extract a subsequence $\{v_{n_i}\}$ such that $v_{n_i} \to v$ a.e. on Ω. Then $v_{n_i} \geq \Psi$ a.e. on Ω implies that

$$v \geq \Psi \text{ a.e. on } \Omega;$$

therefore $v \in K$.

Hence, by Theorem 3.1 of Chapter 1, We have a unique solution for (2.1).

□

2.3 Interpretation of (2.1) as a free boundary problem

For the solution u of (2.1) we define **26**

$$\Omega^+ = \{x : x \in \Omega, u(x) > \Psi(x)\}$$
$$\Omega^0 = \{x : x \in \Omega, u(x) = \Psi(x)\}$$
$$\gamma = \partial\Omega^+ \cap \partial\Omega^0; u^+ = u|_{\Omega^+}; u^0 = u|_{\Omega^0}.$$

Classically the problem (2.1) has been formulated as the problem of finding γ *(the free boundary)* and u such that

$$-\Delta u = f \text{ on } \Omega^+, \tag{2.3}$$

$$u = \Psi \text{ on } \Omega^0, \tag{2.4}$$

$$u = 0 \text{ on } \Gamma, \tag{2.5}$$

$$u^+|_\gamma = u^0|_\gamma \tag{2.6}$$

The physical interpretation of these relations is the following: (2.3) means that on Ω^+ the membrane is strictly over the obstacle; (2.4) means that on Ω^0 the membrane is in contact with the obstacle; (2.6) is transmission relation at the free boundary.

Actually (2.3)–(2.6) are not sufficient to characteristic u since there are an infinity of solutions for (2.3)–(2.6). Therefore it is necessary to add other transmission properties: for instance, if Ψ is smooth enough (say $\Psi \in H^2(\Omega)$), we require the continuity of ∇u at γ (we may ask $\nabla u \in H^1(\Omega) \times H^1(\Omega)$).

REMARK 2.1. *This kind of free boundary interpretation holds for several problems modelled by EVI of first kind and second kind.*

2.4 Regularity of solutions

We state without proof the following regularity theorem for the problem (2.1).

THEOREM 2.2. *(BREZIS- STAMPACCHIA [1]). Let Ω be a bounded domain in $\mathbb{R}^2$ with a smooth boundary. If*

$$L(v) = \int_\Omega fv\,dx \text{ with } f \in L^p(\Omega), 1 < p < \infty \tag{2.7}$$

and

$$\Psi \in W^{2,p}(\Omega), \tag{2.8}$$

27 *then the solution of the problem* (2.1) *is in* $W^{2,p}(\Omega)$.

REMARK 2.2. *Let $\Omega \subset \mathbb{R}^N$ have a smooth boundary. We know that*

$$W^{s,p}(\Omega) \subset C^k(\overline{\Omega}) \text{ if } s > \frac{N}{p} + k \tag{2.9}$$

(cf. NECAS [1]). It follows that the solution u of (2.1) *will be in $C^1(\overline{\Omega})$ if $f \in L^p(\Omega)$, $\Psi \in W^P 2, p(\Omega)$ with $p > 2$ (take $s = 2$, $N = 2$, $k = 1$ in* (2.9)*).*

The proof of this regularity result will be given in the following simple case:

$$L(v) = \int_\Omega fv dx,\ f \in L^2(\Omega), \tag{2.10}$$

$$\Psi = 0 \text{ on } \Omega. \tag{2.11}$$

Before proving that (2.10), (2.11) imply $u \in H^2(\Omega)$, we shall recall a classical lemma (also very useful in the analysis of fourth order problems).

Lemma 2.3. *Let Ω be a bounded domain of $\mathbb{R}^N$ with a boundary Γ sufficiently smooth. Then $\| \Delta v \|_{L^2(\Omega)}$ defines a norm on $H^2(\Omega) \cap H_0^1(\Omega)$ which is equivalent to the norm induced by the $H^2(\Omega)$- norm.*

Exercise 2.1. *Prove the above Lemma 2.3 using the following regularity result due to AGMON-NIRENBERG [1]:*

If $w \in L^2(\Omega)$ and if Γ is smooth then the Dirichlet problem

$$\begin{cases} -\Delta v = w \text{ in } \Omega, \\ v|_\Gamma = 0, \end{cases}$$

has a unique solution in $H_0^1(\Omega) \cap H^2(\Omega)$ (this regularity result also holds if Ω is a convex domain with Γ Lipschitz continuous) .

We shall now apply the Lemma 2.3 to prove the following theorem using a method of BREZIS-STAMPACCHIA [2].

THEOREM 2.3. *If Γ is smooth enough, $\Psi = 0$ and $L(v) = \int_\Omega fv dx$* **28** *with $f \in L^2(\Omega)$ then the solution u of the problem* (2.1) *satisfies*

$$\begin{cases} u \in K \cap H^2(\Omega), \\ \| \Delta u \|_{L^2(\Omega)} \leq \| f \|_{L^2(\Omega)} . \end{cases} \tag{2.12}$$

Proof. From Theorem 2.1 it follows that problem (2.1) has a unique solution u, with L and Ψ as above. Let $\in > 0$, consider the following Dirichlet problem

$$\begin{cases} -\in \Delta u_\epsilon + u_\epsilon = u \text{ in } \Omega, \\ u_\epsilon|_\Gamma = 0. \end{cases} \tag{2.13}$$

Problem (2.13) has a unique solution in $H_0^1(\Omega)$ and the smoothness of Γ assures that u_ϵ belongs to $H^2(\Omega)$. Since $u \geq 0$ a.e. on Ω, by the maximum principle for second order elliptic differential operators, (cf. MECAS [1]) we have $u_\epsilon \geq 0$. Hence □

$$u_\epsilon \in K. \tag{2.14}$$

From (2.14) and (2.1) we obtain

$$a(u, u_\epsilon - u) \geq L(u_\epsilon - u) = \int_\Omega f(u_\epsilon - u)dx. \tag{2.15}$$

The V-ellipticity of $a(.,.)$ implies

$$a(u_\epsilon, u_\epsilon - u) = a(u_\epsilon - u, u_\epsilon - u) + a(u, u_\epsilon - u) \geq a(u, u_\epsilon - u),$$

so that,

$$a(u_\epsilon, u_\epsilon - u) \geq \int_\Omega f(u_\epsilon - u)dx. \tag{2.16}$$

By (2.13) and (2.16) we obtain

$$\in \int_\Omega \nabla u_\epsilon . \nabla(\Delta u_\epsilon)\, dx \geq \in \int_\Omega f \Delta u_\epsilon \, dx$$

29 so that,

$$\int_\Omega \nabla u_\epsilon . \nabla(\Delta u_\in) dx \geq \int_\Omega \Delta u_\epsilon dx. \tag{2.17}$$

By Green's formula, (2.17) implies

$$-\int_\Omega (\Delta u_\epsilon)^2 dx \geq \int_\Omega f \Delta u_\epsilon dx.$$

Thus

$$\| \Delta u_{\epsilon} \|_{L^2(\Omega)} \leq \| f \|_{L^2(\Omega)}, \tag{2.18}$$

using *Schwarz inequality* in $L^2(\Omega)$. By Lemma 2.3 and relations (2.13), (2.18) we obtain

$$\lim_{\epsilon \to 0} u_{\epsilon} = u \text{ weaklyin } H^2(\Omega), \tag{2.19}$$

(which implies that $\lim u_{\epsilon} = u$ strongly in $H^s(\Omega)$, for every $s < 2$ (cf. NECAS [1])), so that $u \in H^2(\Omega)$ with

$$\| \Delta u \|_{L^2(\Omega)} \leq \| f \|_{L^2(\Omega)} . \tag{2.20}$$

2.5 Finite Element Approximations of (2.1)

Henceforth we shall assume that Ω is a polygonal domain of $\mathbb{R}^2$. Consider a " classical" triangulation $\mathscr{C}_h$ of Ω, i.e. $\mathscr{C}_h$ is a finite set of triangles T such that

$$T \subset \overline{\Omega} \forall T \in \mathscr{C} + h, \bigcup_{T \in \mathscr{C}_h} T = \overline{\Omega}. \tag{2.21}$$

$$T_1^0 \cap T_2^0 = \Psi \forall T_1, T_2 \in \mathscr{C}_h \text{ and } T_1 \neq T_2. \tag{2.22}$$

Moreover $\forall T_1, T_2 \in \mathscr{C}_h$ and $T_1 \neq T_2$, exactly one of the following conditions must hold

$$\begin{cases} (1)\ T_1 \cap T_2 = \Psi \\ (2)\ T_1 \text{ and } T_2 \text{ have only one common vertex }, \\ (3)\ T_1 \text{ and } T_2 \text{ have only a whole common edge }. \end{cases} \tag{2.23}$$

30

As usual h will be the length of the largest edge of the triangles in the triangulation.

From now on we restrict ourselves to piecewise linear and piecewise quadratic finite element approximations.

2.5.1 Approximation of V and K.

- P_k: space of polynomials in x_1 and x_2 of degree less than or equal to k.
- $\Sigma_h = \{P \in \overline{\Omega} : P \text{ is a vertex of } T \in \mathscr{C}_h\}$
- $\Sigma_h^0 = \{P \in \Sigma_h : P \notin \Gamma\}$.
- $\Sigma_h' = \{P \in \overline{\Omega} : P \text{ is the mid point of an edge of } T \in \mathscr{C}_h\}$.
- $\Sigma_h'^0 = \{P \in \Sigma_h' : P \notin \Gamma\}$.
- $\Sigma_h^1 = \Sigma_h$ and $\Sigma_h^2 = \Sigma_h \cup \Sigma_h'$.

Figure 2.1 illustrates some further notations associated with an arbitrary triangle T. we have $m_{iT} \in \Sigma_h'$, $M_{iT} \in \Sigma_h$. The centroid of the triangle T is denoted by G_T.

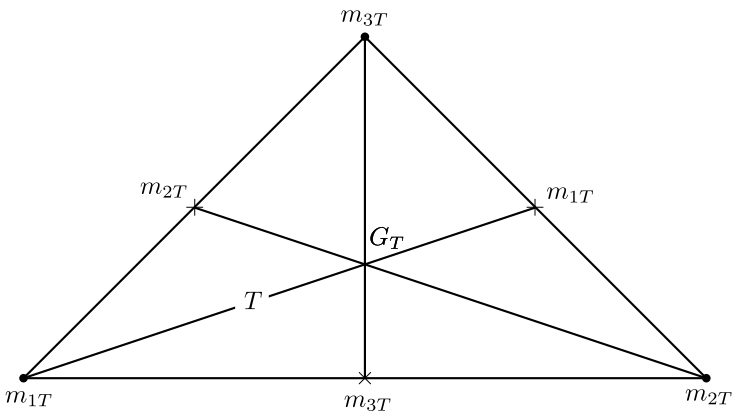

Figure 2.1:

31 The space $v = H_0^1(\Omega)$ is approximated by the family of subspaces $(V_h^k)_h$ with $k = 1$ or 2 where

$$V_h^k = \left\{v_h \in C^0(\overline{\Omega}) : v_h\big|_\Gamma = 0 \ \textit{and} \ v_h\big|_T \in P_k \ \forall T \in \mathscr{C}_h\right\}.k = 1, 2.$$

It is clear that V_h^k are finite dimensional (cf. CIARLET [1]). It is then quite natural to approximate K by

$$K_h^k = \left\{ v_h \in V_h^k : v_h(P) \geq \Psi(p) \; \forall P \in \Sigma_h^k \right\}, k = 1, 2.$$

Proposition 2.1. *The K_h^k for $k = 1, 2$ are closed, convex, non-empty subsets of V_h^k.*

Exercise 2.2. *Prove Proposition 2.1.*

2.5.2 The approximate problems

For $k = 1, 2$ the approximate problems are defined by

$$(P_{1h}^k) \begin{cases} a(u_h^k, v_h - u_h^k) \geq L(v_h - u_h^k) \; \forall v_h \in K_h^k, \\ u_h^k \in K_h^k. \end{cases}$$

From Theorem 3.1 of Chapter 1 and Proposition 2.1, it follows that

Proposition 2.2. *(P_{1h}^k) has a unique solution for $k = 1$ and 2.*

REMARK 2.3. *If the bilinear form $a(\cdot, \cdot)$ is symmetric, (P_{1h}^k) is actually equivalent to (cf. Chap. 1, Remark 3.2) the* quadratic programming *problem*

$$\min_{v_h \in K_h^k} \left[\frac{1}{2} a(v_h, v_h) - L(v_h) \right]. \tag{2.24}$$

2.6 Convergence results

In other to simplify the convergence proof we shall assume in this section that

$$\Psi \in C^0(\overline{\Omega}) \cap H^1 \quad \text{and} \quad \Psi \leq 0 \quad \text{in a neighbourhood of} \quad \Gamma. \tag{2.25}$$

Before proving the convergence results we shall give two important numerical quadrature schemes which will be used to prove the convergence theorem. 32

Exercise 2.3. *With notations as in Fig.2.1, prove the following identities for any triangle T*

$$\int_T w dx = \frac{meas.(T)}{3} \sum_{i=1}^{3} w(M_{iT}) \, \forall w \in P_1. \tag{2.26}$$

$$\int_T w dx = \frac{meas.(T)}{3} \sum_{i=1}^{3} w(m_{iT}) \, \forall w \in P_2. \tag{2.27}$$

Formula (2.26) is called the *Trapezoidal Rule* and (2.27) is known as *Simpson's Integral formula*. These formulae, not only have theoretical importance but also practical utility.

We have the following results about the convergence of u_h^k (solutions of the problem (P_h^k)) as $h \to 0$.

THEOREM 2.4. *Suppose that the angles of the triangles of $\mathscr{C}_h$ are uniformly bounded below by $\theta_0 > 0$ as $h \to 0$; then for $k = 1, 2$*

$$\lim_{h \to 0} \| u_h^k - u \|_{H_0^1(\Omega)} = 0 \tag{2.28}$$

where u_h^k and u are respectively the solutions of P_{1h}^k and (2.1).

Proof. In this proof we shall use the following *density* result to be proved later:

$$\overline{\mathscr{D}(\Omega) \cap K} = K. \tag{2.29}$$

□

To prove (2.28) we shall use Theorem 5.2 of Chap. 1. To do this we have to verify that the following two properties hold (for $k = 1, 2$):

(i) If $(v_h)_h$ is such that $v_h \in K_h^k \, \forall h$ and converges weakly to v as $h \to 0$, then $v \in K$.

(ii) There exists $\chi, \overline{\chi} = K$ and $r_h^k : \chi \to K_h^k$ such that $\lim_{h \to 0} r_h^k v = v$ =*strongly* in $V \, \forall v \in \chi$.

33 **Verification of (i).**Using the notations of Fig.2.1 and considering $\phi \in \mathscr{D}(\Omega)$ with $\phi \geq 0$, we define ϕ_h by $\phi_h = \sum_{T \in \mathscr{C}_h} \phi(G_t)\chi_T$ where χ_T is the *characteristic function* of T and G_T is the centroid of T. It is easy to see from the uniform continuity of ϕ that

$$\lim_{h \to 0} = \phi \ \textit{strongly in} \ \ L^\infty(\Omega). \tag{2.30}$$

Then we approximate Ψ by Ψ_h such that

$$\begin{cases} \Psi_h \in C^0(\overline{\Omega}), \Psi_h|_T \in P_k \forall T \in \mathscr{C}_h, \\ \Psi_h(P) = \Psi(P) \, \forall P \in \Sigma_h^k. \end{cases} \tag{2.31}$$

This function Ψ_h satisfies

$$\lim_{h \to 0} \Psi_h = \Psi \ \textit{strongly in} \ \ L^\infty(\Omega). \tag{2.32}$$

Let us consider a sequence $(v_h)_h$, $v_h \in K_h^k \, \forall h$ such that

$$\lim_{h \to 0} v_h = v \ \textit{weakly in} \ \ V.$$

Then $\lim_{h \to o} v_h = v$ *strongly* in $L^2(\Omega)$, (cf. NECAS [1]) which, using (2.30) and (2.32), implies that

$$\lim_{h \to 0} \int_\Omega (v_h - \Psi_h)\Psi_h dx = \int_\Omega (v - \Psi)\phi dx, \tag{2.33}$$

(actually since $\phi_h \to \phi$ *strongly* in $L^\infty(\Omega)$ the weak convergence of v_h in $L^2(\Omega)$ is enough to prove (2.33)).

We have

$$\int_\Omega (v_h - \phi_h)\phi_h dx = \sum_{T \in \subset_h} \phi(G_t) \int_T (v_h - \Psi_h) dx. \tag{2.34}$$

From (2.26). (2.27) and the definition of Ψ_h we obtain for all $T \in \mathscr{C}_h$.

$$\int_T (v_h - \Psi_h) dx = \frac{meas.(T)}{3} \sum_{i=1}^{3} [v_h(M_{iT}) - \Psi_h(M_{iT})] \text{ if } k = 1, \tag{2.35}$$

$$\int_T (v_h - \Psi_h)dx = \frac{meas.(T)}{3} \sum_{i=1}^{3} [v_h(m_{iT}) - \Psi_h(m_{iT})] \text{ if } k = 2, \quad (2.36)$$

34

Using the fact that $\phi_h \geq 0$, the definition of K_h^k and the relations (2.35), (2.36) it follows from (2.34) that

$$\int_\Omega (v_h - \Psi_h)\phi_h \, dx \geq 0 \, \forall \phi \in \mathscr{D}(\Omega), \phi \geq 0,$$

so that as $h \to 0$

$$\int_\Omega (v - \Psi)\phi dx \geq 0 \, \forall \phi \in \mathscr{D}(\Omega), \phi \geq 0$$

which in turn implies $v \geq \Psi$ a.e. in Ω. Hence (i) is verified.

Verification of (ii). From (2.29) it is natural to take $\chi = \mathscr{D} \cap K$. We define $r_h^k : H_0^1(\Omega) \cap C^0(\overline{\Omega}) \to V_h^k$ as the " linear " interpolation operator when $k = 1$ and " quadratic" interpolation operator when $k = 1$, i.e.

$$\begin{cases} r_h^k v \in V_h^k \, \forall v \in H_0^1(\Omega) \cap C^0(\overline{\Omega}), \\ (r_h^k v)(p) = v(p) \, \forall P \in \sum_h^{\circ k} \ for \quad k = 1, 2. \end{cases} \quad (2.37)$$

On the one hand it is known (cf. for instance CIARLET [1], [2] , STRANG-FIX [1]) that under the assumptions made on $\mathscr{C}_h$ in statement of Theorem 2.4 we have

$$\| r_h^k v - v \|_v \leq C \, h^k \, \| v \|_{H^{k+1}(\Omega)} \ \forall v \in \mathscr{D}(\Omega), k = 1, 2.$$

with C independent of h and v . This implies that

$$\lim_{h \to 0} \| r_h^k v - v \|_v = 0 \, \forall \in \chi, k = 1, 2.$$

On the other hand it is obvious that

$$r_h^k v \in K_h^k \, \forall v \in K \cap C^0(\overline{\Omega}).$$

35 so that

$$r_h^k v \in K_h^k \, \forall v \in \chi, \ \ for \quad k = 1, 2.$$

In conclusion with the above χ and r_h^k, (ii) is satisfied. Hence we have proved the Theorem 2.4 modulo the proof of the density result (2.29).

Lemma 2.4. *Under the assumptions* (2.25) *we have* $\overline{\mathscr{D}(\Omega) \cap K} = K$.

Proof. Let us prove the Lemma in two steps. □

Step 1. *Let us show that*

$$\mathscr{K} = \{v \in K \cap C^0(\overline{\Omega}) : \textit{vcompact support in } \Omega\} \tag{2.38}$$

is dense in K.

Let $v \in K$, $K \subset H_0^1(\Omega)$ implies that exists a sequence $\{\phi_n\}_n$ in $\mathscr{D}(\Omega)$ such that

$$\lim_{n\to\infty} \phi_n = v \ \textit{strongly in} \quad V.$$

Define v_n by

$$v_n = \max(\psi, \phi_n) \tag{2.39}$$

so that

$$v_n = \frac{1}{2}[(\Psi + \phi_n) + |\Psi - \phi_n|].$$

Since $v \in K$, from Corollary 2.1 and relations (2.39) i follows that

$$\lim_{n\to\infty} v_n = \frac{1}{2}[(\Psi + v) + |\Psi - v|] = \max(\Psi, v) = v \ \textit{strongly in} \quad V. \tag{2.40}$$

From (2.25) and (2.39) it follows that

$$\textit{each } v_n \textit{ has a compact support in } \Omega, \tag{2.41}$$

$$v_n \in K \cap C^0(\overline{\Omega}). \tag{2.42}$$

From (2.40) - (2.42) we obtain (2.38) **36**

Step 2. *Let us show that*

$$\mathscr{D}(\Omega) \cap \mathscr{K} \textit{ is dense } \mathscr{K}. \tag{2.43}$$

This proves from Step 1, that $\mathscr{D}(\Omega) \cap K$ is dense in K. Let ρ_n be a sequence of mollifiers, i.e.

$$\begin{cases} \rho_n \in \mathscr{D}(\mathbb{R}^2), \rho_n \geq 0, \\ \int_{\mathbb{R}^2} \rho_n(y)dy = 1 \\ \bigcap_{n=1}^{\infty} \operatorname{Supp} \rho_n = \{0\}, \{\operatorname{Supp} \rho_n\} \text{is a decreasing sequence.} \end{cases} \tag{2.44}$$

Let $v \in \mathscr{K}$. Let $\tilde{v}$ extension of v defined by

$$\tilde{v}(x) = \begin{cases} v(x) \text{ if } x \in \Omega, \\ 0 \text{ if } x \notin \Omega, \end{cases}$$

then $\tilde{v} \in H^1(\mathbb{R}^2)$. Let $\tilde{v}_n = \tilde{v} * \rho_n$ i.e.

$$\tilde{v}_n(x) = \int_{\mathbb{R}^2} \rho_n(x-y)\tilde{v}(y)dy \tag{2.45}$$

then

$$\begin{cases} \tilde{v}_n \in \mathscr{D}(\mathbb{R}^2), \\ \text{Supp } \tilde{v}_n \subset \text{Supp } v + \text{Supp } \rho_n' \\ \lim\limits_{n\to\infty} \tilde{v}_n = \tilde{v} \text{ strongly in } H^1(\mathbb{R}^2). \end{cases} \tag{2.46}$$

Hence from (2.41) and (2.46) we have

$$\text{Supp}(\tilde{v}_n) \subset \Omega \text{ for } n \text{ large enough.} \tag{2.47}$$

37 We also have (since Supp($\tilde{v}$)is bounded)

$$\lim \tilde{v}_n = \tilde{v} \ \textit{strongly in} \ \ L^\infty(\mathbb{R}^2). \tag{2.48}$$

Define $v_n = \tilde{v}_n|_\Omega$, then (2.46)–(2.48) imply

$$\begin{cases} v_n \in \mathscr{D}(\Omega) \\ \lim\limits_{n\to\infty} v_n = v \textit{ strongly in } H_0^1(\Omega) \cap C^0(\overline{\Omega}); \end{cases} \tag{2.49}$$

$v \in \mathscr{K}$ and $\Psi \leq 0$ in a neighbourhood of Γ imply that there exists a $\delta > 0$ such that

$$v = 0, \Psi \leq 0 \textit{ on } \ \Omega_\delta \tag{2.50}$$

where $\Omega_\delta = \{x \in \Omega : d(x, \Gamma) < \delta\}$.

From (2.48) and (2.50) it follows that $\forall \in > 0$, there exists an $n_0 = n_0(\in)$ such that $\forall n \geq n_0(\in)$

$$\begin{cases} v(x) - \in \leq v_n(x) \leq (x) + \in \ \forall x \in \Omega - \Omega_{\delta/2} \\ v_n(x) = v(x) = 0 \text{ for } x \in \Omega_{\delta/2} \end{cases} \tag{2.51}$$

Since $\overline{\Omega} - \Omega_{\delta/2}$ is a compact subset of $\overline{\Omega}$ there exists a functions θ (cf. for instance *H*. CARTAN [1]) such that

$$\begin{cases} \theta \in \mathscr{D}(\Omega), \theta \geq 0 \text{ in } \Omega \\ \theta(x) = 1 \; \forall x \in \overline{\Omega} - \Omega_{\delta/2} \end{cases} \tag{2.52}$$

Finally define $w_n^\epsilon = v_n + \in \theta$.

Then from (2.49), (2.51) and (2.52) we have

$$w_n^\epsilon \in \mathscr{D}(\Omega)$$

$$\lim_{\substack{\in \to 0 \\ n \to \infty \\ n \geq n_0(\in)}} w_n^\epsilon = v \text{ strongly in } H_0^1(\Omega),$$

with $w_n^\epsilon(x) \geq v(x) \geq \Psi(x) \, \forall x \in \Omega$, so that Step 2, is proved. **38**

REMARK 2.4. *Analysing the verification (i) in the proof of Theorem 2.4, we observe that if for $k = 2$ we use, instead of K_h^2, the following convex set*

$$\{v_h \in V_h^2 : v_h(p) \geq \Psi(p) \, \forall P \in \sum_h^{0\prime}\}$$

then the convergence of u_h^2 to u still holds provided $\mathscr{C}_h$ obeys the same assumptions as in the statement of Theorem 2.4.

Exercise 2.4. *Extend the previous analysis if Ω is not a polygonal domain.*

Exercise 2.5. *Let Ω be a bounded domain of $\mathbb{R}^2$ and Γ_0 a "nice" subset of Γ, see Fig. 2.2. Define V by $V = \{v \in H^1(\Omega) : v|\Gamma_0 = 0\}$. Taking the bilinear form $a(\cdot, \cdot)$ like in* (2.1)*, and $L \in V^*$, study the following EVI*

$$\begin{cases} a(u, v - u) \geq L(v - u) \, \forall v \in K, \\ u \in K, \end{cases}$$

where $K = \{v \in V : v \geq \Psi \text{ a.e. in } \Omega\}$ and $\Psi \in C^0(\overline{\Omega}) \cap H^1(\Omega)$, $\Psi \leq 0$ in a neighbourhood of Γ_0. Also study the finite element approximation of the above EVI.

Hint. Use the fact that if Γ and Γ_0 are smooth enough then $\mathscr{V} = V$ where (see Fog. 2.2) $\mathscr{V} = \{v \in C^\infty(\overline{\Omega}) : v = 0$ in a neighbourhood of $\Gamma_0\}$.

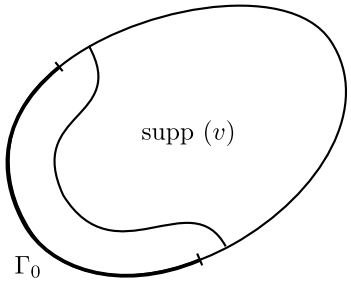

Figure 2.2:

2.7 Comments on the error estimates

39 We do not emphasize too much on this subject since it has been done in detail in CIARLET [1], Chap.9, at least for piecewise linear approximations.

2.7.1 Piecewise linear approximation

Using piecewise linear finite elements and assuming that f, Ψ, $u \in H^2(\Omega)$, $0(h)$ estimates for $\| u-u_h \|_{H^1(\Omega)}$ have been obtained by FALK[1], [2], [3], STRANG [1], STRANG-MOSCO [1]. We also refer to CLARLET [1, Chap. 9], in which the Falk analysis is given.

2.7.2 Piecewise quadratic approximation

Assuming more regularity for f, Ψ, u that in the previous case, assuming also some smoothness hypotheses for the free boundary, an $0(h^{3/2-\epsilon})$ estimate for $\| u_h - u \|_{H61(\Omega)}$ has been obtained by BREZZI-HAGER-RAVIART [1], BREZZI-SACCHI [1] for an approximation by piecewise quadratic finite elements, similar to the described in Section 2.6.

2.8 Iterative solution of the approximation problem

Once the continuous problem has been approximated and the convergence proved, it remains to compute effectively the approximate solution. In the case of the discrete obstacle problem this can be done easily by using an *over- relaxation method with projection* as described in CEA[2].

Let us justify the use of this method. It follows from Remark 2.3 that the discrete problem is of the following type

$$\min_{v\in C}\left[\frac{1}{2}(Av,v)-(b,v)\right] \tag{2.53}$$

where $(\cdot,\cdot)$ denotes the usual inner product in $\mathbb{R}^N$ and $v=\{v_1,\ldots,v_n\}$ and

$$A=(a_{ij}),1\le i\le N,1\le j\le N \tag{2.54}$$

is a symmetric, positive definite $N\times N$ and C is the set given by

$$C=\{v\in\mathbb{R}^N : v_i\ge\Psi_i,1\le i\le N\}. \tag{2.55}$$

Since C is the product of closed intervals of $\mathbb{R}$, the over-relaxation **40** method with projection on C can be used. Let us describe it in detail:

$$\begin{aligned} &u^0\in C, u^0 \textit{ arbitrarily chosen in } C\\ &(u^0=\{\Psi_1,\ldots,\Psi_n\} \textit{ may be a good guess}). \end{aligned} \tag{2.56}$$

Then u^n being known, we compute u^{n+1}, component by component using for $i=1,2,\ldots N$

$$u_i^{-n+1}=\frac{1}{a_{ii}}\left(b_i-\sum_{j=1}^{i-1}a_{ij}u_j^{n+1}-\sum_{j=i+1}^{N}a_{ij}u_j^n\right). \tag{2.57}$$

$$u_i^{n+1}=P_i(u_i^n+w(u_i^{-n+1}-u_i^n)) \tag{2.58}$$

where

$$P_i(x)=\max(x,\Psi_i)\,\forall x\in\mathbb{R}. \tag{2.59}$$

It follows from CEA [2] (see also CEA-GLOWINSKI [1], G.L.T. [1])that

Proposition 2.3. *Let* (u^n) *be defined by* (2.56)–(2.59). *Then for every* $u^0 \in C$ *and* $\forall 0 < w < 2$, *we have* $\lim_{n\to\infty} u^n = u$ *where* u *is the unique solution of* (2.53).

REMARK 2.5. *In the case of the discrete obstacle problem the components of* u *will be the values taken by the approximation solution at the nodes of* $\overset{0}{\underset{h}{\sum}}$ *if* $k = 1$ *and* $\overset{0}{\underset{h}{\sum}} \cup \overset{0}{\underset{h}{\sum}}'$ *if* $k - 2$. *Similarly* Ψ_i *will be the values taken by* Ψ *at the nodes stated above, assuming these nodes have been ordered from 1 to* N.

REMARK 2.6. *The optimal choice for* ω *is a critical but nontrivial point. However it has been observed from numerical experiments that the so-called* Young method *for obtaining the optimal value of* ω *during the iterative process itself, leads to a value of* ω *with good convergence properties. The convergence of this method has been proved for linear equations and requires special properties for the matrix of the system (see YOUNG [1], VARGA [1]). However, empirical justification of its success for the obstacle problem can be made, but will not be given here.*

REMARK 2.7. *From numerical experiments it is found that the optimal value of* ω *is always strictly greater than one.*

3 A Second Example of EVI of The First Kind: The Elasto-Plastic Torsion Problem

3.1 Formulation. Preliminary results

41 Let Ω be a bounded domain of $\mathbb{R}^2$ with a smooth boundary Γ. With the same definition for V, $a(\cdot,\cdot)$, $L(.)$ as in Sec. 2.1 of this Chapter, we consider the following EVI of the first kind

$$\begin{cases} a(u, v-u) \geq L(v-u)\ \forall v \in K, \\ u \in K, \end{cases} \tag{3.1}$$

where

$$K = \{v \in H_0^1(\Omega) : |\nabla_v| \leq 1 a.e. \text{ in } \Omega\}. \tag{3.2}$$

THEOREM 3.1. *Problem* (3.1) *has a unique solution.*

Proof. In order to apply Theorem 3.1 of Chapter I, we only have to verify that K is a *non- empty, closed, convex,* subset of V. □

K is non-empty because $0 \in K$, and the convexity of K is obvious. To prove that K is closed, consider a sequence $\{v_n\}$ in K such that $v_n \to v$ strongly in V. Then there exists a subsequence $\{v_{n_i}\}$ such that

$$\lim_{i\to\infty} \nabla v_{n_i} = \nabla v \text{ a.e.}$$

Since $|\nabla v_n| \leq 1$ a.e. we get $|\nabla v| \leq 1$ a.e. therefore $v \in K$. Hence K is closed.

The following Proposition gives a very useful property of K.

Proposition 3.1. *K is compact in $C^0(\bar{\Omega})$ and*

$$|v(x)| \leq \dot{d}(x,\Gamma)\ \forall x \in \Omega \text{ and } \forall v \in K, \tag{3.3}$$

where $d(x,\Gamma)$ is the distance from x to Γ.

Exercise 3.1. *Prove Proposition 3.1* **42**

REMARK 3.1. *Let us define u_∞ and $u_{-\infty}$ by*

$$u_\infty(x) = d(x,\Gamma)$$
$$u_{-\infty}(x) = -d(x,\Gamma).$$

Then u_∞ and $u_{-\infty}$ to K. We observe that u_∞ is the *maximal* element of K and $u_{-\infty}$ is the *minimal* element of K.

REMARK 3.2. *Since $a(\cdot,\cdot)$ is* symmetric *the solution u of* (3.1) *is characterised (see Section 3.2 of Chap. 1) as the unique solution of the minimization problem*

$$\begin{cases} J(u) \leq J(v)\ \forall v \in K, \\ u \in K, \end{cases} \tag{3.4}$$

with $J(v) = \dfrac{1}{2}a(v,v) - L(v)$.

3.2 Physical motivation

Let us consider an infinitely long cylindrical bar of cross-section Ω where Ω is *simply connected*. Assume that this is made up of an *isotropic, elastic, perfectly plastic* material whose plasticity yield is given by the Von Misses Criterion. (For a general discussion of plasticity problems, see KOITER [1], DUVAUT-LIONS [1, Chap. 5]). Starting from a *zero stress initial state*, an increasing *torsion* moment is applied to the bar. The torsion is characterised by C which is defined as the *torsion angle per unit length*. Then for all C, it follows from the *Harr- Karman Principle* that the determination of the *stress field* is equivalent (in a convenient system of physical units) to the solution of the following variational problem:

$$\min_{v\in K} \frac{1}{2}\int_\Omega |\nabla_v|^2 dx - C\int_\Omega v dx. \tag{3.5}$$

This is a particular case of (3.1) or (3.4) with

$$L(v) = C\int_\Omega v dx. \tag{3.6}$$

43 The *stress vector* σ in a cross - section is related to u by $\sigma = \nabla u$, so that u is a *Stress potential* and we obtain σ once the solution of (3.5) is known.

Proposition 3.2. *Let us denote by u_C the solution of* (3.5) *and let, as before $u_\infty = d(x,\Gamma)$ then $\lim_{C\to\infty} u_C = u_\infty$ strongly in $H_0^1(\Omega)\cap C^0(\overline{\Omega})$.*

Proof. Since u_C is the solution of (3.5), it is characterised by

$$\begin{cases} \int_\Omega \nabla u_C.\nabla(v-u_C)dx \geq C\int_\Omega (v-u_C)dx \;\forall v\in K, \\ u_C \in K. \end{cases} \tag{3.7}$$

□

Since $u_\infty \in K$, from (3.7) we have

$$\int_\Omega \nabla u_C.\nabla(u_\infty - u_C)dx \geq C\int_\Omega (u_\infty - u_C)dx, \tag{3.8}$$

i.e.

$$\begin{cases} C\int_\Omega (u_\infty - u_C)dx + \int_\Omega |\nabla u_C|^2 dx \le \int_\Omega \nabla u_\infty . \bigtriangledown u_C dx \\ \le \int_\Omega |\nabla u_\infty| . \nabla u_C| dx \le (\text{ meas } .\Omega). \end{cases} \tag{3.9}$$

From (3.3) we have $u_\infty - u_C \ge 0$ so that (3.9) implies

$$\| u_\infty - u_C \|_{L^1(\Omega)} \le C^{-1}(\text{ meas}.\Omega).$$

Which in turn implies

$$\lim_{C\to\infty} \| u_\infty - u_C \|_{L^1(\Omega)} = 0. \tag{3.10}$$

Form the definition of K and from the Proposition 3.1 we get that K is bounded and weakly closed in V and hence weakly compact in V. Further K is compact in $C^0(\overline{\Omega})$.

Relation (3.10) implies

$$\begin{cases} \lim_{C\to\infty} u_C = u_\infty \ \textit{strongly in } C^0(\overline{\Omega}), \\ \lim_{C\to\infty} u_C = u_\infty \ \textit{weakly in } V. \end{cases} \tag{3.11}$$

It follows from (3.8) that 44

$$\begin{cases} \int_\Omega \nabla u_\infty . \nabla(u_\infty - u_C) \ge \int_\Omega |\nabla(u_\infty - u_C)|^2 dx + C\int_\Omega (u_\infty - u_C)dx \\ = \| u_C - u_\infty \|_V^2 + C \| u_\infty - u_C \|_{L^1(\Omega)} . \end{cases} \tag{3.12}$$

It follows easily from (3.11) and (3.12) that

$$\lim_{C\to\infty} C \| u_\infty - u_C \|_{L^1(\Omega)} = 0,$$
$$\lim_{C\to\infty} \| u_\infty - u_C \|_V = 0.$$

REMARK 3.3. *In the case of multiply connected cross section, the variational formulation of the torsion problem has to redefined (see LANCHON [1], GLOWINSKI - LANCHON [1], GLOWINSKI [1, Chap 4]).*

3.3 Regularity properties and exact solutions

3.3.1 Regularity results

THEOREM 3.2. *(BREZIS-STAMPACCHIA [2]). Let u be a solution of* (3.1) *or* (3.4) *and* $L(v) = \int_\Omega fv\,dx$.

(1) *Let* Ω *be a bounded convex domain of* $\mathbb{R}^2$ *with* Γ *Lipschitz continuous and* $f \in L^p(\Omega)$ *with* $1 < p < \infty$. *Then we have*

$$u \in W^{2,p}(\Omega). \tag{3.13}$$

(2) *If* Ω *is a bounded domain of* $\mathbb{R}^2$ *with a smooth boundary* Γ; *if* $f \in L^p(\Omega)$ *with* $1 < p < \infty$ *then* $u \in W^{2,p}(\Omega)$.

45 **REMARK 3.4.** *It will be seen in the next section that, in general, there is a limit for the* regularity*of the* solution *of* (3.1) *even if* Γ *and* f *are very smooth.*

REMARK 3.5. *It has been proved by H. BREZIS that, under quite restrictive smoothness assumptions on* Γ *and* f *we may have*

$$u \in W^{2,\infty}(\Omega).$$

3.3.2 Exact solutions

In this section we are going to give some examples of problems (3.1) for which exact solutions are known.

Example 1. *We take* $\Omega = \{x : 0 < x < 1\}$ *and* $L(v) = C\int_0^1 v\,dx$ *with* $c > 0$.

Then the explicit form (3.1) *is*

$$\begin{cases} \int_0^1 u'(v' - u')dx \geq c\int_0^1 (v - u)dx\ \forall v \in K, \\ u \in K, \end{cases} \tag{3.14}$$

where $K = \{v \in H_0^1(\Omega) : |v'| \leq 1$ *a.e. on* $\Omega\}$ *and* $v' = \dfrac{dv}{dx}$.

The exact solutions of (3.14) is given by

$$u(x) = \frac{c}{2}x(1-x)\,\forall x, \text{ if } c \le 2. \tag{3.15}$$

If $c > 2$

$$u(x) = \begin{cases} x \text{ if } 0 \le x \le \frac{1}{2} - \frac{1}{c} \\ \frac{c}{2}[x(1-x) - (\frac{1}{2} - \frac{1}{c})] \text{ if } \frac{1}{2} - \frac{1}{2} \le x \le \frac{1}{2} + \frac{1}{2}, \\ 1 - x \text{ if } \frac{1}{2} + \frac{1}{c} \le x \le 1. \end{cases} \tag{3.16}$$

Example 2. *In this example we consider a two dimensional problem. We take*

$$\Omega = \{x : x_1^2 + x_2^2 < R^2\},$$
$$L(v) = c\int_\Omega v dx \text{ with } c > 0.$$

Then setting $r = (x_1^2 + x_2^2)^{1/2}$ the solution u of (3.1) is given by **46**

$$u(x) = \frac{c}{4}(R^2 - r^2) \text{ if } c \le \frac{2}{R}, \tag{3.17}$$

if $c > \dfrac{2}{R}$ then

$$u(x) = \begin{cases} R - r \text{ if } \frac{2}{c} \le r \le R, \\ \frac{c}{4}[(R^2 - r^2) - (R - \frac{2}{c})^2] \text{ if } 0 \le r \le \frac{c}{2}. \end{cases} \tag{3.18}$$

These examples illustrate Remark 3.4. We see that for c large enough we have

$$u \in W^{2,\infty}(\Omega) \cap H_0^1(\Omega), u \notin H^3(\Omega). \tag{3.19}$$

In fact we have

$$u \in H^s(\Omega)\,\forall s < \frac{5}{2}.$$

Exercise 3.2. *Verify that u given in the above two examples are exact solutions of the corresponding problems.*

3.4 An equivalent variational formulation

In H. BREZIS-M, SIBONY [1] it is proved that if Ω is a bounded domain of $\mathbb{R}^2$ with a smooth boundary Γ and if

$$L(v) = c\int_\Omega v(x)\,dx (c > 0 \textit{ for instance}),$$

$$\begin{cases} a(u, v-u) \geq c\int (v-u)dx \ \forall v \in \hat{K}, \\ u \in \hat{K} = \{v \in H_0^1(\Omega), |v(x)| \leq d(x,\Gamma) a.e.\}. \end{cases} \tag{3.20}$$

The above problem (3.20) is very similar to the obstacle problem considered in Sec. 2 of this Chapter. Since $a(\cdot,\cdot)$ is symmetric, (3.20) is also equivalent to

$$\begin{cases} J(u) \leq J(v) \quad \forall v \in \hat{K}, \\ u \in \hat{K}, \end{cases} \tag{3.21}$$

47 with

$$J(v) = \frac{1}{2}a(v,v) - c\int_\Omega v(x)dx.$$

The numerical solutions of (3.20) and (3.21) is considered in G.L.T [1, Chap. 3] (see also CEA[2, Chap. 4]).

Exercise 3.3. *Study the numerical analysis of* (3.21).

Exercise 3.4. *Assume* $c > 0$ *in* (3.20). *Then prove that the solution u of* (3.20) *is also the solution of the EVI obtained by replacing K by* $\{v \in H_0^1(\Omega) : v(x) \leq d(x,\Gamma) a.e.\}$ *in* (3.20).

3.5 Finite Element Approximations of (3.1).

We consider in this section an approximation of (3.1) be *first order finite elements*. From the view point of applications in mechanics (in which $f = c$) it seems that, given the equivalence of (3.1) and (3.20), it is sufficient to approximate (3.20) (using essentially the same method as in Sec. 2). However, in view of other possible applications, it seems to us that it would be interesting to consider the numerical solution of (3.1) working directly with K instead of $\hat{K}$. For the numerical analysis of (3.20) by Finite Differences sec G.L.T. [1. Chap. 3] and CEA-GLOWINSKI-NEDELEC [1].

3.5.1 Approximation of V and K.

We use the notation of Sec. 2.5 of this Chap. We assume that Ω is a polygonal domain of $\mathbb{R}^2$ (see Remark 3.8 below for the non polygonal case) and we consider a triangulation τ_h of Ω satisfying (2.21)–(2.23). Then V and K are respectively approximated by

$$v_h = \{v_h \in C^0(\overline{\Omega}) : v_h = 0 \text{ on } \Gamma, v_h|_T \in P_1 \ \forall T \in \tau_h\},$$
$$K_h = K \cap V_h$$

Then one can easily prove

Proposition 3.3. *K_h is a closed, convex, non-empty subset of V_h.* **48**

REMARK 3.6. *If $v_h \in V_h$ then ∇v_h is a constant vector on every $T \in \tau_h$.*

3.5.2 The approximate problem

The approximate problem is defined by :

$$\begin{cases} \text{Find } u_h \in K_h \text{ such that} \\ a(u_h, v_h - u_h) \geq L(v_h - u_h) \ \forall v_h \in K. \end{cases} \tag{3.22}$$

One can easily prove

Proposition 3.4. *The approximate problem* (3.22) *has a unique solution.*

One may find in Sec 7. of this chapter practical formulae related to finite element approximation. Using these formulae, (3.22) and the equivalent problem (3.23) can be expressed in a form more suitable for computation.

REMARK 3.7. *Since $a(\cdot,\cdot)$ is symmetric,* (3.22) *is equivalent to the non-linear programming problem*

$$\min_{v_h \in K_h} \left[\frac{1}{2} a(v_h, v_h) - L(v_h) \right]. \tag{3.23}$$

The natural variables in (3.23) are the values taken by v_h over the set $\overset{0}{\sum_h}$ of the interior nodes of τ_h. Then the number of variables in (3.23) is Card $(\overset{0}{\sum_h})$. The number of constraints is the number of triangles i.e. Card (τ_h) and each constraint is quadratic w.r.t. the variables since

$$|\nabla v_h| \leq 1 \text{ iff } |\nabla v_h|^2 \leq 1 \text{ over T.} \tag{3.24}$$

REMARK 3.8. *If Ω is not polygonal, it is always possible to approximate Ω by a polygonal domain Ω_h in such a way that all vertices of $\Gamma_h = \partial\Omega_h$ belong to Γ. Then instead of defining* (3.22) *over Ω we define it over Ω_h.*

3.5.3 Remarks on the use of higher order finite elements

In these notes only an approximation of (3.1) by first order finite elements has been considered. That fact is justified by the existence of a *regularity limitation* for the solution of (3.2), which implies that even very smooth data one may have $u \notin V \cap H^3(\Omega)$ (see examples of Sec. 3.3.2.

49 We refer to G.L.T. [1, Chap. 3] and GLOWINSKI [1, Chap . 4. Sec. 3.5.3] for further discussions on the use of finite elements of order ≥ 2.

3.6 Convergence results . General case

In this sections we take $L(v) =< f, v >$, for $f \in H^{-1}(\Omega) = V$.

3.6.1 A density Lemma

In order to apply the general results of Chap. 1, the following density lemma will be very useful

Lemma 3.1. *We have*

$$\overline{\mathscr{D}(\Omega) \cap K} = K. \tag{3.25}$$

Proof. We use the notation of Lemma 2.4. Let $v \in K$ and $\in > 0$; define v_ϵ by

$$v_\epsilon = (v - \in)^+ - (v + \in)^-. \tag{3.26}$$

□

Then we have $v_\epsilon \in H^1(\Omega)$ with $| \nabla v_\epsilon| \leq 1$ a.e. in Ω. From the inclusion $K \subset \hat{K} = \{v \in V : |v(x)| \leq d(x, \Gamma)$ a.e. in $\Omega\}$ it follows that

$$\begin{cases} v_\epsilon(x) = 0 \text{ if } d(x, \Gamma) \leq \in, \\ |v_\epsilon(x)| \leq d(x, \Gamma) - \in \quad \textit{elsewhere} \end{cases} \tag{3.27}$$

so that from (3.27) it follows that

$$v_\epsilon \in K \textit{and has a compact support in } \Omega. \tag{3.28}$$

From Corollary 2.1 we have

$$\lim_{\in \to 0} v_\epsilon = v \ \textit{ strongly in } \ V. \tag{3.29}$$

From (3.28) and (3.29) it follows that if $\mathscr{K} = \{v \in K : v$ has a compact support in $\Omega\}$, then $\mathscr{K} = K$.

Thus to prove the lemma it suffices to prove that any $v \in \mathscr{K}$ can be approximated by a sequence $(v_n)_n$ of functions in $\mathscr{D}(\Omega) \cap K$. Let ρ_n be a mollifying sequence as defined in Lemma 2.4 of this chapter. **50**
Let $v \in \mathscr{K}$. Denote by $\tilde{v}$ extension of v to $\mathbb{R}^2$ putting outside Ω. Then $\tilde{v} \in H^1(\mathbb{R}^2)$.

Let $\tilde{v}_n = \tilde{v} * \rho_n$ so that

$$\tilde{v}_n(x) = \int_{\mathbb{R}^2} \rho_n(x - y)\tilde{v}(y)dy, \tag{3.30}$$

$$\nabla\tilde{v}_n(x) = \int_{\mathbb{R}^2} \rho_n(x - y) \nabla \tilde{v}(y)dy. \tag{3.31}$$

Then

$$\tilde{v}_n \in \mathscr{D}(\mathbb{R}^2) \textit{and } \lim_{n \to \infty} \tilde{v}_n = \tilde{v} \textit{ strongly in } H^1(\mathbb{R}^2). \tag{3.32}$$

Since Supp $\tilde{v} \subset \Omega$, from (3.30) we get

$$\text{Supp}\,\tilde{v}_n \subset \Omega \textit{ for } n \textit{ sufficiently large}\,. \tag{3.33}$$

Define $v_n = \tilde{v}_n|_\Omega$ for n sufficiently large. Then (3.32) and (3.33) imply

$$\begin{cases} v_n \in \mathscr{D}(\Omega), \\ \lim\limits_{n\to\infty} v_n = v \textit{ strongly in} V. \end{cases} \tag{3.34}$$

From (3.31), $\rho_n \geq 0$, $\int\limits_{\mathbb{R}^2} \rho_n dy = 1$ and $|\nabla\tilde{v}(y)| \leq$ a.e. on $\mathbb{R}^2$, we obtain

$$|\nabla v_n(x)| = |\nabla\tilde{v}_n(x)| \leq \int_{\mathbb{R}^2} |\nabla\tilde{v}(y)|.\rho_n(x-y)dy \leq 1\ \forall x \in \Omega, \tag{3.35}$$

which completes the proof of the Lemma.

3.6.2 A convergence theorem

THEOREM 3.3. *Suppose that the angles of the triangles of* $\mathscr{C}_h$ *are uniformly bounded by* $\theta_0 > 0$ *as* $h \to 0$. *Then*

$$\lim_{h\to 0} u_h = u \text{ strongly in } V \cap C^0(\overline{\Omega}), \tag{3.36}$$

51 *where* u *and* u_h *are respectively the solutions of* (3.1) *and* (3.22).

Proof. To prove the strong convergence in V, we use Theorem 5.2 of Chap. 1, Sec. 5. To do this one has to verify the following properties

(i) If $(v_h)_h$, $v_h \in K_h\ \forall h$, convergence *weakly* to v then $v \in K$.

(ii) There exists χ and r_h with the following properties:

(1) $\overline{\chi} = K$,

(2) $r_h : \chi \to K_h\ \forall h$

(3) For each $v \in \chi$ we can find $h_0 = h_0(v)$ such that for all $h \leq h_0(v)$, $r_h v \in K_h$ and $\lim\limits_{h\to 0} r_h v = v$ strongly in V.

□

Verification of (i). Since $K_h \subset K$ and K is weakly closed (i) is obvious.

Verification of (ii). Let us define χ by

$$\chi = \{v \in \mathscr{D}(\Omega) : |\nabla v(x)| < 1 \; \forall x \in \Omega\}.$$

Then by Lemma 3.1 and from $\lim_{\lambda \to 1} = v$ *strongly in* V, $\forall v \in V$, it follows that $\overline{\chi} = K$.

Define $r_h : V \cap C^0(\overline{\Omega}) \to V_h$ by

$$\begin{cases} r_h v \in V_h \; \forall v \in V \cap C^0(\overline{\Omega}), \\ (r_h v)(P) = v(P) \; \forall P \in \overset{0}{\sum_h}. \end{cases} \tag{3.37}$$

Then $r_h v$ is the "linear" interpolate of v on $\mathscr{C}_h$. From the assumption on $\mathscr{C}_h$ we have (cf. STRANG-FIX [1], CIARLET [1], [2])

$$|\nabla(r_h v - v)| \leq Ch \parallel v \parallel_{W^{2,\infty}(\Omega)} \text{ a.e. } v \in \mathscr{D}(\Omega), \tag{3.38}$$

with C independent of h and v.

This implies

$$\lim_{h \to 0} \parallel r_h v - v \parallel_v = 0 \; \forall v \in \chi. \tag{3.39}$$

$$|\nabla r_h v(x)| \leq |\nabla v(x)| + Ch \parallel v \parallel_{W^{2,\infty}(\Omega)} \textit{ a.e.} \tag{3.40}$$

52

Since $v \in \chi$ it follows from (3.40) that we have $|\nabla r_h v(x)| < 1$ a.e. for $h < h_0(v)$.

This implies $r_h v \in K_h$.

This completes the Verification of (ii)′ and hence by Theorem 5.2 of Chap. 1, we have the *strong convergence* of u_h to u in V.

The strong convergence of u_h to u in the L^∞- norm follows from the convergence in V and from the compactness of K in $C^0(\overline{\Omega})$ (see Proposition 3.1).

3.7 Error estimates

From now on we assume that $f \in L^p$ for some $p \geq 2$.

In Sec. 3.7.1 we consider a one-dimensional problem (3.1). In this case if $f \in L^2(\Omega)$ we derive an $o(h)$ error estimate in the V-norm. In Sec. 3.7.2 we consider a two-dimensional case with $f \in L^p$, $p > 2$ and Ω convex, then we derive an $0(h^{1/2-1/p})$ error estimate in the V-norm.

3.7.1 One-dimensional case

We assume here $\Omega = \{x \in \mathbb{R} : 0 < x < 1\}$ and that $f \in L^2(\Omega)$. Then problem (3.1) can be written

$$\begin{cases} \int_0^1 \frac{du}{dx}\left(\frac{dv}{dx} - \frac{du}{dx}\right)dx \geq \int_0^1 f(v-u)dx \;\forall v \in K, \\ \\ u \in K\{v \in V : \left|\frac{dv}{dx}\right| \leq 1 \;\; a.e. \; in \quad \Omega\}. \end{cases} \tag{3.41}$$

Let N be a positive integer and $h = \dfrac{1}{N}$. Let $x_i = ih$ for $i = 0, 1, \ldots, N$ and

$$e_i = [x_{i-1}, x_i], i = 1, 2, \ldots, N.$$

Let $V_h = \{v_h \in C^0(\overline{\Omega}) : v_h(0) = v_h = 0, v_h|_{e_i} \in P_1, i = 1, 2, \ldots, N\}$,

$$K_h = K \cap V_h = \{v_h \in V_h : |v_h(x_i) - v_h(x_{i-1})|h \text{ for } i = 1, 2, \ldots, N\}$$

The approximate problem is defined by

$$\begin{cases} \int_0^1 \frac{du_h}{dx}\left(\frac{dv_h}{dx} - \frac{du_h}{dx}\right)dx \geq \int_0^1 f(v_h - u_h)dx \;\forall v_h \in K_h, \\ u_h \in K_h. \end{cases} \tag{3.42}$$

53 Obviously this problem has a unique solution. Now we are going to prove

THEOREM 3.4. *Let u and u_h be the respective solutions of* (3.41) *and* (3.42). *If $f \in L^2(\Omega)$ then we have*

$$\| u_h - u \|_v = 0(h).$$

Proof. Since $u_h \in K_h \subset K$ we have from (3.41)

$$a(u, u_h - u) \geq \int_0^1 f(u_h - u)dx. \tag{3.43}$$

Adding (3.42) and (3.43) we obtain

$$a(u_h - u, u_h - u) \leq a(v_h - u, u_h - u) + a(u, v_h - u) - \int_0^1 f(v_h - u)dx \forall v_h \in K_h$$

which in turn implies

$$\frac{1}{2} \| u_h - u \|_V^2 \leq \frac{1}{2} \| v_h - u \|_V^2 + \int_0^1 \frac{du}{dx}\left(\frac{dv_h}{dx} - \frac{du}{dx}\right) - \int_0^1 f(v_h - u)dx \forall v_h \in K_h : \tag{3.44}$$

since $u \in K \cap H^2(0, 1)$ we get

$$\int_0^1 \frac{du}{dx}\frac{d}{dx}(v_h - u)dx = \int_0^1 \left(-\frac{d^2u}{dx^2}\right)(v_h - u)dx \leq \| \frac{d^2u}{dx^2} \|_{L^2} \| v_h - u \|_{L^2} .$$

But we have

$$\| \frac{d^2u}{dx^2} \|_{L^2} \leq \| f \|_{L^2} . \tag{3.45}$$

Therefore (3.44) becomes

$$\frac{1}{2} \| u_h - u \|_v^2 \leq \frac{1}{2} \| v_h - u \|_v^2 + 2 \| f \|_{L^2} \| v_h - u \|_{L^2} \quad \forall v_h \in K_h. \tag{3.46}$$

□

Let $v \in K$. Then the usual linear interpolate $r_h v$ is defined by

$$\begin{cases} r_h v \in V_h, \\ (r_h v)(x_i) = v(x_i) \; i = 0, 1, \ldots, N \end{cases} \tag{3.47}$$

we have **54**

$$\frac{d}{dx}(r_h v)|_{e_i} = \frac{v(x_i) - v(x_{i-1})}{h}$$

$$= \frac{1}{h}\int_{x_{i-1}}^{x_i} \frac{dv}{dx}dx.$$

Hence we obtain

$$|\frac{d}{dx}(r_h v)|_{e_i} \leq 1 \text{ since } |\frac{dv}{dx}| \leq 1 \text{ a.e. in } \Omega. \tag{3.48}$$

Thus $r_h v \in K_h$.

Let us replace v_h by $r_h u$ in (3.46). Then

$$\frac{1}{2} \| u_h - u \|_v^2 \leq \frac{1}{2} \| r_h u - u \|_v^2 + 2 \| f \|_{L^2(\Omega)} \| r_h u - u \|_{L^2(\Omega)} . \tag{3.49}$$

Since

$$\| r_h u - u \|_V \leq Ch \| u \|_{H^2(\Omega)} \leq Ch \| f \|_{L^2(\Omega)}, \tag{3.50}$$

$$\| r_h - u - u \|_{L^2(\Omega)} \leq Ch^2 \| u \|_{H^2(\Omega)} \leq Ch^2 \| f \|_{L^2(\Omega)} \tag{3.51}$$

where C denotes constants independent of u and h; combining (3.49)–(3.51) we get

$$\| u_h - u \|_V = 0(h).$$

This proves the result.

Exercise 3.5. *:Prove* (3.45).

3.7.2 Two-dimensional case

55 We shall assume in this subsection that Ω is a convex, bounded, polygonal domain in $\mathbb{R}^2$ and that $f \in L^p(\Omega)$ with $p > 2$. The last assumption is quite reasonable since in practical applications in mechanics we have f = constant.

THEOREM 3.5. *Suppose that the angles of $\mathscr{C}_h$ are uniformly bounded by $\theta_0 > 0$ as $h \to 0$, the with the above assumptions on Ω and f we have*

$$\| u_h - u \|_v = 0(h^{1/2-1/p}),$$

where u and u_h are respectively the solutions of (3.1) *and* (3.22).

Proof. : Since $f \in L^p(\Omega)$ with $p > 2$ and Ω is bounded, from Theorem 3.2 of this chapter we have

$$u \in W^{2,p}(\Omega).$$

Then as in proof of Theorem 3.4 and using $K_h \subset K$ we obtain

$$\begin{cases} \frac{1}{2} \parallel u_h - u \parallel_v^2 \leq \frac{1}{2} \parallel v_h - u \parallel_v^2 + a(u, v_h - u) - \int_\Omega f(v_h - u)dx, \\ \leq \frac{1}{2} \parallel v_h - u \parallel_v^2 - \int_\Omega (-\Delta u - f)(v_h - u)dx \; \forall v_h \in K_h. \end{cases} \tag{3.52}$$

Then using Holder's inequality it follows from (3.52) that

$$\begin{cases} \frac{1}{2} \parallel u_h - u \parallel_v^2 \leq \frac{1}{2} \parallel v_h - u \parallel_V^2 + \{\parallel \Delta u \parallel_{L^p(\Omega)} + \parallel f \parallel_{L^p}(\Omega)\} \\ \qquad\qquad \parallel v_h u \parallel_{L^{p'}(\Omega)} \; \forall v_h \in K_h \\ \text{with } \frac{1}{p} + \frac{1}{p'} = 1. \end{cases} \tag{3.53}$$

□

Let $1 \leq q \leq \infty$. Assume $\mathscr{C}_h$ satisfies the hypothesis of Theorem 3.5 and that $p > 2$. If $W^{2,p}(T) \subset W^{1,q}(T)$ it follows from CLARLET [2] and the *Sobolev imbedding Theorem* ($W^{2,p}(T) \subset W^{1,\infty}(T) \subset C^0(T)$) that $\forall T \in \mathscr{C}_h$ and $\forall v \in W^{2,p}(T)$ we have

$$\parallel \nabla(v - \pi_T v) \parallel_{L^q(T) \times L^q(T)} \leq C h_T^{1+2(\frac{1}{q} - \frac{1}{p})} \parallel v \parallel_{W^{2,p}(T)} \tag{3.54}$$

In (3.54) $\pi_T v$ is the linear interpolate of v at the three vertices of T, **56** h_T is the diameter of T and C is a constant independent of T and v.

Let $v \in W^{2,p}(\Omega)$ and let $\pi_h : V \cap C^0(\overline{\Omega}) \to V_h$ be defined by

$$\begin{cases} \pi_h v \in V_h & \forall v \in H_0^1(\Omega) \cap C^0(\overline{\Omega}), \\ (\pi_h v)(P) = & v(P) \; \forall P \in \sum_h^0. \end{cases}$$

Since $p > 2$ implies $W^{2,p}(\Omega) \subset C^0(\overline{\Omega})$, one may define $\pi_h v$, but *unlike the one dimensional case, usually*

$$\pi_h v \notin K_h \text{ for } v \in W^{2,p}(\Omega) \cap K.$$

Since $W^{2,p}(\Omega) \subset W^{1,\infty}(\Omega)$ for $p > 2$, it follows from (3.54) that a.e.

$$|\nabla(\pi_h v - v)(x)| \leq r h^{1-\frac{2}{p}} \parallel v \parallel_{W^{2,p}(\Omega)} \ \forall v \in W^{2,p}(\Omega)$$

which in turn implies that a.e.

$$|\nabla(\pi_h v)(x)| \leq 1 + r h^{1-\frac{2}{p}} \parallel v \parallel_{W^{2,p}(\Omega)}, \ \forall v \in K \cap W^{2,p}(\Omega). \tag{3.55}$$

The constant r occurring in (3.55) is independent of v and h. Let us define

$$\begin{cases} r_h : V \cap W^{2,p}(\Omega) \to V_h \text{ by} \\ r_h v = \dfrac{\pi_h v}{1 + r h^{1-2/p} \parallel v \parallel_{W^{2,p}(\Omega)}}. \end{cases} \tag{3.56}$$

It follows from (3.55) and (3.56) that

$$r_h v \in K_h \ \forall v \in W^{2,p}(\Omega) \cap K. \tag{3.57}$$

Since $u \in W^{2,p}(\Omega) \cap K$, It follows from (3.57) that we take $= r_h u$ in (3.53) so that

$$\frac{1}{2} \parallel u_h - u \parallel_V^2 \leq \frac{1}{2} \parallel r_h u - u_h \parallel_V^2 + \{\parallel \Delta_u \parallel_{L^p} + \parallel f \parallel_{L^p}\} \parallel r_h u - u \parallel_{L^{p'}(\Omega)} . \tag{3.58}$$

57 We have

$$r_u - u = \frac{\pi_h u - u - r h^{1-2/p} \parallel u \parallel_{W^{2,p}(\Omega)} .u}{1 + r h^{1-2/p} \parallel u \parallel_{W^{2,p}(\Omega)}}$$

which implies

$$\parallel r_h u - u \parallel_V \leq \parallel \pi_h u - u \parallel_V + r h^{1-2/p} \parallel u \parallel_{W^{2,p}} \parallel u \parallel_V, \tag{3.59}$$

$$\parallel r_h - u \parallel_{L^{p'}(\Omega)} \leq \parallel \pi_h u - u \parallel_{L^{p'}(\Omega)} + r h^{1-2/p} \parallel u \parallel_{W^{2,p}(\Omega)} \parallel u \parallel_{L^{p'}} . \tag{3.60}$$

Since $p > 2$ we have $L^p(\Omega) \subset L^{p'}(\Omega)$ inclusion and from standard approximation results (see STRANG-FIX [1], CIARLET [1], [2]) it follows that under the above assumption on $\mathscr{C}_h$ we have

$$\parallel \pi_h u - u \parallel_V \leq Ch \parallel u \parallel_{W^{2,p}(\Omega)}, \tag{3.61}$$

$$\parallel \pi_h u - u \parallel_{L^{p'}(\Omega)} \leq Ch^2 \parallel u \parallel_{W^{2,p}(\Omega)}, \tag{3.62}$$

with C independent of h and u. Then the $0(h^{1/2-1/p})$ error estimate of the statement of Theorem 3.5 follows directly from (3.58)–(3.62).

REMARK 3.9. *It follows from Theorem 3.5 that if f = constant (which correspond to application in mechanics) and if Ω is a convex polygonal domain, we have "practically" an* $0(\sqrt{h})$ *error estimate.*

REMARK 3.10. *One may find in FALK[1] an analysis of the error estimate for piecewise linear approximations of* (3.1) *when Ω is not polygonal.*

REMARK 3.11. *We may find in FALK-MERCIER [1] a different piecewise linear approximation of* (3.1).*Under appropriate assumptions this approximation leads to an* $0(h)$ *error estimate for* $\| u_h - u \|_v$. *However this approximation seems less suitable for computations than the ap-* **58** *proximations we have studied in this section (see also G.L.T. [1, chap. 3]).*

3.8 A dual iterative method for solving (3.1) and (3.2)

There are several iterative methods for solving (3.1), (3.22) and the reader who is interested in this direction of the problem may consult G.L.T [1, Chap. 3] (see also CEA-GLOWINSKI-NEDELEC [1]). In this section we shall use the material of CEA[2, Chap, 5, Section 5] to describe an algorithm of Uzawa type which has been successfully used to solve the elasto-plastic torsion problem. Another method will be described in Chap. 5, Sec. 6.2.

3.8.1 The Continuous case

Following CEA [2] we observe that K can also be written as

$$K = \{v \in V : |\nabla v|^2 - | \leq 0 \text{ a.e. } \}.$$

Hence it is quite natural to associate to 3.1 the following Lagrangian functional $\mathscr{L}$ defined on $H_0^1(\Omega) \times L^\infty(\Omega)$ by

$$\mathscr{L}(v, \mu) = \frac{1}{2} \int_\Omega |\nabla v|^2 dx - < f, v > + \frac{1}{2} \int_\Omega \mu(|\nabla v|^2 - 1) dx.$$

It follows from CEA [2] that if $\mathscr{L}$ has a saddle point $\{u, \lambda\} \in H_0^1(\Omega) \times L_+^\infty(\Omega) (L_+^\infty(\Omega) = \{q \in L^\infty(\Omega) : q \leq 0 \text{ a.e.}\})$ then u is a solution of (3.1).

Thus λ appears as an infinite dimensional multiplier (of F. John-Kuhn-Tucker type) for (3.1). The existence of such a multiplier in L^∞_+ has been probed by H. BREZIS [2] in the *physical case* (i.e. f_∞ = constant), but in more general cases the existence of such a multiplier in $L^\infty_+(\Omega)$ is still an *open problem*.

Following CEA [2], it is then natural to use, for solving (3.1), a *saddle point solver* like the following algorithm of Uzawa's type:

$$\lambda^0 \in L^\infty_+(\dot{\Omega}), \text{ arbitrarily given (for example } \lambda^0 = 0), \tag{3.63}$$

then by induction assuming λ^n known we obtain u^n and λ^{n+1} by

$$\begin{cases} \mathscr{L}(u^n, \lambda^n) \leq \mathscr{L}(v, \lambda^n) \; \forall v \in H^1_0(\Omega), \\ u^n \in H^1 +_0 (\Omega). \end{cases} \tag{3.64}$$

59

$$\lambda^{n+1} = [\lambda^n + \rho(|\nabla u^n|^2 - 1)]^+ \text{ with } \rho > 0. \tag{3.65}$$

Let us analyse (3.64) in detail; actually (3.64) is a linear Dirichlet problem, the explicit form of which is given (in the divergence form) by

$$\begin{cases} -\nabla.((1 + \lambda^n)\nabla u^n) = f \text{in } \Omega \\ u^n|_\Gamma = 0. \end{cases} \tag{3.66}$$

The problem (3.66) has a unique solution in $H^1_0(\Omega)$ whenever $\lambda^n \in L^\infty_+(\Omega)$. Since we do not know in general about the existence of a multiplier in $L^\infty_+(\Omega)$, the above algorithm in general is purely formal.

3.8.2 The discrete case

In this section we shall follow G.L.T. [1, Chap.3, Sec. 9.2]. Define V_{h_∞} and K_h as in section 3.5.1 of this Chapter. Define L_h (approximation of $L(\Omega)$) and Λ_h (approximation of L^∞_+) by

$$L_h = \{\mu \in L^\infty(\Omega) : \mu = \sum_{T \in \mathscr{C}_h} \mu_T \chi_T, \mu_T \in \mathbb{R}\},$$

and where χ_T is the *characteristic function* of T, and

$$\Lambda_h = \{\mu \in L_h : \mu \geq 0 \text{ a.e. in } \Omega\}.$$

Clearly it follows that for $v_h \in V_h$, $\nabla v_h \in L_h \times L_h$ and for $v_h \in K_h$, $1 - |\nabla v_h|^2 \in \Lambda_h$.

Define the Lagrangian $\mathscr{L}$ on $V_h \times L_h$ as section 3.8.1, then we have

Proposition 3.5. *The Lagrangian $\mathscr{L}$ has a saddle point $\{u_h, \lambda_h\}$ in $V_h \times \Lambda_h$ with*

$$u_n \text{ is solution of } (3.22), \tag{3.67}$$

$$\lambda_h(|\nabla u_h|^2 - 1) = 0. \tag{3.68}$$

Proof. Since V_h and L_h are finite dimensional (3.67) and (3.68) will fol- **60**
low from CEA [2. Chap. 5] (cf. also ROCKAFELLAR [1, Chap. 28]), if we can prove that there exists an element of V_h in the neighbourhood of which the constraints are strictly satisfied. Let us show that there exists a neighbourhood N_h of zero in V_h such that $\forall v_h \in N_h$, $|\nabla v_{h_2}|^2 - 1 < 0$. In order to show this, observe that the functional given by $v_h \to |\nabla v_h|^2 - 1$ is C^∞ and at zero it is equal to -1, Hence the assertion follows. □

To conclude this Section 3, let us describe an algorithm of Uzawa's type which is the discrete version of (3.63)–(3.65)

$$\lambda_h^0 \in \Lambda_h \text{ arbitrarily chosen (for instance } \lambda_h^0 = 0), \tag{3.69}$$

then by induction once λ_h^n is known, we obtain u_h^n and λ_h^{n+1} by

$$\begin{cases} \mathscr{L}(u_h^n, \lambda_h^n) \le \mathscr{L}(v_h, \lambda_h^n) \forall v_h \in V_h, \\ u_h^n \in V_h, \end{cases} \tag{3.70}$$

$$\lambda_h^{n+1} = [\lambda_h^n + \rho(|\nabla u_h^n|^2 - 1)]^+ \text{ with } \rho > 0. \tag{3.71}$$

We observe that if λ_h^n is known then u_h^n is the unique solution of the following approximate Dirichlet problem (given in variational form)

$$\begin{cases} \int_\Omega (1 + \lambda_h^n) \nabla u_h^n \cdot \nabla v_h dx = < f, v_h > \ \forall v_h \in V_h, \\ u_h^n \in V_h. \end{cases} \tag{3.72}$$

It follows from CEA [2, Chap. 5] and G.L.T. [1, Chap. 2] that for $\rho > 0$ and sufficiently small we have $\lim_{n \to \infty} u_h^n = u_h$ where $u - h$ is the solution of (3.22).

REMARK 3.12. *The computations we have done seem to prove that the optimal choice for ρ, is almost independent of h for a given problem. Similarly the number of iterations of Uzawa's algorithm for a given problem is almost independent of h.*

4 A Third Example of EVI of The First Kind: A Simplified Signorini Problem

61 Most of the material in this section can be found in G.L.T. [1, Chap. 4]

4.1 The continuous problem

Existence and uniqueness results. As usual let Ω be a bounded domain of $\mathbb{R}^2$ with a smooth boundary Γ. We define

$$V = H^1(\Omega), \tag{4.1}$$

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v dx + \int_\Omega uv dx, \tag{4.2}$$

$$L(v) =< f, v >, f \in V^*, \tag{4.3}$$

$$K = \{v \in H^1(\Omega) : \gamma v \geq 0 \text{ a.e. on } \Gamma\}, \tag{4.4}$$

where γv denotes the *trace* of v on Γ. We have then the following

THEOREM 4.1. *The variational inequality*

$$\begin{cases} a(u, v-u) \geq L(v-u) \; \forall v \in K, \\ u \in K \end{cases} \quad \textit{has a unique solution.} \tag{4.5}$$

Proof. : Since the bilinear form $a(\cdot,\cdot)$ is the usual scalar product in $H^1(\Omega)$ and L is continuous, from Theorem 3.1 of Chapter 1 we get that (4.5) has a unique solution provided we show that K is a closed. convex, non-empty subset of V.

Since $0 \in K$ (actually $H_0^1(\Omega) \subset K$), K is non-empty. The convexity of K is obvious. If $(v_n)_n \subset K$ and $v_n \to v$ in $H^1(\Omega)$ then $\gamma v_n \to \gamma v$, since $\gamma : H^1(\Omega) \to L^2(\Gamma)$ is continuous. Since $v_n \in K$, $\gamma v_n \geq 0$ a.e. on Γ. Therefore $\gamma v \geq 0$ a.e. on Γ. Hence $v \in K$ which shows K is closed.

This proves the theorem. □

REMARK 4.1. *Since $a(\cdot,\cdot)$ is symmetric, the solution u of* (4.5) *is characterised (see Chap. 1, Sec. 3.2) as the unique solution of the minimisation problem*

$$\begin{cases} J(u) \leq J(v)\ \forall v \in K, \\ u \in K, \end{cases} \tag{4.7}$$

where $J(v) = \frac{1}{2}a(v,v) - L(v)$. **62**

REMARK 4.2. *Actually* (4.5) *or* (4.7) *is a simplified version of a problem, occurring in elasticity, called the Signorini problem for which we refer to DUVAUT-LIONS [1, Chap. 3] and to the references therein. We refer also to DUVAUT-LIONS, loc. cit., Chap. 1, Chap. 2 for other physical and mechanical interpretations of* (4.5) *and* (4.7).

REMARK 4.3. *Assuming that Ω is bounded (at least in one direction of $\mathbb{R}^2$) we consider*

$$\hat{V} = \{v \in H^1(\Omega); v = 0 \text{ a.e.on } \Gamma_0\}. \tag{4.8}$$

$$\hat{a}(u,v) = \int_\Omega \nabla u \cdot \nabla v dx, \tag{4.9}$$

$$\hat{L}(v) = < f, v > \text{ with } f \in (\hat{v})^*, \tag{4.10}$$

$$[\hat{K}] = \{v \in V : \gamma v \geq g \text{ a.e. on } \Gamma_1\}, \tag{4.11}$$

where Γ_0 and Γ_1 are "good" subsets of Γ such that $\Gamma_1 \cap \Gamma_0 = \phi, \Gamma = \Gamma_1 \cup \Gamma_0$ (see fig. 4.1)

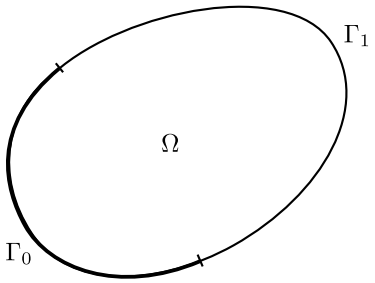

Figure 4.1:

63 Assuming that measure of Γ_0 is positive and that g is sufficiently smooth, it can be proved that the following variant of (4.5)

$$\begin{cases} \hat{a}(u, v-u) \geq \hat{L}(v-u) \, \forall v \in \hat{K}, \\ u \in \hat{K}, \end{cases} \tag{4.12}$$

has a unique solution.

In the proof of this result one uses the fact that $\hat{a}(v, v)$ defined a norm on V which is equivalent to the norm induced by $H^1(\Omega)$.

Exercise 4.1. *Prove that $a(v, v)$ defines a norm equivalent to the norm induced by $H^1(\Omega)$.*

4.2 Regularity of the solution

THEOREM 4.2. *(H BREZIS [3]) Let Ω be a bounded domain of $\mathbb{R}^2$ with a smooth boundary Γ (or Ω is a convex, polygonal domain). If $L(v) = \int_\Omega f v dx$ with $f \in L^2(\Omega)$ then the solution u of* (4.5) *is in $H^2(\Omega)$.*

4.3 Interpretation of (4.5) as a free boundary problem

Let us recall some definitions and results related to cones.

DEFINITION 2.1. *Let X be a vector space, $C \subset X$ and $x \in C$, then C is called a cone with vertex x if for all $y \in C$, $t \geq 0$ implies $x + t(y - x) \in C$.*

Lemma 4.1. *Let H be a real Hilbert space, $b(\cdot, \cdot)$ a bilinear form on $H \times H$; λ a linear form on H and C a convex cone contained in H with vertex at 0. Then every solution of*

$$\begin{cases} b(u, v-u) \geq \lambda(v-u) \, \forall v \in C, \\ u \in C \end{cases} \tag{4.13}$$

is a solution of

$$\begin{cases} b(u, v) \geq \lambda(v) \, \forall v \in C, \\ b(u, u) = \lambda(u), \\ u \in C, \end{cases} \tag{4.14}$$

64 *and conversely.*

Exercise 4.2. *Prove Lemma 4.1*

Proposition 4.1. *Assume that*

$$L(v) = \int_\Omega f v dx + \int_\Gamma g\gamma v d\Gamma \tag{4.15}$$

with f and g sufficiently smooth. Then the solution u of (4.5) *is characterised by*

$$\begin{cases} -\Delta u + u = f \text{ a.e. in } \Omega, \\ \gamma u \geq 0, \frac{\partial u}{\partial n} \geq \text{ a.e. on } \Gamma, \\ \gamma u(\frac{\partial u}{\partial n} - g) = 0 \text{ a.e. on } \Gamma. \end{cases} \tag{4.16}$$

Proof. (1) First we will prove that (4.5) implies (4.16)

Since K is a convex cone with vertex at 0 it follows from Lemma 4.1 that

$$a(u, v) \geq L(v) \, \forall v \in K, \tag{4.17}$$

$$a(u, u) = L(u). \tag{4.18}$$

Since $\mathscr{D}(\Omega) \subset K$ we have from (4.17) that

$$\int_\Omega \nabla u \cdot \nabla \phi dx + \int_\Omega u\phi dx = \int_\Omega f\phi dx \, \forall \phi \in \mathscr{D}(\Omega). \tag{4.19}$$

It follows from (4.19) that

$$-\Delta u + u = f \text{ a.e. in } \Omega. \tag{4.20}$$

□

Let $v \in K$. Multiplying (4.20) by v and using *Green's formula* it follows that

$$a(u, v) = \int_\Omega f v dx + \int_\Gamma \gamma v \frac{\partial u}{\partial n} d\Gamma \, \forall v \in K. \tag{4.21}$$

From (4.17) and (4.21) we obtain

$$\int_\Gamma \left(\frac{\partial u}{\partial n} - g \right) \gamma v d\Gamma \geq 0 \, \forall v \in K. \tag{4.22}$$

Since the cone γK is dense in $L^2_+(\Gamma) = \{v \in L^2(\Gamma) : v \geq 0 \text{ a.e. on } \Gamma\}$ it **65** follows from (4.22) that

$$\frac{\partial u}{\partial n} - g \geq 0 \text{ a.e. on } \Gamma. \tag{4.23}$$

Taking $v = u$ in (4.21) and using (4.18) we obtain

$$\int_\Gamma \gamma u \left(\frac{\partial u}{\partial n} - g \right) d\Gamma = 0. \tag{4.24}$$

Since $\gamma u \geq 0$ and using (4.23) we obtain $\gamma u \left(\frac{\partial u}{\partial n} - g \right) = 0$ a.e. on Γ. This shows that (4.5) implies (4.16).

(1) Let us show that (4.16) implies (4.5). Starting from (4.20) and using Green's formula one can easily prove (4.17) and (4.18). These two relations in turn imply, from Lemma 4.1, that u is the solution of (4.5).

REMARK 4.4. *Similar results may be proved for that variant* (4.12) *of* (4.5) *(see Remark* (4.3).

REMARK 4.5. *From the equivalent formulation* (4.16) *of* (4.5) *it appears that the solution u of* (4.5) *is the solution of a free boundary problem namely*

Find a sufficiently smooth function u and two subsets Γ_0 and Γ_+ such that

$$\Gamma_0 \cup \Gamma_+ = \Gamma, \Gamma_0 m \cap \Gamma_+ = \phi, \tag{4.25}$$

$$\begin{cases} -\Delta u + u = f \text{ in } \Omega, \\ \gamma u = 0 \text{ on } \Gamma_0, \frac{\partial u}{\partial n} \geq g \text{ on } \Gamma_0, \\ \gamma u > o \text{ on } \Gamma_+, \frac{\partial u}{\partial n} = g \text{ on } \Gamma_+. \end{cases} \tag{4.26}$$

4.4 Finite element approximation of (4.5)

66 We consider in this section the approximation of (4.5) by piecewise linear and piecewise quadratic finite elements. We assume that Ω of is a bounded polygonal domain of $\mathbb{R}^2$ and we consider a triangulation $\mathscr{C}_h$ of Ω obeying (2.21)–(2.23) (see Sec 2.5., Chap. 2) ; we use the notation of Sec. 2.5.1 and 3.6 of this chapter.

4.4.1 Approximation of V and K

The space $V = H^1(\Omega)$ may be approximated by the spaces V_h^k where

$$V_h^k = \{v_\epsilon \in C^0(\overline{\Omega}) : v_h|_T \in P_k, \forall T \in \mathscr{C}_h\}, k = 1, 2.$$

$$\begin{aligned} \text{Define } \gamma_h &= \{P \in \Sigma_h \cap \Gamma\} = \Sigma_h - \Sigma_h^0, \\ \gamma_h' &= \{P \in \Sigma_h' \cap \Gamma\} = \Sigma_h' - \Sigma_h'^0, \\ \gamma_h^k &= \begin{cases} \gamma_h \text{ if } k = 1 \\ \gamma_h \cup \gamma_h' \text{ if } k = 2. \end{cases} \end{aligned}$$

Then we approximate K by

$$K_h^k = \{v_h \in V_h^k : v_h(P) \geq 0 \, \forall P \in \gamma_h^k \text{ for } k = 1, 2\}.$$

We have then the obvious

Proposition 4.2. *For $k = 1, 2$ the K_h^k are closed, convex, non-empty subsets of V_h^k and $K_h^1 \subset K \, \forall h$.*

4.4.2 The approximate problem

For $k = 1, 2$ the approximate problems are defined by

$$(P_{1h}^k) \begin{cases} a(u_h^k, v_h - u_h^k) \geq L(v_h - u_h^k) \, \forall v_h \in K_h^k, \\ u_h^k \in K_h^k. \end{cases}$$

Then one can easily prove,

Proposition 4.3. *The problem $(P_{1h}^k)(k = 1, 2)$ has a unique solution.*

REMARK 4.6. *Since $a(\cdot, \cdot)$ is symmetric, (P_{1h}^k) is equivalent (See Chap. 1, Sec. 3.2) to the quadratic programming Problem* 67

$$\min_{v_h \in K_h^k} \left[\frac{1}{2} a(v_h, v_h) - L(v_h) \right].$$

REMARK 4.7. *Using the formula of Sec. 7 One may express* (4.5) *and the equivalent quadratic problem in a form more suitable for computation.*

4.5 Convergence results. (General case)

4.5.1 A density Lemma

To prove the convergence results of the following Sec. 4.5.2 we shall use the following

Lemma 4.2. *Under the above assumptions on* Ω *we have*

$$\overline{K \cap C^\infty(\overline{\Omega})} = K.$$

Proof. Since Γ is Lipschitz continuous we have (see NECAS [1])

$$H^1(\Omega) = \overline{C^\infty(\overline{\Omega})};$$

Using the standard decomposition $v = v^+ - v^-$ it follows from Corollary 2.1 that

$$v \in K \iff v^- \in H_0^1(\omega). \tag{4.27}$$

□

Since $\overline{\mathscr{D}\Omega)} = H_0^1(\Omega)$ in the $H^1(\Omega)-$ topology, it follows from (4.27) that we have only to prove

$$\overline{\hat{K} \cap C^\infty(\Omega)} = \hat{K}, \tag{2.1}$$

where $\hat{K} = \{v \in H^1(\Omega), v \geq 0$ a.e. in $\Omega\}$.

Since Γ is Lipschitz continuous, Ω has (see LIONS [2], NECAS [1]), the so-called 1-extension property which implies

$$\begin{cases} \forall v \in H^1(\Omega), \exists \tilde{v} \in H^1(\mathbb{R}^2) \text{ such that} \\ \tilde{v}|_\Omega = v. \end{cases} \tag{4.29}$$

68 Let $v \in K$ and let $\tilde{v} \in H^1(\mathbb{R}^2)$ be an extension of v obeying (4.29). It follows, from $v \geq o$ a.e. in Ω and Corollary 2.1, that $|\tilde{v}|$ is also an extension of v obeying (4.29). Therefore if $v \in \tilde{K}$, it has always an extension $\tilde{v} \geq 0$ a.e. obeying (4.29). Consider such a non-negative

extension $\tilde{v}$ and a mollifying sequence ρ_n (like in Lemma 2.4 of this Chap.). Define $\tilde{v}_n$ by

$$\tilde{v}_n = \tilde{v} * \rho_n. \tag{4.30}$$

we have

$$\begin{cases} \tilde{v}_n \in \mathscr{D}(\mathbb{R}^2), \\ \lim \tilde{v}_n = \tilde{v} \text{ strongly in } H^1(\mathbb{R}^2). \end{cases} \tag{4.31}$$

From $\rho_n \geq 0$ and $\tilde{v} \geq 0$ a.e. we obtain from (4.30) that

$$\tilde{v}_n(x) \geq 0 \; \forall x \in \mathbb{R}^2. \tag{4.32}$$

Define v_n by

$$v_n = \tilde{v}_n|_{\overline{\Omega}};$$

from (4.31) and (4.32) it follows that

$$v_n \in C^\infty(\overline{\Omega}), \lim_{n\to\infty} v_n = v \text{ strongly } in H^1(\Omega), v_n \geq 0 \text{ a.e. in } \Omega.$$

This proves the Lemma.

4.5.2 Convergence theorem

THEOREM 4.3. *Suppose that the angles of $\mathscr{C}_h$ are uniformly bounded below by $\theta_0 > 0$ as $h \to 0$, then*

$$\lim_{h\to 0} u_h^k = u \text{ strongly } in H^1(\Omega), \tag{4.33}$$

where u, u_h^k are respectively the solutions of (4.5) and (P_{1h}^k) for $k = 1, 2$.

Proof. To prove (4.33) we use Theorem 5.2 of Chap. 1. To do this we only have to verify that the following two properties hold:

(i) If $(v_h)_h$, $v_h \in K_h^k$, converges *weakly* to v then $v \in K$. **69**

(ii) There exist $\chi \subset K$ and $r_h^k : \chi \to K_h^k$ such that $\overline{\chi} = K$ and $\lim_{h\to 0} r_h^k v = v$ *strongly* in V, $\forall v \in \chi$.

□

Verification of (i). If $k = 1$, then (i) is trivially satisfied, since $K_h^1 \subset K$.

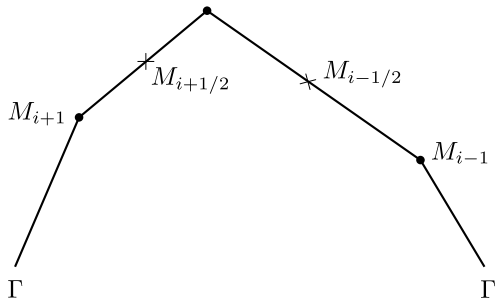

Figure 4.2:

If $k = 2$, using the notation of Fig. 4.2. we consider $\phi \in C^0(\Gamma)$, $\phi \geq 0$, and we define ϕ_h by

$$\phi_h = \Sigma_i \phi(M_{i+1/2}) \chi_{i+1/2} \tag{4.34}$$

where $\chi_{i+1/2}$ denotes the *characteristic function* of the open segment $]M_i, M_{i+1}[$. Then

$$\begin{cases} \phi_h \geq 0 \text{ a.e. on } \Gamma, \\ \lim\limits_{h \to 0} \| \phi_h - \phi \|_{L^\infty(\Gamma)} = 0. \end{cases} \tag{4.35}$$

Let us consider a sequence $(v_h)_h$, $v_h \in K_h^2 \; \forall h$, such that

$$\lim v_h = v \text{ weakly } in V. \tag{4.36}$$

70 It follows from (4.36) (see NECAS [1]) that $\lim\limits_{h \to 0} \gamma v_h = \gamma v$ *strongly* in $L^2(\Gamma)$. This implies in turn that

$$\lim_{h \to 0} \int_\Gamma \gamma v_h \phi_h d\Gamma = \int_\Gamma \gamma v \phi d\Gamma. \tag{4.37}$$

It follows from Simpson's rule that

$$\begin{cases} \int_\Gamma \gamma v_h \phi_h d\Gamma = \frac{1}{6}\Sigma_i |\overrightarrow{M_i M_{i+1}}| \phi(M_{i+1/2}) \\ \qquad [v_h(M_i) + 4v_h(M_{i+1/2}) + v_h(M_{i+1})] \geq 0 \\ \forall v_h \in K_h^2,\ \forall \phi \in C^0(\Gamma), \phi \geq 0. \end{cases} \tag{4.38}$$

We obtain from (4.37) and (4.38) that

$$\int_\Gamma \gamma v \phi d\Gamma \geq 0\ \forall \phi \in C^0(\Gamma), \phi \geq 0,$$

which implies $\gamma v \geq 0$ a.e. on Γ.

This proves (i).

Verification of (ii). From Lemma 4.2, it is natural to take $\chi = K \cap C^\infty(\overline{\Omega})$. Define $r_h^k; H^1(\Omega) \cap C^0(\overline{\Omega}) \to V_h^k$ by

$$\begin{cases} r_h^k v \in V_h^k\ \forall v \in H^1(\Omega) \cap C^0(\overline{\Omega}), \\ r_h^k v(P) = v(P)\ \forall P \in \Sigma_h^k, k = 1, 2. \end{cases} \tag{4.39}$$

On one hand, under the assumptions made on $\mathscr{C}_h$ we have (see STRANG-FIX [1])

$$\| r_h^k v - v \|_V \leq C h^k \| v \|_{H^{k+1}(\omega)}\ \ \forall v \in C^\infty(\overline{\Omega}), k = 1, 2. \tag{4.40}$$

with C independent of h and v.

This implies

$$\lim_{h \to 0} \| r_h^k v - v \|_v = 0\ \forall v \in \chi, k = 1, 2. \tag{4.41}$$

On the other hand it is obvious that $r_h^k v \in K_h^k\ \forall v \in K \cap C^0(\overline{\Omega})$, so that $r_h^k v \in K_h^k\ \forall v \in \chi, k = 1, 2$.

In conclusion, with the above χ and r_h^k, (ii) is satisfied.

REMARK 4.8. *For error estimates in the approximation of* (4.5) *by piecewise linear finite elements, it has been shown by BREZZI-HAGER-RAVIART [1] that we have* **71**

$$\| u_h - u \|_{H^1(\Omega)} = 0(h),$$

assuming reasonable smoothness hypothesis for u on Ω.

4.6 Iterative methods for solving the discrete problem

We shall briefly describe two types of methods which seem to be appropriate for solving the approximate problem of Sec. 4.4.

4.6.1 Solution by an over-relaxation method

The approximate problem (P_{1h}^k) are, for $k = 1, 2$, equivalent to the quadratic programming problems described in Remark 4.6. By virtue of the properties of K_h^k (see Sec. 4.4.1.) we can use, for the solution of (P_{1h}^k), the over-relaxation method with projection, which has already been used in Sec. 2.8 to solve the approximate obstacle problem and is described in CEA [2, Chap. 4]. From the properties of our problem the method will converge provided $0 < \omega < 2$.

4.6.2 Solution by a duality method

We first consider the *continuous case.* Let us define a Lagrangian $\mathscr{L}$ by

$$\mathscr{L}(v, q) = \frac{1}{2}a(v, v) - L(v) - \int_\Gamma q\gamma v d\Gamma. \tag{4.42}$$

and let λ be the *positive cone* of $L^2(\Gamma)$, i.e.

$$\Lambda = \{q \in L^2(\Gamma) : q \geq 0 \text{ a.e. on } \Gamma\}.$$

Then we have

THEOREM 4.4. *Let* $L(v) = \int_\Omega fvdx + \int_\Gamma g\gamma v d\Gamma$ *with* f *and* g *sufficiently smooth. Suppose that the solution* u *of* (4.5) *and* (4.7) *belongs to* $H^2(\Omega)$*; then* $\{u, \frac{\partial u}{\partial n} - g\}$ *is the unique saddle point of* $\mathscr{L}$ *over* $H^1(\Omega) \times \Lambda$.

72 *Proof.* We divide the proof into two parts. In the first part we will show that $\{u, \frac{\partial u}{\partial n} - g\}$ is a saddle point of $\mathscr{L}$ over $H^1(\Omega) \times \lambda$ and in the second part we will prove the uniqueness.

(1) Let $p = \frac{\partial u}{\partial n} - g$. From the definition of a saddle point we have to prove that

$$\begin{cases} \mathscr{L}(u, q) \leq \mathscr{L}(u, p) \leq \mathscr{L}(v, p) \forall \{v, q\} \in V \times \Lambda, \\ \{u, p\} \in V \times \Lambda. \end{cases} \tag{4.43}$$

□

Since $u \in H^2(\Omega)$ we have $\frac{\partial u}{\partial n} \in H^{1/2}(\Gamma) \subset L^2(\Gamma)$ (see LIONS - MAGENES [1]). Then if g is smooth enough we have $p = \frac{\partial u}{\partial n} - g \in L^2(\Gamma)$. From Proposition 4.1 we have

$$\begin{cases} p = \frac{\partial u}{\partial n} - g \geq 0 \text{ on } \Gamma, \\ p - \gamma u = 0 \text{ a.e. on } \Gamma. \end{cases} \tag{4.44}$$

This implies that we have $\{u, p\} \in H^1(\Omega) \times \Lambda$ and $\int_\Gamma p \cdot \gamma u d\Gamma = 0$. Since $\gamma u \geq 0$ on Γ we have

$$\int_\Gamma q \cdot \gamma u d\Gamma \geq 0 \, \forall q \in \Lambda. \tag{4.45}$$

It follows from (4.44) and (4.45) that

$$\begin{cases} \mathscr{L}(u, q) & = \frac{1}{2}a(u, u) = L(u) - \int_\Gamma q \cdot \gamma u d\Gamma \leq \frac{1}{2}a(u, u) - L(u) = \\ & = \frac{1}{2}a(u, u) - L(u) - \int_\Gamma p \cdot \gamma u d\Gamma = \mathscr{L}(u, p) \, \forall q \in \Lambda \end{cases}$$

which proves the first inequality of (4.43).

To prove the second inequality of (4.43) we observe that the solution u^* of the minimisation problem

$$\begin{cases} \mathscr{L}(u^*, p) \leq \mathscr{L}(v, p) \, \forall v \in H^1(\Omega), \\ u^* \in H^1(\Omega), \end{cases} \tag{4.46}$$

is unique and is actually the solution of the linear variational equation **73**

$$\begin{cases} a(u^*, v) = L(v) + \int_\Gamma p\gamma v d\Gamma \, \forall v \in H^1(\Omega), \\ u^* \in H^1(\Omega). \end{cases} \tag{4.47}$$

Since $L(v) = \int_\Omega f v dx + \int_\Gamma g\gamma v d\Gamma$, u^* is actually the solution of the Neumann problem

$$\begin{cases} -\Delta u^* + u^* = f \text{ in } \Omega, \\ \frac{\partial u^*}{\partial n} = p + g = \frac{\partial u}{\partial n} \text{ on } \Gamma, \end{cases} \tag{4.48}$$

Since from Proposition 4.1 we obviously have

$$\begin{cases} -\Delta u + u = f \text{ in } \Omega, \\ \frac{\partial u}{\partial n} = \frac{\partial u^*}{\partial n} \text{ on } \Gamma, \end{cases}$$

it follows from the uniqueness property of the Neumann problem (4.48) that $u = u^*$. Using (4.46) and $u = u^*$, we obtain the second inequality in (4.43). This proves that $\{u, p\}$ is a saddle point of $\mathscr{L}$ over $H^1(\Omega) \times \Lambda$.

Uniqueness. Let $\{u^*, p^*\}$ be a saddle point of $\mathscr{L}$ over $H^1(\Omega) \times \Lambda$. We will show that $u^* = u, p^* =$. It follows from (4.42) and (4.43) that

$$\int_\Gamma (p - q)\gamma u d\Gamma \leq 0 \, \forall q \in \Lambda. \tag{4.49}$$

We have similarly,

$$\int_\Gamma (p^*, q)\gamma u^* d\Gamma \leq 0 \, \forall q \in \Lambda. \tag{4.50}$$

Taking $q = p^*$ (respectively $q = p$) in (4.49) (respectively (4.50)) we obtain

$$\int_\Gamma (p^* - p)\gamma(u^* - u) d\Gamma \leq 0. \tag{4.51}$$

74 It follows from the second inequality of (4.43) that u is the solution of

$$\begin{cases} a(u, v) = L(v) + \int_\Gamma p \cdot \gamma v d\Gamma \, \forall v \in H^1(\Omega), \\ u \in H^1(\Omega). \end{cases} \tag{4.52}$$

ans similarly

$$\begin{cases} a(u^*, v) = L(v) + \int_\Gamma p^* \cdot \gamma v d\Gamma \, \forall v \in H^1(\Omega), \\ u^* \in H^1(\Omega). \end{cases} \tag{4.53}$$

Taking $v = u^* - u$ (respectively $v = u - u^*$) in (4.52) (respectively (4.53)) we obtain

$$a(u^* - u, u^* - u) = \int_\Gamma (p^* - p)\gamma(u^* - u) d\Gamma. \tag{4.54}$$

Using the V-ellipticity of $a(\cdot,\cdot)$ it follows then from (4.51)–(4.54) that $u^* = u$ and

$$\int_\Gamma (p^* - p)\gamma v d\Gamma = 0 \; \forall v \in H^1(\Omega),$$

which implies that $p^* = p$.

Hence $\{u, p\}$ is the unique saddle point of $\mathscr{L}$ over $H^1(\Omega) \times \Lambda$.

It follows from Theorem 4.4 that we can apply Uzawa's algorithm to solve (4.5) (see CEA [2, Chap. 5], G.L.T. [1, Chap.2], [2, Chap. 4, Sec 3.6]). In the present case this algorithm is written as follows:

$$p^0 \in \Lambda \text{ is arbitrarily chosen (for instance } p^0 = 0). \tag{4.55}$$

By induction, after knowing p^n we compute $\{u^n, p^{n+1}\}$ by

$$\mathscr{L}(u^n, p^n) \le \mathscr{L}(v, p^n) \; \forall v \in H^1(\Omega), u^n \in H^1(\Omega), \tag{4.56}$$

$$p^{n+1} = P_\Lambda(p^n - \rho\gamma u^n), \tag{4.57}$$

where P_Λ is *projection operator* from $L^2(\Gamma)$ to Λ in the $L^2(\Gamma)$ norm and $\rho > 0$. It follows from (4.56) that u^n is in fact the solution of the Neumann problem 75

$$\begin{cases} -\Delta u^n + u^n = f \text{ in } \Omega, \\ \frac{\partial u^n}{\partial n}|\Gamma = p^n + g. \end{cases} \tag{4.58}$$

The projection P_Λ is given by

$$P_\Lambda(q) = q^+ \; \forall q \in L^2(\Gamma). \tag{4.59}$$

Since $\gamma : H^1(\Omega) \to L^2(\Gamma)$ is a continuous linear map we have

$$\| \gamma v \|_{L^2(\Gamma)} \le \| \gamma \| \cdot \| v \|_{H^1(\Omega)} \; \forall v \in H^1(\Omega). \tag{4.60}$$

It follows then from CEA, G.L.T., loc. cit., that

$$\lim_{n\to\infty} u^n = u \text{ strongly in } H^1(\Omega), \tag{4.61}$$

where u is the solution of the problem (4.5) provided that $0 < \rho < \frac{2}{\|\gamma\|^2}$.

Let us give a direct proof for this convergence result. This proof will use the characterisation (4.5) given in Proposition 4.1 (even if $a(\cdot,\cdot)$ is not symmetric the same result follows). It will be convenient to take (4.56), (4.58) in the following equivalent form:

$$\begin{cases} a(u^n, v) = L(v) + \int_\Gamma p^n \gamma v d\Gamma \; \forall v \in H^1(\Omega), \\ u^n \in H^1(\Omega). \end{cases} \tag{4.62}$$

Let u be the solution of (4.5) and $p = \frac{\partial u}{\partial n} - g$. It follows from Proposition 4.1 that

$$\begin{cases} a(u, v) = L(v) + \int_\Gamma p\gamma v d\Gamma \; \forall \mathscr{L} \in H^1(\Omega), \\ u \in H^1(\Omega), \end{cases} \tag{4.63}$$

$$\int_\Gamma (q - p)\gamma u d\Gamma \geq 0 \; \forall q \in \Lambda, \, p \in \Lambda. \tag{4.64}$$

76 Relation (4.64) can also be written as

$$\int_\Gamma (q - p)(p - \rho\gamma u - p) d\Gamma \leq 0 \; \forall q \in \Lambda, \rho > 0.$$

which is classically equivalent to

$$p = P_\Lambda(p - \rho\gamma u). \tag{4.65}$$

Let consider:

$$\overline{u}^n = u^n - u, \overline{p}^n = p^n - p.$$

Since P_Λ is a contraction , we have from (4.57) and (4.65)

$$\| \overline{p}^{n+1} \|_{L^2(\Gamma)} \leq \| \overline{p}^n - \rho\gamma\overline{u}^n \|_{L^2(\Gamma)} . \tag{4.66}$$

It follows from (4.66) that

$$\| \overline{p}^n \|^2_{L^2(\Gamma)} \| \overline{p}^{n+1} \|_{L^2(\Gamma)} \geq 2\rho \int_\Gamma \gamma\overline{u}^n \overline{p}^n d\Gamma - \rho^2 \| \gamma\overline{u}^n \|^2_{L^2(\Gamma)} . \tag{4.67}$$

Taking $v = \overline{u}^n$ (4.62) and (4.63) we obtain

$$a(\overline{u}^n, \overline{u}^n) = \int_\Gamma \overline{p}^n \gamma\overline{u}^n d\Gamma. \tag{4.68}$$

It follows then from (4.67) and (4.68) that

$$\| \overline{p}^n \|^2_{L^2(\Gamma)} - \| \overline{p}^{n+1} \|^2_{L^2(\Gamma)} \geq \rho(2 - \rho \| \gamma \|^2) \| \overline{u}^n \|^2_{H^1(\Omega)} . \qquad (4.69)$$

If $0 < \rho < \frac{2}{\|\gamma\|^2}$ we observe that the sequence $\{\| \overline{p}^n \|^2_{L^2(\Gamma)}\}_n$ is decreasing and hence converges. Therefore we have

$$\lim_{n\to\infty} (\| \overline{p}^n \|^2_{L^2(\Gamma)} - \| \overline{p}^{n+1} \|^2_{L^2(\Gamma)}) = 0$$

so that

$$\lim_{n\to\infty} \| \overline{u}^n \|_{H^1(\Omega)} = 0.$$

Since $\overline{u}^n = u^n - u$, we have proved the convergence. 77

Similarly we can solve the approximate problem (P^k_{1h}), $k = 1, 2$, using the discrete version of algorithm (4.55)–(4.57). We shall limit ourselves to $k = 1$, since the extension here to $k = 2$ is almost trivial.

We use here the notations of Sec. 4.1. Assume that $\gamma_h = \Sigma_h - \Sigma^0_h$ has been ordered.

Let $\gamma_h = \{M_i\}_i$.

We approximate Λ and $\mathscr{L}$ by

$$\Lambda^1_h = \{q_h : q_h = \{q_i\}_i, q_i \geq 0\} \text{ and}$$

$$\begin{cases} \mathscr{L}^1_h(v_h, q_h) = & \frac{1}{2}a(v_h, v_h) - L(v_h) \\ & -\frac{1}{2}\Sigma_i |M_i M_{i+1}| [q_i v_h(M_i) + q_{i+1} v_h(M_{i+1})]. \end{cases} \qquad (2.2)$$

We can prove that $\mathscr{L}^1_h$ has a unique saddle point $\{u_h, p_h\}$ where p_h is a F. John-Kuhn-Tucker vector for (P^1_{1h}) over $V^1_h \times \Lambda^1_h$ and u_h is precisely the solution of (P^1_{1h}). The discrete analogue of (4.55)–(4.57) is then

$$p^0_h \in \Lambda^1_h. \qquad (4.71)$$

$$\begin{cases} \mathscr{L}^1_h(u^n_h, p^n_h) \leq \mathscr{L}^1_h(v_h, p^n_h) \forall v_h \in V^1_h, \\ u^n_h \in V^1_h. \end{cases} \qquad (4.72)$$

$$p^{n+1}_i = [p^n_i - \rho u^n_h(M_i)]^+ \ \forall i, \rho > 0. \qquad (4.73)$$

One can prove that if $0 < \rho < \beta, \beta$ small enough, then $\lim_{n\to+\infty} u_h^n = u_h$ where u_h is the solution of (P_{1h}^1). One may find in G.L.T. [2, Chap.4] numerical applications of the above iterative methods for piecewise linear and piecewise quadratic approximations of (4.5).

Exercise 4.3. *Extend the above considerations to* (P_{1h}^2).

5 An Example of EVI of The Second Kind: A Simplified Friction Problem

5.1 The continuous problem. Existence and Uniqueness results

78

Let Ω be a bounded domain of $\mathbb{R}^2$ with a smooth boundary $\Gamma = \partial\Omega$. Using the same notations as in Sec. 4 we define

$$V = h^1(\Omega), \tag{5.1}$$

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v dx + \int_\Omega uv dx. \tag{5.2}$$

$$L(v) =< f, v >, f \in V^*, \tag{5.3}$$

$$j(v) = g\int_\Gamma |\gamma v| d\Gamma, \text{ where } g > 0. \tag{5.4}$$

We have then the following

THEOREM 5.1. *The variational inequality*

$$\begin{cases} a(u, v-u) + j(v) - j(u) \geq L(v-u)\ \forall v \in V. \\ u \in V. \end{cases} \tag{5.5}$$

has a unique solution.

Proof. In order to apply Theorem 4.1 of Chap. 1, it is enough to verify that $j(\cdot)$ is convex, proper and l.s.c. Actually $j(\cdot)$ is a seminorm on

V. Therefore using Schwartz inequality in $L^2(\Gamma)$ and the fact that $\gamma \in \mathscr{L}(H^1(\Omega), L^2(\Gamma))$ we have

$$|j(u) - j(v)| \leq |j(u-v)| \leq g(meas \cdot \Gamma)^{1/2} \parallel \gamma(v-u) \parallel_{L^2(\Gamma)} \leq C \parallel u - v \parallel_V, \tag{5.6}$$

for some constant C. □

Hence $j(\cdot)$ is Lipschitz continuous on V, so that $J(\cdot)$ is l.s.c ; $j(\cdot)$ is obviously convex and proper. Hence the Theorem is proved.

REMARK 5.1. *If $g = 0$, it is easy to prove that* (5.5) *reduce to the variational equation*

$$\begin{cases} a(u,v) = L(v)\ \forall v \in V, \\ u \in V. \end{cases}$$

This is related to the variational formulation of the Neumann problem.

REMARK 5.2. *Since $a(\cdot,\cdot)$ is symmetric,. the solution u of a* (5.5) *is characterised, using Lemma 4.1 of Chap. 1, as the unique solution of the minimization problem* 79

$$\begin{cases} J(u) \leq J(v),\ \forall v \in V, \\ u \in V, \end{cases} \tag{5.7}$$

where $J(v) = \dfrac{1}{2}a(v,v) + j(v) - L(v)$.

REMARK 5.3. *The problem* (5.5) *(and* (5.7)*) is the simplified version of a* friction *problem occurring in* elasticity. *For this types of problems we refer to DUVAUTLIONS [1]and the bibliography therein.*

Exercise 5.1. *Let us denote by u_g the solution of* (5.5). *Then prove that*

$$\lim_{g \to +\infty} u_g = \hat{U} \text{ strongly in } H^1(\Omega),$$

where $\hat{u}$ is the unique solution of

$$\begin{cases} a(\hat{u}, v) = L(v)\ \forall v \in H_0^1(\omega), \\ \hat{u} \in H_0^1(\Omega). \end{cases}$$

5.2 Regularity of the solution

THEOREM 5.2. *(H. BREZIS [3]). If Ω is a bounded domain with a smooth boundary and if $L(v) = \int_\Omega fvdx$ with $f \in L^2(\Omega)$, then the solution u of* (5.5) *is in $H^2(\Omega)$.*

5.3 Existence of a multiplier

Let us define Λ by

$$\Lambda = \{\mu \in L^2(\Gamma) : |\mu(x)| \leq 1 \text{ a. e. in } \Gamma\}.$$

Then we have

THEOREM 5.3. *The solution u of* (5.5) *is characterised by the existence of λ such that*

$$\begin{cases} a(u,v) + g\int_\Gamma \lambda\gamma v d\Gamma = L(v)\ \forall v \in V, \\ u \in V, \end{cases} \tag{5.8}$$

$$\begin{cases} \lambda \in \Lambda, \\ \lambda\gamma u = |\gamma u| \text{ a. e. in } \Gamma. \end{cases} \tag{5.9}$$

80

Proof. We will prove first that (5.5) implies (5.8) and (5.9).

Taking $v = 0$ and $v = 2u$ in (5.5) we have

$$a(u,u) + j(u) = L(u). \tag{5.10}$$

It follows then from (5.5), (5.10) that

$$L(v) - a(u,v) \leq j(v)\ \forall v \in V,$$

which implies

$$|L(v) - a(u,v)| \leq j(v) = g\int_\Gamma |\gamma v| d\Gamma\ \forall v \in V. \tag{5.11}$$

□

We have $H^1(\Omega) = H_0^1(\Omega) \oplus [H_0^1(\Omega)]^\perp$ where $[H_0^1(\Omega)]^\perp$ is the orthogonal complement of $H_0^1(\Omega)$in $H^1(\Omega)$.

Since $\gamma : [H_0^1(\Omega)]^\perp \to H^{1/2}(\Gamma)$ is an isomorphism, it follows from (5.11) that

$$L(v) - a(u, v) = \ell(\gamma v) \, \forall v \in V, \tag{5.12}$$

where $\ell(\cdot)$ is a continuous linear functional on $H^{1/2}(\Gamma)$. It follows then from (5.11), (5.12) that

$$|\ell(\mu)| \leq g \parallel \mu \parallel_{L^1(\Gamma)} \ \forall \mu \in H^{1/2}(\Gamma). \tag{5.13}$$

Since $H^{1/2}(\Gamma) \subset L^1(\Gamma)$ it follows from (5.13) that, we can apply to $\ell(\cdot)$, the Hanh-Banach Theorem (see for instance YOSIDA [1]) to obtain the existence of $\lambda \in L^\infty(\Gamma)$, $|\lambda(x)| \leq 1$ a.e. in Γ such that

$$\ell(\mu) = g \int_\Gamma \lambda \mu d\Gamma \, \forall \mu \in H^{1/2}(\Gamma). \tag{5.14}$$

Therefore it follows from (5.12) and (5.14) that **81**

$$a(u, v) + g \int_\Gamma \lambda \gamma v d\Gamma = L(v) \, \forall v \in V,$$

which proves (5.8).

Taking $v = u$ in (5.8) we obtain

$$a(u, u) + g \int_\Gamma \lambda \gamma u d\Gamma = L(u).$$

Using (5.10) and the above equation we obtain

$$\int_\Gamma (|\gamma u| - \lambda \gamma u) d\Gamma = 0. \tag{5.15}$$

Since $|\lambda| \leq 1$ a.e. we have

$$|\gamma u| - \lambda \gamma u \geq 0 \text{ a.e.} \tag{5.16}$$

It follows from (5.15) and (5.16) that

$$|\gamma u| = \lambda \gamma u \ a.e.$$

This completes the proof of (5.8) and (5.9). Assuming (5.8) and (5.9) we will show that (5.5) holds.

Let $\{u, \lambda\}$ be a solution of (5.8), (5.9). It follows from (5.8) that

$$a(u, v-u) + g\int_\Gamma \lambda\gamma(v-u)d\Gamma = L(v-u)\,\forall v \in V,$$

which can also be written as

$$a(u, v-u) + g\int_\Gamma \lambda\gamma v d\Gamma - g\int_\Gamma \lambda\gamma u d\Gamma = L(v-u)\,\forall v \in V. \tag{5.17}$$

From (5.9) and (5.17) we obtain

$$a(u, v-u) + g\int_\Gamma \lambda\gamma v d\Gamma - g\int_\Gamma |\gamma u| d\Gamma = L(v-u)\,\forall v \in V. \tag{5.18}$$

But since $\lambda\gamma v \le |\gamma v|$ a.e. in Γ, it follows from (5.18) that

$$a(u, v-u) + j(v) - j(u) \ge L(v-u)\,\forall v \in V.$$

82 This proves the characterization.

REMARK 5.4. *Assuming that*

$$L(v) = \int_\Omega f_0 v dx + \int_\Gamma f_1\gamma v d\Gamma,$$

with f_0, f_1 *sufficiently smooth, we can express* (5.8) *by*

$$\begin{cases} -\Delta u + u = f_0 \text{ in } \Omega, \\ \frac{\partial u}{\partial n} + g\lambda = f_1 \text{ a. e. on } \Gamma. \end{cases} \tag{5.19}$$

It follows from (5.19) *that* λ *is unique.*

Exercise 5.2. *Prove that* λ *is unique* $\forall L \in V^*$.

5.4 Finite element approximation of (5.5)

Let Ω be a bounded domain of $\mathbb{R}^2$. The notation used here is mostly the same as in Sec. 4.4 of this Chapter.

5.4.1 Approximation of V

We use the piecewise linear and piecewise quadratic approximations of $V = H^1(\Omega)$ described in Section 4.4.1 of this chapter.

5.4.2 Approximation of $j(\cdot)$

We use the notation of Figure 4.2. Then we approximate $j(\cdot)$ by

$$j_h^1(v_h) = \frac{g}{2} \sum_i |\overrightarrow{M_i M_{i+1}}|(|\gamma v_h(M_i)| + |\gamma v_h)M_{i+1})|) \; \forall v_h \in V_h^1, \tag{5.20}$$

$$\begin{cases} j_h^2(v_h) & = \frac{g}{6} \sum_i |\overrightarrow{M_i M_{i+1}}|(|\gamma v_h(M_i)| + 4|\gamma v_h(M_{i+1/2}| \\ & \quad +|\gamma v_h(M_{i+1})|) \; \forall v_h \in V_h^2. \end{cases} \tag{5.21}$$

In (5.20) and (5.21) we have $M_i \in \gamma_h$ and $M_{i+1/2} \in \gamma_h'$.

REMARK 5.5. *Clearly* (5.20), (5.21) *are respectively obtained from $j(\cdot)$ by using Trapezoidal and Simpson's numerical integration formulae.*

5.4.3 The approximate problem

For $k = 1, 2$ the problem (5.5) is approximated by **83**

$$(P_{2h}^k) \begin{cases} a(u_n^k, v_h - u_h^k) + j_h^k(v_h) - j_h^k(u_h^k) \geq L(v_h - u_h^k) \; \forall v_h \in V_h, \\ u_h^k \in V_h^k. \end{cases}$$

Then,

Proposition 5.1. *The problem (P_{2h}^k) has a unique solution.*

REMARK 5.6. *Since $a(\cdot, \cdot)$ is symmetric, (P_{2h}^k) is equivalent to the non-linear programing problem*

$$\min_{v_h \in V_h^k} \left[\frac{1}{2} a(v_h, v_h) + j_h^k(v_h) - L(v_h) \right]. \tag{5.22}$$

REMARK 5.7. *Using* (5.20), (5.21) *and* (7.1)–(7.4) *of Section 7 of this chapter, we may express (P_{2h}^k) and* (5.22) *in a form more suitable for computations.*

5.5 Convergence results

THEOREM 5.4. *Suppose that the angles of $\mathscr{C}_h$ are uniformly bounded below by $\theta > 0$ as $h \to 0$, then*

$$\lim_{h\to 0} u_h^k = u \text{ strongly in } H^1(\Omega), \tag{5.23}$$

where u and u_h^k are respectively the solutions of (5.5) *and* (P_{2h}^k) *for* $k = 1, 2$.

Proof. To prove (5.23) it is enough to verify the following (see Theorem 6.3 of Chapter 1)

(i) There exists $U \subset V$, $\overline{U} = V$ and

$$r_h^k : U \to V_h^k \text{ such that}$$
$$\lim_{h\to 0} r_h^k v = v \text{ strongly in } V \; \forall v \in U,$$

(ii) If $V_h \to v$ *weakly* in V then

$$\liminf_{h\to 0} j_h^k(v_h) \geq j(v).$$

84 (iii) $\lim_{h\to 0} j_h^k(r_h^k v) = j(v) \; \forall v \in U$.

□

Verification of (i). Since Γ is Lipschitz continuous we have (see NECAS [1])

$$\overline{C^\infty(\overline{\Omega})} = H^1(\Omega). \tag{5.24}$$

Therefore it is natural to take $U = C^\infty(\overline{\Omega})$. Define r_h^k by (4.39) if Theorem 4.3, chap. 2; under the above assumption on $\mathscr{C}_h$ it follows from STRANG-FIX [1] that

$$\| r_h^k v - v \|_V \leq Ch^k \| v \|_{H^{k+1}(\Omega)} \quad \forall v \in V, \tag{5.25}$$

where C is a constant independent of h and v. This implies (i).

Verification of (ii).

(1) *Case* $k = 1$. We use again the notation of Figure 4. 2. Since the trace of v_h restricted to $[M_i, M_{i+1}]$ is affine it follows that

$$\begin{cases} \gamma v_h(M) = \frac{1}{|M_i M_{i+1}|}(|\overrightarrow{MM}_{i+1}|\gamma v_h(M_i) + |\overrightarrow{MM}_i|\gamma v_h(M_{i+1})), \\ \forall v_h \in V_h^1,\ \forall M \in [M_i, M_{i+1}]. \end{cases} \tag{5.26}$$

$$\text{Since } \frac{|\overrightarrow{MM}_{i+1}|}{|\overrightarrow{M_iM}_{i+1}|} + \frac{|\overrightarrow{MM}_i|}{|\overrightarrow{M_iM}_{i+1}|} = 1,$$

the convexity of $\xi \to |\xi|$ implies

$$\begin{cases} |\gamma v_h(M)| \le \frac{1}{|\overrightarrow{M_iM}_{i+1}|}(|\overrightarrow{MM}_{i+1}||\gamma v_h(M_i)| \\ \quad +|\overrightarrow{M_iM}|\,|\gamma v_h(M_{i+1})|)\,\forall v_h \in V_h^1,\ \forall M \in [M_i, M_{i+1}]. \end{cases} \tag{5.27}$$

Interesting (5.27) on $\widehat{M_iM}_{i+1}$ we obtain

$$\int_{\widehat{M_iM}_{i+1}} |\gamma v_h| d\Gamma \le \frac{|\overrightarrow{M_iM}_{i+1}|}{2}(|\gamma v_h(M_i)| + |\gamma v_h(M_{i+1})|)$$

which implies that $\forall v_h \in V_h^1$ we have

$$\begin{cases} j(v_h) = g\int_\Gamma |\gamma v_h| d\Gamma & = g\sum_i \int_{\widehat{M_iM}_{i+1}} |\gamma v_h| d\Gamma \le \\ & \le \frac{g}{2}\sum_i |\overrightarrow{M_iM}_{i+1}|(|\gamma v_h(M_i)| + |\gamma v_h(M_{i+1})|) \\ & = j_h^1(v_h). \end{cases}$$

Thus we have proved **85**

$$j(v_h) \le j_h^1(v_h)\,\forall v_h \in V_h^1. \tag{5.28}$$

Let $v_h \to v$ *weakly* in V. Then $\lim_{h\to 0} \gamma(v_h) = \gamma(v)$ *strongly* in $L^2(\Gamma)$, which implies

$$\lim_{h\to 0} j(v_h) = j(v). \tag{5.29}$$

It follows then from (5.28) and (5.29) that $\lim_{h\to 0} \inf j_h^1(v_h) \ge j(v)$, which proves (ii) if $k = 1$.

(2) *case* $k = 2$. Let us define $M_{i+1/6}, M_{i+5/6}$ by (see Figure 5.1)

$$\overrightarrow{M_iM_{i+1/6}} = \frac{1}{6}\overrightarrow{M_iM_{i+1}}, \overrightarrow{M_iM_{i+5/6}} = \frac{5}{6}\overrightarrow{M_iM_{i+1}}.$$

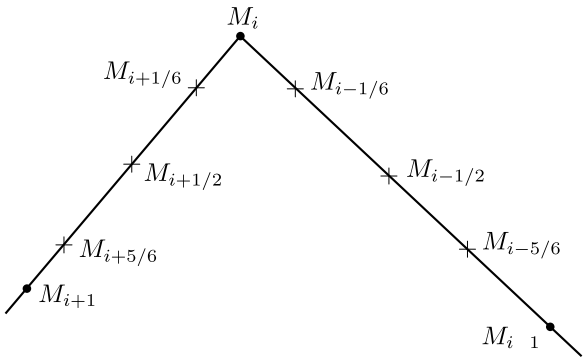

Figure 5.1:

86 Then we define $q_h : C^0(\Gamma) + L^\infty(\Gamma)$ by

$$\begin{cases} q_h(\mu) = \sum\limits_{M_i \in \gamma_h} \mu(m_i)X_i + \sum\limits_{M_{\frac{i+1}{2}} \in \gamma_h'}, \mu(M_{i+1/2})X_{i+1/2} \\ \forall \mu \in C^0(\Gamma). \end{cases} \tag{5.30}$$

where X_i (*respectively* $X_{i+1/2}$) is the *characteristic* function of $M_{i-6}M_iM_{i+1/6}$ (respectively $M_{i+1/6}\widehat{M_{i+5/6}}$). We have then the following obvious properties :

$$\lim_{h\to 0} q_h(\mu) = \mu \text{ strongly in } L^\infty(\Gamma)\ \ \forall \mu \in C^\circ(\Gamma), \tag{5.31}$$

$$j_h^2(v_h) = g\int_\Gamma |q_h\gamma v_h| d\Gamma = g \parallel q_h\gamma v_h \parallel_{L^1(\Gamma)} \quad \forall v_h \in V_h^2, \tag{5.32}$$

$$C_1 \parallel \gamma v_h \parallel_{L^2(\Gamma)} \leq \parallel q_h\gamma v_h \parallel_{L^2(\Gamma)} \leq C_2 \parallel \gamma v_h \parallel_{L^2(\Gamma)} \quad \forall v_h \in V_h^2, \tag{5.33}$$

where in (5.33), C_1 and C_2 are positive constants independent of v_h, h and Γ (values for C_1 and C_2 may be found in G. L. T. [2, Chap. 4]).

Now taking $\mu \in C^0(\Gamma)$, we define $s_h(\mu)$ by

$$\begin{cases} s_h(\mu) \in L^\infty(\Omega), \\ s_h(\mu)|_{]M_i, M_{i+1}[} = \mu(M_{i+1/2}). \end{cases} \tag{5.34}$$

Then

$$\lim_{h\to 0} s_h(\mu) = \mu \text{ strongly in } L^\infty(\gamma), \tag{5.35}$$

and from Simpson's integration formula we have

$$\int_\Gamma s_h(\mu) q_h \gamma_h v_h d\Gamma = \int_\Gamma s_h(\mu) \gamma v_h d\Gamma \quad \forall \mu \in C^\circ(\Gamma), \forall v_h \in V_h^2. \tag{5.36}$$

Let $v_h \to v$ weakly in V, $v_h \in V_h^2 \, \forall h$, then

$$\lim_{h\to 0} \gamma v_h = \gamma v \text{ strongly in } L^2(\Gamma). \tag{5.37}$$

On the one hand it follows from (5.33) that **87**

$$\| q_h \gamma v_h \|_{L^2(\Gamma)} \leq C, \tag{5.38}$$

where C is independent of h.

On the other hand (5.31), (5.35)–(5.38) imply that

$$\lim_{h\to 0} \int_\Gamma s_h(\mu) q_h \gamma v_h d\Gamma = \int_\Gamma \mu \gamma v d\Gamma \quad \forall \mu \in C^0(\Gamma). \tag{5.39}$$

In turn (5.38) and (5.39) imply that

$$\lim_{h\to 0} q_h \gamma v_h = v \text{ weakly in } L^2(\Gamma). \tag{5.40}$$

Since the functional $\mu \to \| \mu \|_{L^1(\Gamma)}$ is convex and continuous on $L^1(\Gamma)$ it follows from (5.40) that

$$\liminf_{h\to 0} \| q_h \gamma v_h \|_{L^1(\Gamma)} \geq \| \gamma v \|_{L^1(\Gamma)} \tag{5.41}$$

Combining (5.41) with (5.32) we obtain $\liminf_{h\to 0} j_h^2(v_h) \geq j(v)$ which proves (ii) when $k = 2$.

Verification of (iii). Let $v \in U = C^\infty(\overline{\Omega})$. From (5.25) and from the uniform continuity of γv on Γ it follows almost immediately that

$$\lim_{h\to 0} j_h^k(r_h^k v_h) = j(v), k = 1, 2.$$

Since the condition (i), (ii) and (iii) are satisfied the strong convergence of u_h to u follows from the Theorem 6.2 of Chap. 1.

REMARK 5.8. *It is proved in G. L. T. [2, Chap. 4] that* $v_h \to v$ weakly *in* $v, v/2 \in V_h^k$, *implies* $\lim_{h\to 0} j_h^k(v_h) = j(v), k = 1, 2.$

Since the proof of this result is rather technical we have used in these notes a simpler approach from which it follows that

$$\liminf_{h\to 0} j_h^k(v_h) \geq j(v), k = 1, 2.$$

As we have seen before this result was sufficient for proving Theorem 5.4.

5.6 Iterative methods for solving (P_{2h}^k).

88 In this section we briefly describe some iterative methods which may be useful for solving the approximate problems (P_{2h}^k).

5.6.1 Solution of (P_{2h}^k) by relaxation methods

It follows from (5.20)–(5.22) (see Remark 5.6) that $(P_{2h}^k), k = 1, 2$, are particular cases of

$$\min_{v\in\mathbb{R}^N} f(v), \tag{5.42}$$

where, with $v = (v_1, \ldots, v_n)$,

$$f(v) = \frac{1}{2}(Av, v) - (b, v) + \sum_{i=1}^{N} \alpha_i |v_i|. \tag{5.43}$$

In (5.43), $(\cdot, \cdot)$ denotes the usual inner product of $\mathbb{R}^N$, A is a $N \times N$ symmetric positive definite matrix and $\alpha_i \geq 0 \ \forall i = 1, \ldots, N$.

It follows then from CEA [2, Chap. 4], CEA-GLOWINSKI [1], G. L. T. [1, Chap, 2] that we can use a relaxation method for solving (5.42). Actually from the computation parameter ω, $\omega > 1$, will increase the speed of convergence.

Finally the algorithm we used is the following :

$$u^0 \text{ arbitrarily given in } \mathbb{R}^N, \tag{5.44}$$

then for $i = 1, 2, \ldots, N$,

$$f(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, u_i^{-n+1}, u_{i+1}^n, \ldots) \leq$$
$$f(u_1^{n+1}, \ldots, u_{i-1}^{n+1}, v_i, u_{i+1}^{n+1}, \ldots) \; \forall v_i \in \mathbb{R}, \tag{5.45}$$
$$u_i^{n+1} = u_i^n + \omega(u_i^{n+1} - u_i^n). \tag{5.46}$$

If $\omega = 1$, (5.44)–(5.46) reduces to a relaxation method. Numerical solutions of (5.5) using (5.44)–(5.46) are given in G. L. T. [2, Chap. 4].

REMARK 5.9. *If $\alpha_i > 0, u_i^{-n+1}$ is the solution of a* one variable, non differentiable minimization problem *which can be exactly computed by hand calculation.*

Exercise 5.3. *Express u_i^{-n+1} as a function of A, b, u^n, u^{n+1}.*

5.6.2 Solution of (P_{2h}^k by duality method)

We first examine the continuous case. Define a Lagrangien $\mathscr{L} : H^1(\Omega) \times$ **89**
$L^2(\Gamma) \to \mathbb{R}$ by

$$\mathscr{L}(v, \mu) = \frac{1}{2} a(v, v) - L(v) + g \int_\Gamma \mu \gamma v d\Gamma. \tag{5.47}$$

Then using the notation of Sec. 5.3 it follows from Theorem 5.3 that

THEOREM 5.5. *Let $\{u, \lambda\}$ be a solution of* (5.8), (5.9) *; then $\{u, \lambda\}$ is the unique saddle point of $\mathscr{L}$ over $H^1(\Omega) \times \Lambda$*

Exercise 5.4. *Prove Theorem 5.5.*

From Theorem 5.5 it follows that to solve (5.5) *we can use the following Uzawa's algorithm,*

$$\lambda^0 \in \Lambda \text{ arbitrarily chosen (for instance } \lambda^0 = 0), \tag{5.48}$$

then by induction, knowing λ^n *we compute* u^n *and* λ^{n+1} *by*

$$\begin{cases} \mathscr{L}(u^n, \lambda^n) \leq \mathscr{L}(v, \lambda^n) \, \forall v \in H^1(\Omega), \\ u^n \in V, \end{cases} \tag{5.49}$$

$$\left\{ \lambda^{n+1} = P_\Lambda(\lambda^n + \rho g \gamma u^n), \rho > 0. \right. \tag{5.50}$$

The minimization problem (5.49) is actually equivalent to the *Neumann variational problem*

$$\begin{cases} a(u^n, v) = L(v) - g \int_\Gamma \lambda^n \gamma v d\Gamma \, \forall v \in H^1(\Omega), \\ u^n \in H^1(\Omega). \end{cases} \tag{5.51}$$

In (5.50). P_Λ is the projection operator from $L^2(\Gamma)$ to Λ in the L^2- norm, then

$$P_\Lambda(\mu) = \sup(-1, \inf(1, \mu)) \ \forall \mu \in L^2(\Gamma). \tag{5.52}$$

Using CEA [2], G. L. T. [1, Chap. 2] it follows that for $0 < \rho < \frac{2}{g^2 \|\gamma\|^2}$. we have

$$\begin{cases} \lim_{n \to \infty} u^n = u \text{ strongly in } H^1(\Omega), \\ u \text{ solution of (5.5), (5.7).} \end{cases}$$

90

Like in Section 4.6.2 a direct proof of the convergence of (5.48)–(5.50) can be given : it will however use the results of Theorem 5.3.

Exercise 5.5. *Using Theorem 5.3, given a direct proof of the convergence of* (5.48)–(5.50).

The adaptation of (5.48)–(5.50) *to the discrete problem* $(P^k_{2h}), k = 1, 2$*, is straightforward (see G. L.T. [2, Chap, 4]), since it is a simple variant of the discrete algorithm described in section 4.6.2.*

Exercise 5.6. *Study the discrete analogues of* (5.48)–(5.50) *related to* $(P^k_{2h}), k = 1, 2$.

6 A Second Example of EVI of The Second Kind: The Flow of A Viscous, Plastic Fluid in A Pipe

Most of the material in this section can be found in G.L.T. [1, Chap. 5] and in GLOWINSKI [1], [3].

6.1 The continuous problem. Existence and uniqueness results

Let Ω be a bounded domain of $\mathbb{R}^2$ with a smooth boundary Γ. We define

$$\begin{cases} V = H_0^1(\Omega). \\ a(u,v) = \int_\Omega \nabla u \cdot \nabla v dx, \\ L(v) =< f, v >, f \in V^* \\ j(v) = \int_\Omega |\nabla v| dx \end{cases}$$

Let μ, g be two *positive* parameters ; then

THEOREM 6.1. *The variational inequality*

$$\begin{cases} \mu a(u, v-u) + g\, j(v) - g\, j(u) \geq L(v-u) \ \ \forall v \in V, \\ u \in V \end{cases} \tag{6.1}$$

has a unique solution. **91**

Proof. In order to apply Theorem 4.1 of Chap. 1, we only have to verify that $j(\cdot)$ is convex, proper and l.s.c.

It is obvious that $j(\cdot)$ is convex and proper, Let $u, v \in V$; then

$$|j(v) - j(u)| \geq \sqrt{meas.\Omega}\, \| u - v \|_V, \tag{6.2}$$

hence $j(\cdot)$ is l.s.c.

This proves the Theorem. □

Exercise 6.1. *Prove that $j(\cdot)$ is a norm on V.*

REMARK 6.1. *If we take* $g = 0$ *in* (6.1) *we recover the variational formulation of the Dirichlet problem*

$$\begin{cases} -\mu\Delta u = f \text{ in } \Omega, \\ u = 0 \text{ on } \Gamma. \end{cases}$$

REMARK 6.2. *Since* $a(\cdot,\cdot)$ *is symmetric, the solution* u *of* (6.1) *is characterized, using Lemma 4.1 of Chap. 1, as the unique solution of the minimization problem*

$$\begin{cases} J(u) \leq J(v) \ \ \forall v \in V, \\ u \in V, \end{cases} \tag{6.3}$$

where $J(v) = \frac{\mu}{2}a(v,v) + g\,j(v) - L(v)$.

6.2 Physical motivation

If $L(v) = C\int_\Omega v dx$ (for instance $C > 0$), it is proved in DUVAUT-LIANS [1, Chap. 6] that (6.1) models the *laminar, stationary flow* of a *Bingham fluid* in a cylindrical pipe of cross - section Ω, $u(x)$ being the velocity at $x \in \Omega$ (We refer to PRAGER [1], GERMAIN [1] and DUVAUT - LIONS [1, Chap. 6] for the definition of Bingham fluid). The constant C is the *linear decay of pressure* and μ, g are respectively the *viscosity* and *plasticity yield* of the fluid. The above medium behaves like a viscous fluid (of viscosity μ) in $\Omega^+ = \{x \in \Omega : |\nabla u(x)| > 0\}$ and like a
92 rigid medium in $\Omega^0 = \{x \in \Omega : \nabla u(x) = 0\}$. We refer to MOSSOLOV-MIASNIKOV [1], [2], [3] for a detailed study of the properties of Ω^+ and Ω^0. We observe that (6.1) appears also as a *free boundary problem.*

6.3 Regularity properties

THEOREM 6.2. *(H. BREZIS [4]). If* $L(v) = \int_\Omega f v dx$, $f \in L^2(\Omega)$ *then the solution* u *of* (6.1) *satisfies*

$$u \in V \cap H^2(\Omega)$$

and if Ω *is convex, we have*

$$\| u \|_{H^2(\Omega)} \leq \frac{\gamma(\Omega)}{\mu} \| f \|_{L^2(\Omega)} \tag{6.4}$$

6.4 Further properties

Let us denote by α the following quantity

$$\alpha = \inf_{\substack{v \in H_0^1(\Omega) \\ v \neq 0}} \frac{j(v)}{\| v \|_{L^1(\Omega)}} \tag{6.5}$$

Then $\alpha > 0$.

We derive from this the following

Proposition 6.1. *Let u be the solution of (6.1) and $f \in L^\infty(\Omega)$, then $u = 0$ if*

$$\| f \|_{L^\infty(\Omega)} \leq g\alpha. \tag{6.6}$$

Proof. By taking $v = 0$, $v = 2u$ in (6.1) we obtain

$$\mu a(u,u) + g(u) = \int_\Omega f u dx. \tag{6.7}$$

□

It follows then from (6.5) and from $\int_\Omega f u dx \leq \| f \|_{L^\infty(\Omega)} \| u \|_{L^1(\Omega)}$, that

$$\mu a(u,u) + (g\alpha - \| f \|_{L^\infty(\Omega)}) \| u \|_{L^1(\Omega)} \leq 0. \tag{6.8}$$

If f obeys (6.6) it follows then from (6.8) that $u = 0$. We also have

Proposition 6.2. *Let u be the solution of* (6.1) *and $f \in L^p(\Omega)$, $p > 1$.* **93**
Then if $f \geq 0$, we have $u \geq 0$.

Exercise 6.2. *Prove Proposition 6.2.*

Proposition 6.3. *Let u be the solution of* (6.2) *and $f = c$, a constant. Then $u = 0 \iff c \leq g$ and $u \neq 0$ if $c > g\alpha$.*

Exercise 6.3. *Prove Proposition 6.3*

Proposition 6.4. *Let u be the solution of* (6.1) *and* $f \in L^2(\omega)$*; then* $u = 0$ *if*

$$\| f \|_{L^2(\omega)} \leq g\beta, \tag{6.9}$$

where

$$\beta = \inf_{\substack{v \in H_0^1(\Omega) \\ v \neq 0}} \frac{j(v)}{\| v \|_{L^2(\Omega)}}.$$

Proof. It follows from a result of L. NIRENBERG and M. STRAUSS (cf. STRAUSS [1]) that $\beta > 0$.

By taking $v = 0$ and $v = u$ in (6.1) we obtain

$$\mu a(u, u) + g\, j(u) = \int_\Omega f u dx. \tag{6.10}$$

□

Using $\int_\Omega f u dx \leq \| f \|_{L^2(\Omega)} \| u \|_{L^2(\Omega)}$ and $\beta > 0$ we obtain

$$\mu a(u, u) + (g\beta - \| f \|_{L^2(\Omega)}) \| u \|_{L^2(\Omega)} \leq 0. \tag{6.11}$$

Hence if f satisfies (6.9) we have from (6.11) that $a(u, u) = 0$. This implies $u = 0$, hence the Proposition.

6.5 Exact solutions

6.5.1

Example 1. *We take* $\omega =]0, 1[$ *and* $f = c$*, c a positive constant. In this case the solution of*

$$\begin{cases} \mu \int_0^1 u(v' - u')dx + g \int_0^1 |v'|dx - g \int_0^1 |u'|dx \geq \\ \qquad\qquad\qquad\qquad c \int_0^1 (v - u)dx \ \ \forall v, H_0^1(\Omega), \\ u \in H_0^1(\Omega), \end{cases} \tag{6.12}$$

94 *(where* $v' = \dfrac{dv}{dx}$*) is given by*

$$u = 0 \text{ if } g \geq \frac{c}{2}. \tag{6.13}$$

If $g \leq \dfrac{c}{2}$ then

$$\begin{cases} u(x) = \frac{c}{2\mu}x(1-x) - \frac{gx}{\mu} \text{ if } 0 \leq x \leq \frac{1}{2} - \frac{g}{c} \\ u(x) = \frac{c}{2\mu}(\frac{1}{2} - \frac{g}{c})^2 \text{ if } \frac{1}{2} - \frac{g}{c} \leq x \leq \frac{1}{2} + \frac{g}{c}, \\ u(x) = \frac{c}{e\mu}x(1-x) - \frac{g}{\mu}(1-x) \text{ if } \frac{1}{2} + \frac{g}{c} \leq x \leq 1. \end{cases} \tag{6.14}$$

We observe that if $g < \frac{c}{2}$ then $u \in H_0^1(\Omega) \cap W^{2,\infty}(\omega)$, but $u \notin H^3(\Omega)$.

6.5.2

Example 2. *Let $\omega = \{x : x_1^2 + x_2^2 < R^2\}$, $f = C$, C a positive constant. Then the solution of* (6.1) *is given by*

$$u = 0 \text{ if } g \geq \frac{CR}{2}, \tag{6.15}$$

$$\begin{cases} \text{If } g \leq \frac{CR}{2} \text{ then} \\ u(x) = (\frac{R-r}{2\mu})[\frac{C}{2}(R+r) - 2g] \text{ if } R' \leq r \leq R, \\ u(x) = (\frac{R-R'}{2\mu})[\frac{C}{2}(R+R') - 2g] \text{ if } 0 \leq r \leq R, \end{cases} \tag{6.16}$$

where

$$r = \sqrt{x_1^2 + x_2^2}, R' = \frac{2g}{C}$$

We observe also that if $g < \frac{CR}{2}$ then $u \in H_0^1(\Omega) \cap W^{2,\infty}(\Omega)$ but $u \notin H^3(\Omega)$ (we have actually $u \in H^s(\Omega)$, $s < \frac{5}{2}$).

Exercise 6.4. *Verify that* (6.13), (6.14) *and* (6.15), (6.16) *are solutions of* (6.12) *and* (6.1) *respectively.*

6.6 Existence of multipliers

Let us define Λ by

$$\Lambda = \{q : q \in L^2(\Omega) \times L^2(\Omega), |q(x)| \leq 1 \text{ a.e.}\}$$

$$|q(x)| = \sqrt{q_1^2(x) + q_2^2(x)}; \text{ then we have}$$

THEOREM 6.3. *The solution u of* (6.1) *is characterised by the existence of p such that*

$$\begin{cases} \mu a(u,v) + g\int_\Omega p\cdot\nabla v dx =< f, v > \ \forall v \in V, \\ u \in V, \end{cases} \tag{6.17}$$

$$\begin{cases} p\cdot\nabla u = |\nabla u| a.e., \\ p \in \Lambda. \end{cases} \tag{6.18}$$

Proof. There are classical proofs of the above Theorem using Min-Max of Hahn-Banach Theorems (see for instance CEA [2, Chap. 5], G. L. T/ [1 , Chap. 1], EKELAND- TEMAM [1]). In the sequel following G. L. T. [2 , Chap. 5] we shall give an "almost constructive " proof making use of a regularisation technique.

(1) *We first prove that* (6.17), (6.18) *imply* (6.1). It follows from (6.17) that

$$\begin{cases} \mu a(u, v-u) + g\int_\Omega p\cdot\nabla(v-u)dx = \mu a(v-u) \\ \qquad\qquad +g\int_\Omega p\cdot\nabla v dx - g\int_\Omega p\cdot\nabla u dx = \\ =< f, v-u > \ \forall v \in V. \end{cases} \tag{6.19}$$

It follows from (6.18) that

$$\int_\Omega p\cdot\nabla u dx = \int_\Omega |\nabla u| dx, \tag{6.20}$$

and from the definition of Λ that

$$\int_\Omega p\cdot\nabla v dx \le \int_\Omega |p|\cdot|\nabla v| dx \le \int_\Omega |\nabla v| dx \ \ \forall v \in V. \tag{6.21}$$

□

Then from (6.17), (6.19)–(6.21) we obtain that

$$\begin{cases} \mu a(u, v-u) + g j(u) \ge < f, v-u > \ \forall v \in V, \\ u \in V. \end{cases}$$

Thus (6.17) and (6.18) implies (6.1).

96 (2) *Necessity of* (6.17) *and* (6.18). Taking $u = 0$ and $v = 2u$ in (6.1) we obtain

$$\mu a(u,u) + g\,j(u) =< f, u > . \tag{6.22}$$

Let $\in > 0$. Regularise $j(\cdot)$ by $j_\epsilon(\cdot)$ defined by $j_\epsilon(v) = \int_\Omega \sqrt{\in^2 + v^2}dx$. Since $j_\epsilon(\cdot)$ is convex and continuous on V, the regularised problem

$$\begin{cases} \mu a(u_\epsilon, v - u_\epsilon) + g\,j_\epsilon(u_\epsilon) \geq L(v - u_\epsilon)\ \forall v \in V, \\ u_\epsilon V, \end{cases} \tag{6.23}$$

has a *unique* solution. Let us show that $\lim_{\in \to 0} u_\epsilon = u$ *strongly* in V.

From (6.1) and (6.23) it follows that

$$\mu a(u_\epsilon, u - u_\epsilon) + g\,j_\epsilon(u) - g\,j_\epsilon(u_\epsilon) \geq L(u - u_\epsilon),$$
$$\mu a(u, u_\epsilon - u) + g\,j(u_\epsilon) - g\,j(u) \geq L(u_\epsilon - u).$$

Adding these inequalities we obtain

$$\mu a(u_\epsilon - u, u_\epsilon - u) + g(j_\epsilon(u_\epsilon) - j(u_\epsilon)) \leq g(j_\epsilon(u) - j(u)). \tag{6.24}$$

From $0 < \sqrt{t^2 + \in^2} - |t| = \frac{\in^2}{\sqrt{\in^2 + t^2} + |t|} \leq \in\ \forall t \in R$ it follows that $\mu a(u_\epsilon, u_\epsilon - u) \leq g \in \text{meas.}(\Omega)$, so that

$$\| u_\epsilon - u \|_V \leq \sqrt{meas.(\Omega)}(\frac{g}{u}. \in)^{1/2} \tag{6.25}$$

From (6.25) we obtain

$$\lim_{\in \to 0} u_\epsilon = u \text{ strongly in } V. \tag{6.26}$$

Since $j_\epsilon(\cdot)$ is differentiable on V, the problem (6.23) is equivalent (see CEA [1]) to the following non-linear variational equation:

$$\begin{cases} \mu a(u_\epsilon, v) + g < j'_\epsilon(u), v >= L(v)\ \forall v \in V, \\ u_\epsilon \in V, \end{cases} \tag{6.27}$$

with **97**

$$< j'_\epsilon(w), v >= \int_\Omega \frac{\nabla_w \cdot \nabla_v}{\sqrt{\epsilon^2 + |\nabla w|^2}} dx \forall v, w \in V. \tag{6.28}$$

If we define (p_ϵ) by

$$P_\epsilon = \frac{\nabla u_\epsilon}{\sqrt{\epsilon^2 + |\nabla u_\epsilon|^2}} \tag{6.29}$$

then

$$p_\epsilon \in \Lambda. \tag{6.30}$$

From (6.27)–(6.30) we obtain

$$\begin{cases} \mu a(u_\epsilon, v) + g \int_\Omega p_\epsilon \cdot \nabla_v dx = L(v) \ \ \forall v \in V, \\ u_\epsilon \in V. \end{cases} \tag{6.31}$$

Since Λ is a bounded, closed, convex subset of $L^2(\Omega) \times L^2(\Omega)$ it is *weakly compact* , so that from $(p_\epsilon)_\epsilon$ we can extract a subsequence, still denoted by $(p_\epsilon)_\epsilon$ such that

$$\begin{cases} \lim_{\epsilon \to 0} p_\epsilon = p \text{ weakly in } L^2(\Omega) \times L^2(\Omega), \\ p \in \Lambda. \end{cases} \tag{6.32}$$

Actually we have $p_\epsilon \to p$ in $L^\infty(\Omega) \times L^\infty(\Omega)$ *weakly* *.

Taking the limit as $\epsilon \to 0$ in (6.31) we have from (6.26) and (6.32)

$$\begin{cases} \mu a(u, v) + g \int_\Omega p \cdot \nabla v dx = L(v) \ \forall v \in V, \\ u \in V, \end{cases} \tag{6.33}$$

so that (6.17) is proved.

98 To complete the proof of the Theorem we have only to prove that

$$p \cdot \nabla u = |\nabla u| \text{ a.e.} \tag{6.34}$$

Taking $v = u$ in (6.33) and comparing with (6.22) we obtain

$$\int_\Omega |\nabla u| dx - \int_\Omega p \cdot \nabla u dx = \int_\Omega (|\nabla u| - p \cdot \nabla u) dx = 0. \tag{6.35}$$

Since $p \in \Lambda$, it follows from Schwartz inequality in $\mathbb{R}^2$ that

$$p \cdot \nabla u \leq |\nabla u| a.e.$$

Combining (6.35) and this inequality we obtain

$$p \cdot \nabla u = |\nabla u| a.e.$$

This proves (6.18) and hence the Theorem.

REMARK 6.3. *The function p occurring in* (6.17), (6.18) *is* not unique *if* $\Omega \in \mathbb{R}^2$*; this is shown in G. L. T. [2, Chap. 5].*

REMARK 6.4. *Relation* (6.17) *implies*

$$\begin{cases} -\mu\Delta u - g\nabla \cdot p = f \text{ in } \Omega, \\ u|_\Gamma = 0. \end{cases} \tag{6.36}$$

6.7 Finite element approximation of (6. 1)

In this section we follow G. L. T. [2, Chap. 5]. For the sake of simplicity we shall assume that Ω is a *polygonal* domain of $\mathbb{R}^2$.

6.7.1 Definition of the approximate problem

Let $\mathscr{C}_h$ be as in Sec. 2 of this chapter. We approximate V by

$$V_h = \{v_h \in C^0(\overline{\Omega}) : v_h = 0 \text{ on } \Gamma, v_h|_T \in P_1 \; \forall T \in \mathscr{C}_h\}$$

and (6.1) by

$$\begin{cases} \mu a(u_h, v_h - u_h) + g j(v_h) - g j(u_h) \geq < f, v_h - u_h > \; \forall v_h \in V_h, \\ u_h \in V_h. \end{cases} \tag{6.37}$$

Then **99**

THEOREM 6.4. *The approximate problem* (6.37) *has a unique solution.*

REMARK 6.5. *In these notes, only an approximation by piecewise linear finite elements has been considered. This fact is justified by the existence of a regularity limitation for the solutions of* (6.1) *which implies*

that even with very smooth data we may have $u \notin H^3(\Omega)$ *(see Sec. 6.5. Nevertheless one may find in FORTIN [1], BRISTEAU [1], G. L. T. [1, Chap. 5], BRISTEAU-GLOWINSKI [1], applications of piecewise quadratic finite elements, straight or isoparametric, for solving* (6.1). *The numerical results which have been obtained, seem to prove that for the same* number of degrees of freedom *the accuracies at the nodes are of the same order for the finite elements of order 1 and 2. From the above works it appears also that the second order finite elements are much more costly to use (storage, computational time etc.). We have also to notice that when using first order finite elements,* $\int_\Omega |\nabla v_h| dx$ *can be expressed* exactly *with respect to the values of* v_h *at the nodes of* $\mathscr{C}_h$*, while with second order finite elements we need a numerical integration procedure.*

REMARK 6.6. *From the symmetry of* $a(\cdot,\cdot)$, (6.37) *is equivalent to the minimization problem*

$$\begin{cases} J(u_h) \leq J(v_h) \ \ \forall v_h \in V_h, \\ u_h \in V_h, \end{cases} \tag{6.38}$$

where

$$J(v_h) = \frac{\mu}{2} a(v_h, v_h) + g \int_\Omega |\nabla v_h| dx - < f, v_h > . \tag{6.39}$$

6.7.2 Convergence of the approximate solutions. (General case).

We use the notations of the previous sections.

THEOREM 6.5. *If, as* $h \to 0$, *the angles of* $\mathscr{C}_h$ *are bounded below uniformly in h, by* θ_0 *then*

$$\lim_{h \to 0} \| U_h - u \|_V = 0, \tag{6.40}$$

100 *where u and* u_h *are respectively the solutions of* (6.1) *and* (6.37).

Proof. In order to prove (6.40) we use Theorem 6.3 of Chap. 1. Here, we have to verify that the following three properties hold:

(i) There exist $U \subset V$, $\overline{U} = V$ and $r_h : U \to V_h$ such that $\lim_{h\to 0} r_h v = v$ *strongly in* V, $\forall v \in U$.

(ii) If $v_h \to v$ weakly in V as $h \to 0$ then, $\liminf j_h(v_h) \geq j(v)$.

(iii) $\lim_{h\to 0} j_h(r_h v) = j(v)\, \forall v \in U$.

□

Verification of (i). We take $U = \mathscr{D}(\Omega)$. Then $\overline{U} = H_0^1(\Omega) = V$. Define $r_h v$ by

$$\begin{cases} r_h v \in V_h \ \ \forall v \in H_0^1(\Omega) \cap C^0(\overline{\Omega}), \\ (r_h v)(P) = v(P) \ \ \forall P \in \sum_h^0 . \end{cases} \tag{6.41}$$

Then since $r_h v$ is the linear interpolate of v on $\mathscr{C}_h$ it follows from CIARLET [1], [2], STRANG-FIX [1] that under the above assumptions on $\mathscr{C}_h$ we have

$$\| r_h v - v \|_{H^1(\Omega)} \leq Ch \| v \|_{W^{2,\infty}(\Omega)} . \tag{6.42}$$

Then from (6.42) we obtain $\lim_{h\to 0} r_h v = v$ *strongly* in $H_0^1(\Omega)\, \forall v \in U$.

Verification of (ii). Since $j_h(v_h) = j(v_h)\, \forall v_h \in V_h$, (ii) is trivially satisfied.

Verification of (iii). Since $j_h(v_h) = j(v_h)\ \forall v_h \in V_h$ and from the continuity of $j(\cdot)$ on V, (iii) is trivially satisfied .

Hence form (i), (ii) and (iii) it follows that $u_h \to u$ *strongly* in V.

6.7.3 Convergence of the approximate solutions. ($f \in L^2(\Omega)$)

From the regularity Theorem 6.2 of this chapter we have **101**

$$\| u \|_{H^2(\Omega)} \leq \frac{\gamma_0(\Omega)}{\mu} \| f \|_{L^2(\Omega)} \tag{6.43}$$

if Ω is convex and Γ is sufficiently smooth. This property still holds if Ω is a convex polygonal set.

In this section we always assume Ω to be a convex polygonal domain. We have the following

THEOREM 6.6. *The assumptions on $\mathscr{C}_h$ being those of Theorem 6.5 we have then*

$$\| u_h - u \|_V = 0(h^{1/2}). \tag{6.44}$$

Proof. It follows from (6.1) and (6.37) that $\mu a(u, u_h - u) + g j(u_h) - g j(u) \geq (f, u_h - u), \mu a(u_h, v_h - u_h) + g j(v_h) - g j(u_h) \geq (f, v_h - u_h)\ \forall v_h \in V_h$, which imply, by addition,

$$\begin{cases} \mu a(u_h - u, u_h - u) \leq \mu a(u_h - u, v_h - u) + g j(v_h) - g j(u) \\ \qquad\qquad + \mu a(u, v_h - u) - (f, v_h - u), \\ \forall v_h \in V_h. \end{cases} \tag{6.45}$$

□

Since $j(\cdot)$ is a norm on V, using Schwartz inequality in V and (6.45) we obtain

$$\frac{\mu}{2} \| u_h - u \|_V^2 \leq \frac{\mu}{2} \| v_h - u \|_V^2 + g j(v_h - u) + \mu a(u, v_h - u) - (f, v_h - u)\ \forall v_h \in V_h. \tag{6.46}$$

Since $u \in H_0^1(\Omega) \cap H^2(\Omega)$ we have

$$a(u, v) = -\int_\Omega \Delta u v dx\ \forall v \in V.$$

Hence from (6.46) we have

$$\frac{\mu}{2} \| u_h - u \|_V^2 \leq \frac{\mu}{2} \| v_h - u \|_V^2 + g j(v_h - u) + \int_\Omega (-\mu \Delta u - f)(v_h - u) dx\ \forall v_h \in V,$$

so that

$$\begin{cases} \frac{\mu}{2} \| u_h - u \|_V^2 \leq \frac{\mu}{2} \| v_h - u \|_V^2 + g j(v_h - u) \\ \qquad + (\| \Delta u \|_{L^2(\Omega)} + \| f \|_{L^2(\Omega)}) \| v_h - u \|_{L^2(\Omega)} \\ \forall v_h \in V_h. \end{cases} \tag{6.47}$$

102

Since $\| \Delta u \|_{L^2(\Omega)}$ is a norm equivalent to the $H^2(\Omega)-$ norm over $H^2(\Omega) \cap H_0^1(\Omega)$, it follows from (6.2), (6.43) and (6.47) that

$$\begin{cases} \frac{\mu}{2} \| u_h - u \|_V^2 \leq \frac{\mu}{2} \| v_h - u \|_V^2 + g \sqrt{meqs.(\Omega)} \| v_h - u \|_V + \\ \qquad + (1 + \gamma_0(\Omega)) \| f \|_{L^2(\Omega)} \| v_h - u \|_{L^2(\Omega)}\ \forall v_h \in V_h. \end{cases} \tag{6.48}$$

Since Γ is Lipschitz continuous, we have (cf. NECAS [1]) $H^2(\Omega) \subset C^0(\overline{\Omega})$. The solutions u of (6.1) $\in H^2(\Omega)$ and using the above inclusion, we can define $r_h u$ by

$$\begin{cases} r_h u \in V_h \\ (r_h u)(P) = v(P) \ \ \forall P \in \overset{0}{\Sigma}_h. \end{cases}$$

From the above assumptions on $\mathscr{C}_h$ we have (cf. CIARLET [1], [2], STRANG-FLIX [1])

$$\| r_h u - u \|_V \leq \gamma_1 h \| u \|_{H^2(\Omega)}, \tag{6.49}$$

$$\| r_h u - u \|_{L^2(\Omega)} \leq \gamma_2 h^2 \| u \|_{H^2(\Omega)}, \tag{6.50}$$

where γ_1 and γ_2 are constants independent of h and u.

Taking $v_h = r_h u$ in (6.48) it follows from (6.43), (6.49) and (6.50) that

$$\frac{\mu}{2} \| u_h - u \|_V^2 \leq \frac{\gamma_0 \| f \|_{L^2(\Omega)}}{\mu} [(\frac{\gamma_1^2 \gamma_0}{2} + (1 + \gamma_0)\gamma_2) \| f \|_{L^2} h^2 + g \sqrt{M(\Omega)} \gamma_1 h] \tag{6.51}$$

with $\gamma_0 = \gamma_0(\Omega)$ and $M(\Omega) = meas.(\Omega)$. Hence from (6.51) we have

$$\| u_h - u \| = 0(h^{1/2}).$$

This proves the theorem.

6.8 The case of a circular domain with f= constant

In this section we consider a particular case of the general problem (6.1) **103** by taking

$$\Omega = \{x \in \mathbb{R}^2 : \sqrt{x_1^2 + x_2^2} < R\}, \tag{6.52}$$

$$L(v) = C \int_\Omega v dx, C > 0. \tag{6.53}$$

6.8.1 Exact solutions and regularity properties

The solution of (6.1) corresponding to (6.52), (6.53) is given in sec. 2 of this Chap. We recall that, if $f < \frac{CR}{2}$ then

$$\begin{cases} u \in V \cap W^{2,\infty}(\Omega), \\ u \notin V \cap H^3(\Omega). \end{cases} \tag{6.54}$$

In the sequel we assume that $g < \frac{CR}{2}$.

6.8.2 Approximation by finite element of order 1

Let $\mathscr{C}_h$ be a finite triangulation of Ω satisfying (2.22), (2.23) of Sec. 2.5, Chap. 2 and

$$\forall T \in \mathscr{C}_h, T \subset \overline{\Omega}. \tag{6.55}$$

Define Ω_h and Γ_h by

$$\Omega_h = \bigcup_{T \in \mathscr{C}_h} T, \Gamma_h = \partial\Omega_h.$$

Then $\Omega_h \subset \Omega$ and in the sequel we assume that Γ_h satisfies

$$\text{all the vertices of } \Gamma_h \text{ belongs to } \Gamma. \tag{6.56}$$

Then we approximate V by

$$V_h = \{v_h \in C^0(\overline{\Omega}_h) : v_h = 0 \text{ on } \Gamma_h, v_h|_T \in P_1 \ \forall T \in \mathscr{C}_h\}.$$

Now V_h can be considered as a subspace of V, obtained by extending $v_h \in V_h$ to Ω by taking zero in $\Omega - \Omega_h$. It is then possible to approximate (6.1) by

$$\begin{cases} \mu a(u_h, v_h - u_h) + g\, j(v_h) - g\, j(u_h) \geq C \int_\Omega (v_h - u_h) dx \ \forall v_h \in V_h, \\ u_h \in V_h. \end{cases} \tag{6.57}$$

 This is a finite dimensional problem which has a unique solution.

6.8.3 Error estimate

In this section we will obtain an error estimate of order $h\sqrt{-\log h}$. The three following lemmas play an important role in obtaining the above error estimate.

Lemma 6.1. *Let* $[\overrightarrow{p}], [\overrightarrow{q}] \in \mathbb{R}^2 - \{\{\overrightarrow{0}\}$. *Then*

$$\left|\frac{\overrightarrow{p}}{|\overrightarrow{p}|} - \frac{\overrightarrow{q}}{|\overrightarrow{q}|}\right| \leq 2\frac{|\overrightarrow{p} - \overrightarrow{q}|}{|\overrightarrow{p}| + |\overrightarrow{q}|}. \tag{6.58}$$

Proof. We have

$$(|\overrightarrow{p}| + |\overrightarrow{q}|)\left(\frac{\overrightarrow{p}}{|\overrightarrow{q}|} - \frac{\overrightarrow{q}}{|\overrightarrow{q}|}\right) = (\overrightarrow{p} - \overrightarrow{q}) + \left(\frac{|\overrightarrow{q}|}{|\overrightarrow{p}|}\overrightarrow{p} - \frac{|\overrightarrow{p}|}{|\overrightarrow{q}|}\overrightarrow{q}\right).$$

But

$$\left|\frac{|\overrightarrow{q}|}{|\overrightarrow{p}|}\overrightarrow{p} - \frac{|\overrightarrow{p}|}{\overrightarrow{q}}\overrightarrow{q}\right|^2 = |\overrightarrow{p}|^2 + |\overrightarrow{q}|^2 - 2\overrightarrow{p}\cdot\overrightarrow{q} = |\overrightarrow{p} - \overrightarrow{q}|^2.$$

Consequently,

$$(|\overrightarrow{p}| + |\overrightarrow{q}|)\left|\frac{\overrightarrow{p}}{|\overrightarrow{q}|} - \frac{\overrightarrow{q}}{|\overrightarrow{q}|}\right| \leq 2|\overrightarrow{p} - \overrightarrow{q}|$$

which obviously implies (6.58). □

REMARK 6.7. *In* (6.58), *2 is the best possible constant (take* $\overrightarrow{p} = -\overrightarrow{q}$*). Moreover* (6.58) *is also true in* $\mathbb{R}^N$, $\forall N$

Lemma 6.2. *Let u and* u_h *be the solutions of* (6.1) *and* (6.57) *respectively.*

Let p satisfy (6.17), (6.18); then we have

$$\begin{cases} \mu a(u_n - u, u_n - u) \leq \mu a(u_n - u, v_h - u) \\ \qquad\qquad + g\int_\Omega (p_h - p)\cdot\nabla(v_h - u)dx \;\forall v_h \in V_h \\ \text{and } \forall p_h \in \Lambda \text{ such that } p_h\cdot\nabla v_h = |\nabla v_h| a.e \end{cases} \tag{6.59}$$

Proof. We shall prove (6.59) with $f \in V^*$. **105**

From (6.45) we have

$$\begin{cases} \mu a(u_h - u, u_h - u) & \leq \mu a(u_h - u, v_h - u) + g\, j(v_h) \\ & \qquad -g\, j(u) + \mu a(u, v_h - u) \\ & - < f, v_h - u > \ \forall v_h =\in V_h. \end{cases}$$

□

Taking account of (6.17) we obtain

$$\begin{cases} \mu a(u_h - u, u_h - u) \leq \mu a(u_h - u, v_h - u) + g\, j(v_h) - g\, j(u) \\ \qquad\qquad -g \int_\Omega p \cdot \nabla(v_h - u) dx \\ \forall v_h \in V_h. \end{cases} \tag{6.60}$$

Let $v_h \in V_h$ and $p_h \in \Lambda$ such that $p_h \cdot \nabla v_h = |\nabla v_h| a.e.$; such a p_h always exists. Substituting this in (6.60) and using the following relations

$$j(v_h) = \int_\Omega p_h \cdot \nabla v_h dx, \tag{6.61}$$

$$j(u) = \int_\Omega |\nabla u| dx \geq \int_\Omega p_h \cdot \nabla u dx, \tag{6.62}$$

we obtain (6.59).

This proves the Lemma.

Let u be the solution of (6.1) and $\delta > 0$. Define $\Omega^\delta \subset \Omega$ by

$$\Omega^\delta = \{x \in \Omega : |\nabla u(x)| > \delta\}.$$

In the case of the problem (6.1) associated with (6.52), (6.53) (assuming $g < \frac{CR}{2}$) we have

Lemma 6.3. *We have the following identity*

$$\int_{\Omega^\delta} \frac{dx}{|\nabla u|} = \frac{4\pi\mu}{C} \left[-\frac{2\mu}{C}\delta + \left(R - \frac{2g}{c}\right) + \frac{2g}{C} log\left(R - \frac{2g}{C}\right) - \frac{2g}{C} \log \frac{2\mu}{C}\delta \right]. \tag{6.63}$$

Proof. We obtain from (6.16)

$$\left|\frac{du}{dr}\right| = \frac{1}{\mu}\left(\frac{Cr}{2} - g\right) \text{ if } \frac{2g}{C} \leq r \leq \mathbb{R},$$

so that

$$\int_{\Omega}^{\delta} \frac{dx}{|\nabla u|} = 2\pi\mu \int_{\frac{2}{C}(\mu\delta+g)}^{R} \frac{rdr}{\frac{Cr}{2} - g},$$

which implies (6.63). □ **106**

From the above Lemmas we shall deduce

THEOREM 6.7. *Let u be the solution of the problem* (6.1) *associated with* (6.52), (6.53). *Let u_h be the solution of the problem* (6.57) *with $\mathscr{C}_h$ satisfying* (6.55), (6.56). *Assume that as $h \to 0$ the angles of $\mathscr{C}_h$ are bounded from below uniformly in h by $\theta_0 > 0$. Then we have*

$$\| u_h - u \|_V = 0(h\sqrt{-\log h}). \tag{6.64}$$

Proof. Starting from Lemma 6.2, we obtain from (6.59)

$$\begin{cases} \frac{\mu}{2} \| u_h - u \|_V^2 \leq \frac{\mu}{2} \| r_h u - u \|_V^2 + g \int_{\Omega} |p_h - p| \cdot |\nabla(r_h u - u)| dx \forall p_h \in \Lambda \\ \text{such that } p_h \cdot \nabla r_h u = |\nabla r_h u|, \end{cases} \tag{6.65}$$

where $r_h u$ is defined by

$$\begin{cases} r_h u \in V_h \\ (r_h u)(P) = u(p) \, \forall P \in \text{ vertex of } \mathscr{C}_h. \end{cases}$$

We have $r_h u = 0$ on $\Omega - \Omega_h$ so that

$$\| r_h u - u \|_V = \int_{\Omega} |\nabla(r_h u -)|^2 dx = \int_{\Omega - \Omega_h} |\nabla u|^2 dx + \int_{\Omega_h} |\nabla(r_h u - u)|^2 dx. \tag{6.66}$$

□

Let us define

$$X_1 = \frac{\mu}{2} \int_{\Omega - \Omega_h} |\nabla u|^2 dx,$$

$$X_1 = \frac{\mu}{2} \int_{\Omega_h} |\nabla(r_h u - u)|^2 dx.$$

It is easily shown that **107**

$$meas.(\Omega - \Omega_h) < \frac{\pi}{4} h^2. \tag{6.67}$$

Furthermore (6.16) implies that

$$|\nabla u(x)| \leq \frac{C}{2\mu}\left(R - \frac{2g}{C}\right) \forall x \in \Omega. \tag{6.68}$$

It follows from (6.67) and (6.68) that

$$X_1 \leq \frac{\pi}{32\mu} C^2 \left(R - \frac{2g}{C}\right)^2 h^2. \tag{6.69}$$

Since $u \in W^{2,\infty}(\Omega)$, on each triangle $T \in \mathscr{C}_h$ we have (cf. CIARLET - WAGSHAL [1])

$$|\nabla(r_h u - u)(x)| \leq \frac{2h}{\sin\theta_0} \| \rho(D_2 u) \|_{L^\infty(\Omega)}, \tag{6.70}$$

where $D_2 u(x)$ is the *Hessian matrix* of u at x, defined by

$$D_2 u(x) = \begin{pmatrix} \frac{\partial^2 u}{\partial x_1^2}(x) & \frac{\partial^2 u}{\partial x_1 \partial x_2}(x) \\ \frac{\partial^2 u}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 u}{\partial x_2^2}(x) \end{pmatrix}$$

and $\rho(D_2 u(x))$ is the *spectral radius* of $D_2 u(x)$

We have

$$D_2 u(x) = 0 \text{ if } 0 < r < \frac{2g}{C} \tag{6.71}$$

so that $\rho(D_2 u(x)) = 0$ and it is easily verified that

$$\rho(D_2 u(x)) = \frac{C}{2\mu} \text{ if } \frac{2g}{C} < r < R. \tag{6.72}$$

Then (6.70)–(6.72) imply

$$X_2 \leq \frac{\pi}{\mu} R(R - \frac{2g}{C} + h) C^2 \left(\frac{h}{\sin\theta_0}\right)^2. \tag{6.73}$$

108 So it remains to estimate in (6.65) the term $g\int_\Omega |p_h - p|\cdot|\nabla(r_h u - u)|dx$ We have

$$\overline{\Omega} = \bigcup_{i=3}^{6} \overline{\Omega}_i, \tag{6.74}$$

where

$$\Omega_3 = \Omega - \Omega_h,$$
$$\Omega_4 = \{x \in \Omega_h : r > \frac{2g}{C} + h\},$$
$$\Omega_5 = \{x \in \Omega_h : \frac{2g}{C} - h < r < \frac{2g}{C} + h\},$$
$$\Omega_6 = \{x \in \Omega_h : 0 \le r < \frac{2g}{C} - h\}.$$

Let us define for $3 \le i \le 6$

$$X_i = g\int_{\Omega_i} |p_h - p|\cdot|\nabla(r_h u - u)|dx. \tag{6.75}$$

We have $r_h u = 0$ over $\Omega - \Omega_h$ so that in (6.65) we can take

$$p_h = 0 \text{ over } \Omega - \Omega_h. \tag{6.76}$$

From (6.67), (6.68), (6.70) and $p \in \Lambda$ it follows that

$$X_3 \le g\int_{\Omega-\Omega_h} |\nabla u|dx \le \frac{\pi}{8\mu} gC(R - \frac{2g}{C})h^2. \tag{6.77}$$

From (6.70)–(6.72) and since $p, p_h \in \Lambda$ we have

$$X_5 \le 2g\frac{C}{\mu} meas.(\Omega_5)\frac{h}{\sin\theta_0},$$

so that

$$X_5 \le \frac{16\pi}{\mu} g^2 \frac{h^2}{\sin\theta_0}. \tag{6.78}$$

From the definition of h(h = maximal length of the edges of $T \in \mathscr{C}_h$) we have $r_h u = u$ = constant over Ω_6, so that

$$X_6 = 0. \tag{6.79}$$

It remains to estimate X_4. Since the equipotential of u in Ω_4 are circular, for h sufficiently small we have from (6.16) that **109**

$$\begin{cases} r_h u|_T \neq \text{constant} & \forall T \in \mathscr{C}_h, T \subset \overline{\Omega}_4, \text{so that} \\ \nabla r_h u|_T \neq 0 & \forall T \in \mathscr{C}_h, T \subset \overline{\Omega}_4. \end{cases} \tag{6.80}$$

Taking account of (6.80) it follows from (6.65) that

$$p_h|_T = \frac{\nabla r_h u}{|\nabla r_h u|}\bigg|_T \forall T \in \mathscr{C}_h, T \subset \overline{\Omega}_4. \tag{6.81}$$

Furthermore we observe that (6.16) implies that

$$|\nabla u(x)| \geq \frac{Ch}{2\mu} > 0 \, \forall x \in \Omega_4. \tag{6.82}$$

This in turn implies that

$$p = \frac{\nabla_u(x)}{|\nabla_u(x)|} \forall x \in \Omega_4.$$

It follows from (6.80)–(6.82), applying Lemma 6.1. to the pair $\{\nabla r_h u, \nabla_u\}$ and Lemma 6.3 with $\delta = \frac{Ch}{2\mu}$, that

$$X_4 \leq g \parallel \nabla(r_h u - u) \parallel_\infty^2 \int_{\Omega_4} \frac{dx}{|\nabla_u| + |\nabla r_h u|},$$

where

$$\parallel \nabla_v \parallel_\infty = \parallel \nabla_v \parallel_{L^\infty(\Omega)\times L^\infty(\Omega)} .$$

It follows from (6.70) and (6.82) that

$$X_4 \leq g \frac{C^2}{\mu^2} \left(\frac{Ch}{\sin\theta_0}\right)^2 \int_{\Omega^0} \frac{dx}{|\nabla u|},$$

which implies, using (6.63), that

$$X_4 \leq \frac{4\pi}{\mu} gC \left(\frac{h}{\sin\theta_0}\right)^2 \left[-h + \left(R - \frac{2g}{C}\right) + \frac{2g}{C} \log\left(R - \frac{2g}{C}\right) \frac{2g}{C} \log h\right]$$

or more simply

$$X_4 \leq \frac{4\pi}{\mu} gC \left(\frac{h}{sin\theta_0}\right)^2 \left(R - \frac{2g}{C} \log \frac{h}{R}\right). \tag{6.83}$$

110

Taking into account (6.65), the estimate (6.64) is obtained by addition of the $X_i, i = 1, \ldots, 6$. More precisely, for sufficiently small h,

$$\| u_h - u \|_V \leq 4\frac{g}{\mu} \sqrt{\pi} \frac{h}{\sin \theta_0} \sqrt{-\log h}. \tag{6.84}$$

6.8.4 Generalization

From the numerical experiments, we have done, it seems that in a great of cases (important from the point of view of application we have the following properties for u:

(1) $u \in V \cap W^{2,\infty}(\Omega)$,

(2) $\Omega_0 = \{x : \nabla u(x) = 0\}$ is a compact subset of Ω with smooth boundary,

(3) Ω_0 has a finite number of connected components.

Moreover it seems that in the above cases we can conjecture that for $\delta > 0$ we still have

$$\int_{\Omega^\delta} \frac{dx}{|\nabla(x)|} = 0(-\log \delta). \tag{6.85}$$

With these properties we can easily prove the following error estimate:

$$\| u_h - u \|_V = 0\left(h\sqrt{-\log h}\right).$$

REMARK 6.8. *Using an equivalent formulation of* (6.1) *(less suitable for computations) FALK-MERCIE [1] have obtained an* $0(h)$ *estimate for* $\| u_h - u \|_{H^1(\Omega)}$ *for the piecewise linear approximation.*

6.9 Iterative solution of the continuous and approximate problems by Uzawa's algorithm

We begin with the continuous problem (6.1). Let us define $\mathscr{L} : V \times H \to \mathbb{R}$ by

$$\mathscr{L}(v,q) = \frac{\mu}{2}a(v,v) - L(v) + g\int_\Omega q \cdot \nabla v dx \; \forall v \in V, \; \forall q \in H,$$

where $H = L^2(\Omega) \times L^2(\Omega)$.

111 Let $\{u, p\}$ be the solution of (6.17), (6.18). Then we have

THEOREM 6.8. *The pair $\{u, p\}$ is a saddle point of $\mathscr{L}$ over $V \times \Lambda \iff$ $\{u, p\}$ satisfies* (6.17) *and* (6.18).

Exercise 6.5. *Prove the Theorem 6.8.*

It follows from CEA [2, Chap. 5] (see also G.L.T.[2, Chap. 5]) that to solve (6.1) we can use the following Uzawa's algorithm.

$$p^0 \in \Lambda \text{ arbitrarily chosen (for example } p = 0\text{)}, \tag{6.86}$$

then by induction knowing p^n we compute u^n and p^{n+1} by

$$\begin{cases} \mu a(u^n, v) =< f, v > -g\int_\Omega p^n \cdot \nabla v dx \forall v \in V, \\ u^n \in V, \end{cases} \tag{6.87}$$

$$p^{n+1} = P_\Lambda(p^n + \rho g \nabla u^n), \tag{6.88}$$

where $P_\Lambda : H \to \Lambda$ is the projection operator in the H-norm, defined by

$$P_\Lambda(q) = \frac{q}{\sup(1, |q|)}.$$

Since u^n is a solution of (6.87), u^n is actually the unique solution in V of

$$\begin{cases} -\mu \Delta u^n = f + g\nabla \cdot p^n, \\ u^n|_\Gamma = 0. \end{cases} \tag{6.89}$$

We shall give a direct proof for the convergence of (6.86)–(6.88) based on the theorem 6.3 of Sec. 6.6

THEOREM 6.9. *Let u^n be the solution of* (6.87). *Then if*

$$0 < \rho < \frac{2\mu}{g^2}, \tag{6.90}$$

we have

$$\lim_{n\to\infty} \| u_n - u \|_V = 0, \tag{6.91}$$

where u is the solution of (6.1). **112**

Proof. Let $\{u, p\}$ satisfies (6.17) and (6.18). Then (6.18) implies

$$p = P_\Lambda(p + \rho g \nabla u). \tag{6.92}$$

We define $u^{-n} = u^n - u$, $p^{-n} = p^n - p$. Using the fact that P_Λ is a contraction mapping and from (6.88), (6.92) we obtain

$$|p^{-n+1}|^2 \leq |p^{-n}|^2 + 2\rho g \int_\Omega p^{-n} \cdot \nabla u^{-n} dx + \rho^2 g^2 \int_\Omega |\nabla u^{-n}|^2 dx, \tag{6.93}$$

where

$$|q| = \| q \|_{L^2(\Omega)\times L^2(\Omega)} .$$

It follows from (6.17), (6.87) that

$$\mu a(u^{-n}, v) + g \int_\Omega p^{-n}.\nabla v dx = 0 \forall v \in V, \tag{6.94}$$

Replacing v by u^{-n} in (6.94) we obtain

$$\mu a(u^{-n}, u^{-n}) + g \int_\Omega p^{-n} \cdot \nabla v^{-n} dx = 0. \tag{6.95}$$

From (6.93) and (6.95) we have

$$|p^{-n}|^2 - |p^{-n+1}|^2 \geq \rho(2\mu - \rho g^2) \| u^{-n} \|_V^2 . \tag{6.96}$$

If $0 < \rho < \frac{2\mu}{g^2}$ then using a standard reasoning, we obtain that

$$\lim_{n\to\infty} \| u^{-n} \|_V = 0,$$

which proves the Theorem. □

Let us describe the adaptation of (6.86)–(6.88) to the approximate problem (6.37). We define $L_n \subset L^2(\Omega) \times L^2(\Omega)$ by

$$L_h = \{q_h : q_h = \sum_{T \in \mathscr{C}_h} q_T X_T, q_T \in \mathbb{R}^2 \ \forall T \in \mathscr{C}_h\}$$

113 where X_T is the characteristic function of T.

It is then clear that $\forall v_h \in V_h$, $\nabla v_h \in L_h$. We also define Λ_h by $\Lambda_h = \Lambda \cap L_h$. We can easily prove that

$$P_{\Lambda_h}(q_h) = P_\Lambda(q_h) \ \forall q_h \in L_h.$$

Then (6.86)–(6.88) is approximated by

$$p_h^0 \in \Lambda_h \text{ arbitrarily chosen }, \tag{6.97}$$

by induction knowing p_h^n *we obtain* u_h^n *and* p_h^{n+1} *by*

$$\begin{cases} \mu a(u_h^n, v_h) = L(v_h) - g \int_\Omega p_h^n \cdot \nabla_{v_h} dx \ \forall v_h \in V_h, \\ u_h^n \in V_h, \end{cases} \tag{6.98}$$

$$p_h^{n+1} = P_\Lambda(p_h^n + \rho g \nabla u_h^n). \tag{6.99}$$

Then for $0 < \rho < \dfrac{2\mu}{g^2}$ we obtain the convergence of u_h^n to u_h.

Exercise 6.6. *Study the convergence of* (6.97)–(6.99)

REMARK 6.9. *The above methods have been numerically applied for solving* (6.1) *in CEA-GLOWINSKI [2], BRISTEAU [1], BRISTEAU-GLOWINSKI [1], G. L. T. [2, Chap. 5]. They appear to be very efficient and particularly well suited to take into account non-differentiable functionals like* $\int_\Omega |\nabla_v| dx$.

7 On Some Useful Formulae

Let T be the triangle of Figure 7.1. We denoted by $M(T)$ the measure of T.

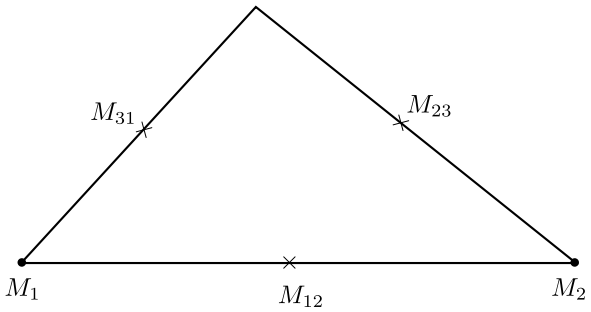

Figure 7.1:

Let v be a function defined on T. We define v_i and v_{jk} by **114**

$$v_i = v(M_i), v_{jk} = v(M_{jk}).$$

Then we have the following formulae

$$\int_T uvdx = \frac{M(T)}{12}\{(u_1+u_2)(v_1+v_2)+(u_2+u_3)(v_2+v_3) + (u_3+u_1)(v_3+v_1)\}\ \forall u, v \in P_1, \quad (7.1)$$

$$\begin{cases} |\nabla_v|^2 = \frac{1}{4M(T)^2}\{|\overrightarrow{M_2M_3}|^2v_1^2 + |\overrightarrow{M_3M_1}|^2v_2^2 + |\overrightarrow{M_1M_2}|^2v_3^2 \\ \qquad +2\overrightarrow{M_2M_3}\cdot\overrightarrow{M_3M_1}v_1v_2 \\ +2\overrightarrow{M_1M_2}\cdot\overrightarrow{M_2M_3}v_3v_1 + \overrightarrow{2M_3M_1}\cdot\overrightarrow{M_1M_2}v_2v_3\}, \\ \forall v \in P_1, \end{cases} \quad (7.2)$$

$$\begin{cases} \int_T |v|^2dx \\ = \frac{M(T)}{3}\{\frac{1}{10}(v_1^2+v_2^2+v_3^2) + \frac{8}{15}(v_{22}^2+v_{23}^2+v_{31}^2) - \frac{1}{30}(v_1v_2+v_2v_3+v_3v_1) \\ +\frac{8}{15}(v_{12}v_{23}+v_{23}v_{31}+v_{31}v_{12}) - \frac{2}{15}(v_1v_{23}+v_2v_{31}+v_3v_{12})\},\ \forall v \in P_2, \end{cases} \quad (7.3)$$

$$\begin{cases} \int_T |\nabla v|^2 dx = \frac{1}{12M(T)}\{|v_1\overrightarrow{M_2M_3} - v_2\overrightarrow{M_3M_1} + v_3\overrightarrow{M_1M_2} \\ \qquad\qquad +2(v_{12}+v_{23}-v_{31})\overrightarrow{M_3M_1}|^2 \\ +|v_1\overrightarrow{M_2M_3} + v_2\overrightarrow{M_3M_1} - v_3\overrightarrow{M_1M_2} + 2(v_{23}+v_{13}-v_{12})\overrightarrow{M_1M_2}|^2 + \\ +|-v_1\overrightarrow{M_2M_3} + v_2\overrightarrow{M_3M_1} + v_3\overrightarrow{M_1M_2} + 2(v_{31}+v_{12}-v_{23}) \\ \qquad\qquad \overrightarrow{M_2M_3}|^2\}\ \forall v \in P_2.\} \end{cases} \tag{7.4}$$

115

The above formulae may be useful to express the approximations of the problems of this chapter, in a from suitable for computations.

Chapter 3

On The Approximation of Parabolic Variational Inequalities

1 Introduction References

In this chapter we would like to give some indications on the approxima- **116**
tion of *Parabolic Variational Inequalities* (PVI) (mostly without proof). For a detailed treatment see G. L. T. [2 Chap. 6], TREMOLIERES [1], and for further reference see FORTIN [1], BRISTEAU [1], BRISTEAU-GLOWINSKI [1], C. JOHNSON [1] and A. BERGER [1]. See also LASCAUX [1] for the numerical analysis of *time dependent equations.*

2 Formulation And Statement of The Main Results

Let H and V be two real Hilbert spaces that $V \subset H$, $\overline{V} = H$. Assuming that $H = H^*$ we have then $V \subset H \subset V^*$.

The scalar product in H (resp. . in V) and the corresponding norms are denoted by $(\cdot,\cdot)$,$|\cdot|$ (resp. . $((\cdot,\cdot))$, $\|\cdot\|$). Moreover we also use $(\cdot,\cdot)$ for the duality between V^* and V.

We now introduce:

- A time interval $[0, T]$ with $0 < T < \infty$, a bilinear from $a : V \times V \to \mathbb{R}$, continuous and *elliptic* in the following sense :$\exists \alpha > 0$ and $\lambda \geq 0$ such that

$$a(v, v) + \lambda |v|^2 \geq \alpha \|v\|^2 \ \forall v \varepsilon V,$$

- $$f \varepsilon L^2(0, T; V), u^0 \in H,$$

  (for the definition of $L^2(0, T; X)$) see LIONS [1], [3])

- K : closed, convex, non-empty subset of V,

- $j : V \to \overline{\mathbb{R}}$ convex, proper, l.s.c

We consider then the following two families of PVI:

$$\begin{cases} \text{Find } u(t) \text{ such that} \\ \left(\frac{\partial u}{\partial t}, v - u\right) + a(u, v - u) \geq (f, v - u) \ \forall v \varepsilon K, \ a.e.\ t \in]0, T[, \\ u(t) \in K \ a.e.\ t \in]0, T[, u(0) = u_0, \end{cases} \tag{2.1}$$

117 and

$$\begin{cases} \text{Find } u(t) \text{ such that} \\ (\frac{\partial u}{\partial t}, v - u) + a(u, v - u) + j(v) - j(u) \geq (f, v - u) \ \forall v \in V, \ a.e.\ t \in]0, T[, \\ u(t) \in V \ a.e.\ t \in]0, T[, u(0) = u_0. \end{cases} \tag{2.2}$$

REMARK 2.1. *If $K = V$ and $J \equiv 0$ then* (2.1) *and* (2.2) *reduce to the standard parabolic variational equation*

$$\begin{cases} (\frac{\partial u}{\partial t}, v) + a(u, v) = (f, v) \ \forall v \in V, \ a.e. \ \ in\ t \in]0, T[, \\ u(t) \in V \ a.e.\ t \in]0, T[, u(0) = u_0. \end{cases} \tag{2.3}$$

Under appropriate assumptions on u_0, K and $j(\cdot)$ it is proved that (2.1), (2.2) have unique solutions in $L^2(0, T; V) \cap C^0([0, T], H)$. For the proof of this we refer to BREZIS [4], [5]; LIONS [1],DUVAUT-LIONS[1].

In the following sections of this chapter we would like to give some discretisation schemes for (2.1), (2.2) and then in sec. 6 study the asymptotic properties in time of a specific example, for the continuous and discrete cases.

3 Numerical Schemes For Parabolic Linear Equations

Let us assume that V and H have been approximated (as $h \to 0$) by the same family $(V_h)_h$ of closed subspaces of V (in practice the V_h are finite dimensional). We also approximate $(\cdot,\cdot)$, $a(\cdot,\cdot)$ by $(\cdot,\cdot)_h$, $a_h(\cdot,\cdot)$ is such a way that ellipticity, symmetry etc. are preserves. We also assume that u_0 is approximated by $(u_{oh})_h$ such that $u_{oh} \in V_h$ and $\lim_{h\to 0} u_{0h} = u_0$ strongly in H.

We now introduce a *time step* Δt; then denoting u_h^n the approximation of u at time $t = n\Delta t\,(n = 0, 1, 2, \ldots,)$, we approximate (2.3) using the classical step by step numerical schemes (i.e. we describe how to compute u_n^{n+1}) if u_h^n and u_h^{n-1} are known).

1. **Explicit scheme.** **118**

$$\begin{cases} \left(\frac{u_n^{n+1}-u_h^n}{\Delta t}, v_h\right)_h + a_h(u_h^n, v_h) = (f_h^n, v_h)_h \ \forall v_h \in v_h, \\ n = 0, 1, \ldots, \\ u_h^0 = u_{0h}. \end{cases} \tag{3.1}$$

Stability. (see LASCAUK [1]) for the terminology) *conditional.* *Accuracy.* $0(\Delta t)$ (we just consider the influence of the time discretisation).

2. **Ordinary implicit scheme.**

$$\begin{cases} \left(\left(\frac{u_n^{n+1}-u_h^n}{\Delta t}, v_h\right)\right)_h + a_h(u_h^{n+1}, v_h) = (f_h^{n+1}, v_h)_h \ \forall v_h \in V_h, \\ n = 0, 1, 2, \ldots, \\ u_h^0 = u_{0h}. \end{cases} \tag{3.2}$$

Stability. Unconditional.

Time accuracy. $0(\Delta t)$

3. **Cranck-Nicholson scheme.**

$$\begin{cases} \left(\frac{u_h^{n+1}-u_h^n}{\Delta t}, v_h\right)_h + a_h\left(\frac{u_h^{n+1}+u_h^n}{2}, v_h\right) = \left(\frac{f_h^{n+1}+f_h^n}{2}, v_h\right)_h \ \forall v_h \in V_h \\ \text{or } = (f_h^{n+1/2}, v_h)_h \ \forall v_h \in V_h \\ n = 0, 1, 2, \ldots, ; u_h^0 = u_{oh}. \end{cases} \tag{3.3}$$

Stability. Unconditional.

Time accuracy. $0(|\Delta t|^2)$.

4. **Two steps implicit scheme.**

$$\begin{cases} (\frac{\frac{3}{2}u_h^{n+1}-2u_h^n+\frac{1}{2}u_n^{n-1}}{\Delta t}, v_h)_h + a_h(u_h^{n+1}, v_h) = (f_h^{n+1}, v_h)_h \ \forall v_h \in V_h, \\ n = 1, 2, \ldots, u_h^0 = u_{0h}, u_h^1 \text{ given .} \end{cases} \tag{3.4}$$

119 *Stability. Unconditional.*

Time accuracy. $0(|\Delta t|^2)$.

Unlike the three previous schemes, this latter scheme requires the use of a *starting procedure* to obtain u_h^1 from $u_h^0 = u_{0h}$; to compute u_h^1 we can use for example one of the scheme (3.1), (3.2) or (3.3); we recommend (3.3) since it is also an $0(|\Delta t|^2)$-scheme. Similarly the generalisations of scheme (3.4) discussed in Sec. 4, 5 will require the use of a starting procedure which can be the corresponding generalization of schemes (3.1), (3.2) or (3.4).

REMARK 3.1. *The vector* f_h^n *(or* $f_h^{n+1/2}$*) occurring in the right hand sides of* (3.1)- (3.4) *is a convenient approximation of* f *at* $t = n\Delta t$ *(or* $t = (n + \frac{1}{2})\Delta t$*).*

In some cases it may be defined as follows (we just consider f_h^n *since the technique described below is also applicable to* $f_n^{n+1/2}$*).*

First we define $f^n \in V^*$, by

$$f^n = f(n\Delta t) \text{ if } f \in C^0[0, T; V^*].$$

In the general case, it is defined by

$$f^0 = \frac{2}{\Delta t}\int_0^{\frac{\Delta t}{2}} f(t)dt,$$
$$f^n = \frac{1}{\Delta t}\int_{(n-\frac{1}{2})\Delta t}^{(n+\frac{1}{2})\Delta t} f(t)dt \text{ if } n \geq 1.$$

Then since $(\cdot,\cdot)_h$ is a scalar product on V_n one may define f_h^n by

$$(f^n, v_h) = (f_h^n, v_h)_h \ \forall v_h \in V_h,\ \ f_h^n \in V_h.$$

In some cases we have to use more sophisticated methods to define f_h^n.

REMARK 3.2. *At each step* $(n+1)$ *we have to solve a linear system to compute* u_h^{n+1}*; however if we can use a scalar product* $(\cdot,\cdot)_h$ *leading to a* diagonal matrix*, with regard to the variables defining* v_h*, then the use explicit scheme will only require to solve* one variable linear equations *at each step.*

REMARK 3.3. *We can also use nonconstant time steps* Δt_n.

REMARK 3.4. *If we are interested in the numerical integration of* **120**
"Stiff" phenomenon *or in* long range integration *we can briefly say that*

- *Schemes* (3.1), (3.2) *are* too *dissipative, moreover the stability condition in* (3.1) *may be a serious drawback.*
- *Scheme* (3.3) *is, in some sense,* not sufficiently dissipative.
- *Scheme* (3.4) *avoids the above inconveniences and is highly recommended for "Stiff" problems and long range integration. In most cases the extra storage it requires is not a serious drawback.*

REMARK 3.5. *There are many works related to the numerical analysis of parabolic equations via finite differences in time and finite elements in space approximations. We refer to RAVIART [1], [2], CROUZEIX [1], STRANG-FIX [1], ODEN-REDDY [1, CHAP. 9] and the bibliographies therein.*

4 Approximation of PVI of The First Kind

We assume that K in (2.1) has been approximated by (K_h), $K_h \subset V_h \forall h$, like in the elliptic case (see Chap. 1). We also suppose that the bilinear form a is possibly dependent on the time t and has been approximated by $a(t; u_h, v_h)$.

1. **Explicit scheme.**

$$\begin{cases} \left(\frac{u_h^{n+1}-u_h^n}{\Delta t}, v_h - u_h^{n+1}\right)_h + a_h(n\Delta t; u_h^n, v_h - u_h^{n+1}) \\ \qquad\qquad \geq (f_h^n, v_h - u_h^{n+1})_h \ \forall v_h \in K_h, \\ u_h^{n+1} \in K_h, \\ n = 0, 1, 2, \ldots, u_h^0 = u_{0h}. \end{cases} \tag{4.1}$$

Stability. Conditional (see G.L.T [2, Chap 6]). This scheme is almost never used in practice since it is conditional stable and that the computation of u_h^{n+1} will require in general, the use of an iterative method *even* if the matrix corresponding to $(\cdot,\cdot)_h$ is *diagonal*.

121 2. **Ordinary implicit scheme.**

$$\begin{cases} \left(\frac{u_h^{n+1}-u_h^n}{\Delta t}, v_h - u_h^{n+1}\right)_h + a_h((n+1)\Delta t; u_h^{n+1}, v_h - u_h^{n+1}) \\ \qquad\qquad \geq (f_h^{n+1}, v_h - u_h^{n+1}) \ \forall v_h \in K_h, \\ u_h^{n+1} \in K_h, \\ n = 0, 1, 2, \ldots, u_h^0 = u_{oh}. \end{cases} \tag{4.2}$$

Stability. Unconditional.

At each step we have to solve an EVI of the first kind in K_h to compute u_h^{n+1}. This scheme is very much used in practice.

3. **Crank-Nicholson scheme.**

$$\begin{cases} (\frac{u_h^{n+1}-u_h^n}{\Delta t}, v_h - u_h^{n+1/2})_h + a_h((n+\frac{1}{2})\Delta t; u_h^{n+1/2}, v_h - u_h^{n+1/2}) \\ \geq f_h^{n+1/2}, v_h - u_h^{n+1/2}) \ \forall v_h \in K_h, \\ u_h^{n+1/2} \in K_h, u_h^{n+1/2} = \frac{u_h^n + u_h^{n+1}}{2}, nn = 0, 1, 2, \ldots, u_h^0 = u_{0h}. \end{cases} \tag{4.3}$$

Stability. Unconditional.

Since $\dfrac{u_h^{n+1}-u_h^n}{\Delta t} = \dfrac{u_h^{n+1/2}-u_h^n}{\dfrac{\Delta t}{2}}$, we observe that at *each step* we have to solve an EVI of the *first kind* to compute $u_h^{n+1/2}$. We observe also that possibly $u_h^n \notin K_h$. We do not recommend this scheme if the regularity in time of the continuous solution is poor.

4. **Two steps implicit schemes.**

$$\begin{cases} (\frac{\frac{3}{2}u_h^{n+1}-2u_h^h+\frac{1}{2}u_h^{n-1}}{\Delta t}, v_h - u_h^{n+1})_h + a_h((n+1)\Delta t; u_h^{n+1}, v_h - u_h^{n+1}) \\ \geq (f_h^{n+1}, v_h - u_h^{n+1})_h \ \forall v_h \in K_h, u_h^{n+1} \in K_h, \\ \qquad n = 1, 2, \ldots, u_h^0 = u_{0h}, u_h^1 \text{ given.} \end{cases} \tag{4.4}$$

Stability. Unconditional. We have to solve at each step an EVI of the first kind in K_h to compute u_h^{n+1}. Remark 3.4 applies to this scheme also.

5 Approximation of PVI of The Second Kind

1. **Explicit scheme** **122**

$$\begin{cases} (\frac{u_h^{n+1}-u_h^n}{\Delta t}, v_h - u_h^{n+1})_h + a_h(n\Delta t; u_h^n, v_h - u_h^{n+1}) + j_h(v_h)(u_h^{n+1}) \geq \\ \geq f_h^n, v_h - u_h^{n+1}) \ \forall v_h \in V_h, u_h^{n+1} \in V_h, n = 0, 1, 2, \ldots, u_h^0 = u_{0h}. \end{cases} \tag{5.1}$$

Stability. Conditional.

This scheme is also almost never used in practice since it is conditionally stable and the the computation of u_h^{n+1} will require the solution of an EVI of the *second kind* in V_h (in general by an iterative method) even if the matrix corresponding to $(\cdot,\cdot)_h$ is *diagonal*.

2. **Implicit scheme.**

$$\begin{cases} \left(\frac{u_h^{n+1}-u_h^n}{\Delta t}, v_h - u_h^{n+1}\right)_h + a_h((n+1)\Delta t; u_h^{n+1}, v_h - u_h^{n+1}) \\ \qquad + j_h(v_h) - j_h(u_h^{n+1}) \\ \geq \left(f_h^{n+1}, v_h - u_h^{n+1}\right) \ \forall v_h \in V_h, u_h^{n+1} \in V_h, \\ n = 0, 1, 2, \ldots, u_h^0 = u_{0h}. \end{cases} \tag{5.2}$$

Stability.Unconditional.

At each *step* we have to solve an EVI of the *second kind* in V_h to compute u_h^{n+1}.

3. **Cranck-Nicholson scheme.**

$$\begin{cases} (\frac{u_h^{n+1}-u_h^n}{\Delta t}, v_h - u_h^{n+1/2})_h + a_h(n+\frac{1}{2})\Delta t; \\ \qquad u_h^{n+1/2} + j_h(v_h) - j_h(u_h^{n+1/2}) \\ \geq f_h^{n+1/2}, v_h - u^{n+1/2})_h \ \forall v_h \in V_h, u_h^{n+1/2} \in V_h, u_h^{n+1/2} = \frac{u_h^n + u_h^{n+1}}{2}, \\ n = 0, 1, 2, \ldots, u_h^0 = u_{0h}. \end{cases} \tag{5.3}$$

Stability. Unconditional.

123 Since $\dfrac{u_h^{n+1} - u_h^n}{\Delta t} = \dfrac{u_h^{n+1/2} - u_h^n}{\frac{\Delta t}{2}}$ we observe that at *each step* we have to solve an EVI of the *second kind* to compute $u_h^{n+1/2}$. If the regularity in time of the solution is poor we do not recommend this scheme.

4. **Two steps implicit scheme.**

$$\begin{cases} \frac{\frac{3}{2}u_h^{n+1} - 2u_h^n + \frac{1}{2}u_h^{n-1}}{\Delta t}, v_h - u_h^{n+1})_h + j_h(v_h) - j_h(u_h^{n+1}) \\ \qquad + a_h((n+1)\Delta t; u_h^{n+1}, v_h - u_h^{n+1}) \\ \geq (f_h^{n+1}, v_h - u_h^{n+1}) \ \forall v_h \in V_h, \\ n = 0, 1, 2, \ldots, ; u_h^0 = u_{0h}, u_h^1 \text{ given } . \end{cases} \tag{5.4}$$

We use one of above schemes (5.1)-(5.3) to compute u_h^1, starting from $u_h^0 = u_{0h}$. *Stability. Unconditional.*

We have to solve *at each step* an EVI of the *second kind* in V_h to compute u_h^{n+1}. Remark 3.4 applies for this scheme also.

Comments. The properties of stability and convergence of the various schemes of Sec. 4, 5 are studied in the references given in Sec. 1. In some cases error estimates also have been obtained.

In FORTIN [1], G.L.T [2, Chap. 6], applications to more complicated PVI than (2.1), (2.2) are also given. For the numerical analysis of hyperbolic variational inequalities see G.L.T[2, Chap.6], TREMOLIERES [1].

6 Application to a Specific Example: Time Dependent Flow of a Bingham Fluid in a Cylindrical Pipe

Following GLOWINSKI [4], we consider the time dependent problem associated to the EVI of Chap. 2. 6, and study its asymptotic properties.

6.1 Formulation of the problem. Existence and uniqueness Theorem

Let Ω be a bounded domain of $\mathbb{R}^2$ with a smooth boundary Γ. We consider:

- $V = H_0^1(\Omega) H = L^2(\Omega)$, $V^* = H^{-1}(\Omega)$,
- $a(u,v) = \int_\Omega \Delta\nabla u.\Delta\nabla v dx$,
- A time interval $[0,T]$, $0 < T < \infty$, **124**
- $f \in L^2(0,T;V^*)$, $u_0 \in H$,
- $j(v) = \int_\Omega |\nabla v| dx$,
- $\mu > 0$, $g > 0$.

We have then the following

THEOREM 6.1. *The PVI*

$$\begin{cases} (\frac{\partial u}{\partial t}, v-u) + \mu a(u, v-u) + g\,j(v) \geq (f, v-u)\ \forall v \in V\ a.e\ t \in]0, T[, \\ u(x,0) = u_0(x), \end{cases} \tag{6.1}$$

has a unique solution u such that

$$\begin{cases} u \in L^2(0,T;V) \cap C^0([0,T];H), \\ \frac{\partial u}{\partial t} \in L^2(0,T,V^*) \end{cases}$$

and this $\forall u_0 \in H$, $\forall f \in L^2(0,T;V^*)$. *For a proof of this see LIONS-DUVAUT [1, Chap.6].*

6.2 The asymptotic behaviour of the continuous solution.

Assume that if f is independent of t and that $f \in L^2(\Omega)$. We consider the following stationary problem

$$\begin{cases} \mu a(u, v-u) + g\,j(v) - g\,j(u) \geq (f, v-u)\ \forall v \in V, \\ u \in V. \end{cases} \tag{6.2}$$

It is proved in LIONS-DUVAUT [1, Chap.6] (see also Chap.2, Sec.6 of these notes), that

$$u \equiv 0 \text{ if } g\beta \geq \|f\|_{L}{}^{2}(\Omega), \tag{6.3}$$

where

$$\beta = \inf_{v \in V} \frac{j(v)}{\|v\|_{L^2(\Omega)}}. \tag{6.4}$$

125 Then we can prove the following

THEOREM 6.2. *Assume that $f \in L^2(\Omega)$ with $\|f\|_{L^2(\Omega)} < \beta g$, then if u is the solution of* (6.1)*, we have*

$$u(t) = 0 \text{ for } t \geq \frac{1}{\lambda_0 \mu} \log(1 + \lambda_0 \mu \frac{\|u_0\|_{L^2}}{\beta g - \|f\|_{L^2}}) \tag{6.5}$$

where λ_0 is the smallest eigenvalue of $-\Delta$ in $H_0^1(\Omega)(\lambda_0 > 0)$.

Proof. We use $|\cdot|$ for the $L^2(\Omega)$ -norm and $\|\cdot\|$ for the $H_0^1(\Omega)$-norm. Since $f \in L^\infty(\mathbb{R}^+, L^2(\Omega))$ it follows from Theorem 6.1 that the solution of (6.1) is defined on the whole of $\mathbb{R}^+$. □

We observe now that if $g\beta > |f|$ the zero is the unique solution of (6.2); if follows then from Theorem 6.1 that if $u(t_0) = 0$ for some $t_0 \geq 0$ then

$$u(t) = 0 \forall t \geq t_0. \tag{6.6}$$

Taking $v = 0$ and $v = 2u$ in (6.1) we obtain

$$(\frac{\partial u}{\partial t}, u) + \mu a(u,u) + g\, j(u) = (f,u) \; a.e. \text{ in } t. \tag{6.7}$$

But since $v \in L^2(0,T;V)$, $v' \in L^2(0,T,V^*)$ implies (this is a general result) that $t \to |v(t)|^2$ is *absolutely continuous* with $\frac{d}{dt}|v|^2 = 2(\frac{dv}{dt}, v)$; we obtain from (6.7) that

$$\begin{cases} \frac{1}{2}\frac{d}{dt}|u|^2 + \mu a(u,u) + g\, j(u) & = (f,u) \\ & \leq |f| \cdot |u| \; a.e \text{ in } t. \end{cases} \tag{6.8}$$

Since $a(v,v) \geq \lambda_0 |v|^2 \; \forall v \in V$, and $j(v) \geq \beta|v| \; \forall v \in V$ (from(6.4)), we obtain from (6.8) that

$$\frac{1}{2}\frac{d}{dt}|u|^2 + \mu\lambda_0 |u|^2 + (g\beta - |f|)|u| \; a.e \text{ in } t \in \mathbb{R}^+. \tag{6.9}$$

Assume that $u(t) \neq 0 \forall t \geq 0$; since $t \to |u(t)|^2$ is absolutely continuous with $|u(t)| > 0$ it follows that $\to |u(t)|$ is also absolutely continuous. Therefore (6.9) we obtain

$$\frac{d}{dt}|u(t)| + \mu\lambda_0 |u(t)| + (g\beta - |f|) \leq 0 \; a.e \; t \in \mathbb{R}^+. \tag{6.10}$$

It follows from (6.10) that **126**

$$\frac{\frac{d}{dt}|u(t)|}{|u(t)| + \frac{g\beta - |f|}{\mu\lambda_0}} \leq -\mu\lambda_0 \; a.e. \; t \in \mathbb{R}^+. \tag{6.11}$$

Define γ by $\gamma = \dfrac{g\beta - |f|}{\mu\lambda_0}$, then $\gamma > 0$. It follows then by integrating (6.11) that

$$|u(t)| + \gamma \leq (|u_0| + \gamma)e^{-\mu\lambda_0 t} \ \forall t \in \mathbb{R}^+; \tag{6.12}$$

(6.12) is absurd for t large enough. Actually we have $u(t) = 0$ if

$$-\gamma \geq (|u_0| + \gamma)e^{-\mu\lambda_0 t},$$

i. e.

$$t \geq \frac{1}{\lambda_0\mu} \log(1 + \frac{\lambda_0\mu\|u_0\|_{L^2(\Omega)}}{g\beta - \|f\|_{L^2(\Omega)}}). \tag{6.13}$$

Exercise 6.1. *Let $f \in L^2(\Omega)$ with possibly $|f| \geq g \cdot \beta$. Let us denote by u_∞ the solution of* (6.2)*; theorem prove that*

$$|u(t) - u_\infty| \leq |u_0 - u_\infty|e^{-\lambda_0\mu t}$$

where $u(t)$ is the solution of (6.1) .

6.3 On the asymptotic behaviour of the discrete solution.

We still assume that $f \in L^2(\Omega)$. To approximate (6.1) we proceed as follows : assuming that Ω is a polygonal domain, we use the same approximation with regard to the space variables as in Chap. 2, Sec. 6 (i.e. by means of piecewise linear finite elements, see Chap. 2, Sec. 6). Hence we have

$$\begin{cases} a_h(u_h, v_h) = a(u_h, v_h) \ \forall u_h, v_h \in V_h, \\ j_h(v_h) = j(v_h) \ \forall v_h \in V_h, \end{cases}$$

127 and from the formulae of Chap. 2, Sec. 7 we can also take

$$(u_h, v_h)_h = (u_h, v_h) \ \forall u_h, v_h \in V_h.$$

Then we approximate (6.1) by the *implicit scheme* (5.2) and we obtain

$$\begin{cases} (\frac{u_h^{n+1} - u_h^n}{\Delta t}, v_h - u_h^{n+1}) + \mu \int_\Omega \nabla u_h^{n+1} \cdot \nabla(v_h - u_h^{n+1})dx + j(v_h) - j(u_h^{n+1}) \\ \geq (f_h, v_h - u_h^{n+1}) \forall v_h \in V_h, u_h^{n+1} \in V_h; n = 0, 1, 2, \ldots, ; u_h^0 = u_{0h}. \end{cases} \tag{6.14}$$

We assume that $u_{0h} \in V_h \ \forall h$ and

$$\lim_{h\to 0} u_{0h} = u_o \text{ strongly in } L^2(\Omega). \tag{6.15}$$

Similarly we assume that f is approximated by $(f_h)_h$ in such a way that (f_h, v_h) can be computed easily and

$$\lim_{h\to 0} f_h = f \text{ strongly in } L^2(\Omega). \tag{6.16}$$

THEOREM 6.3. *Let $|f| < \beta g$. If* (6.15) *and* (6.16) *hold, then if h is sufficiently small we have $u_h^n = 0$ for n large enough.*

Proof. As in the proof of Theorem 6.2, taking $v_h = 0$ and $v_h = 2u_h^{n+1}$ in (6.14) we obtain

$$\left(\frac{u_h^{n+1} - u_h^n}{\Delta t}, u_h^{n+1}\right) + \mu \int_\Omega |\nabla u_h^{n+1}|^2 dx + g \int_\Omega |\nabla u_h^{n+1} dx = \int_\Omega f_h u_h^{n+1} dx \ \forall n \geq 0; \tag{6.17}$$

using Schwarz inequality in $L^2(\Omega)$, if follows from (6.17) that

$$\frac{|u_h^{n+1}| - |u_h^n|}{\Delta t}|u_h^{n+1}| + \mu\lambda_0 |u_h^{n+1}|^2 + (g\beta - |f_h|)|u_h^{n+1}| \leq 0 \ \forall n \geq 0. \tag{6.18}$$

□

Since $f_h \to f$ *strongly* in $L^2(\Omega)$ we have

$$g\beta - |f_h| > 0 \text{ for } h \text{ sufficiently small .} \tag{6.19}$$

It follows then from (6.18), (6.19) that

$$u_h^{n_0} = 0 \Rightarrow u_h^n = 0 \text{ for } n \geq n_0 \text{ if } h\text{is small enough.} \tag{6.20}$$

Assume that $u_h^n \neq 0 \forall n$; then (6.18) implies **128**

$$\frac{|u_h^{n+1}| - |u_h^n|}{\Delta t} + \mu\lambda_0 |u_h^{n+1}| + g\beta - |f_h| \leq 0 \forall n \geq 0. \tag{6.21}$$

We define γ_h by $\gamma_h = g\beta - |f_h|$ then,

$$\gamma_h > 0 \text{ for } h \text{ small enough and } \lim_{h\to 0} \gamma_h = \gamma = g\beta - |f|. \tag{6.22}$$

It follows follows from (6.21) that

$$\left(|u_h^{n+1}| + \frac{\gamma_h}{\lambda_0\mu}\right)(1 + \lambda_0\mu\Delta t) \leq |u_h^n| + \frac{\gamma_h}{\lambda_0\mu} \ \forall n \geq 0$$

which implies that

$$\left(|u_h^n| + \frac{\gamma_h}{\lambda_0\mu}\right) \leq (1 + \lambda_0\mu\Delta t)^{-n}\left(|u_h^0| + \frac{\gamma_h}{\lambda_0\mu}\right). \tag{6.23}$$

Since $\gamma_h > 0$ for h small enough, (6.23) is impossible for n large enough. More precisely we shall have $u_h^n = 0$ if

$$\frac{\gamma_h}{\lambda_0\mu} \geq (1 + \lambda_0\mu\Delta t)^{-n}\left(|u_h^0| + \frac{\gamma_h}{\lambda_0\mu}\right),$$

which implies:

$$\text{If } h \text{ is small enough, then } u_h^n = 0 \text{ if } n \geq \frac{\log(1 + \lambda_0\mu\frac{|u_h^0|}{\gamma_h})}{\log(1 + \lambda_0\mu\Delta t).} \tag{6.24}$$

Relation (6.24) makes the statement of Theorem 6.3 more precise. Moreover in terms of *time*, (6.24) implies that u_h^n is equal to zero if

$$n\Delta t \geq \Delta t\frac{\log(1 + \lambda_0\mu\frac{|u_h^0|}{\gamma_h})}{\log(1 + \lambda_0\mu\Delta t)}. \tag{6.25}$$

We observe that

$$\lim_{\substack{h\to 0\\ \Delta t\to 0}} \Delta t\frac{\log(1 + \lambda_0\mu\frac{|u_h^0|}{\gamma_h})}{\log}\left(1 + \lambda_0\mu\Delta t = \frac{1}{\lambda_0\mu}\log(1 + \lambda_0\mu)\frac{|u^0|}{\gamma}\right)$$

129

Hence taking the limit in (6.25) we obtain another proof (assuming that u_h^n converges to u in some topology) of the estimate (6.5) given in the statement of Theorem 6.2.

Exercise 6.2. *Let u_h^∞ be the solution of the time independent problem associated to f_h, possibly with $|f_h| \geq \beta \cdot g$, then prove that*

$$|u_h^n - u_h^\infty| \leq (1 + 2\mu\lambda_0\Delta t)^{-n/2}|u_h^0 - u_h^\infty| \quad n \geq 0.$$

6.4 Remarks

REMARK 6.1. *We can generalize Theorem 5.1 to the case of a Bingham flow in a 2-dimensional bounded cavity.*

REMARK 6.2. *In GLOWINSKI [5], BRISTEAU[1], BEGIS[1], numerical verification of the above asymptotic properties have been performed and found to be consistent with the theoretical predictions.*

REMARK 6.3. *One may find in H.BREZIS [5], many results on the asymptotic behaviour of various PVI as* $t \to \infty$.

Chapter 4

Applications of elliptic variational Inequality methods to the solution of some nonlinear elliptic equations

1 Introduction

For solving some non-linear elliptic equations it may be convenient, from the theoretical and numerical points of view, to see them as EVI's. **130**

We shall consider in this chapter two examples of such situations:

(1) A family of mildly non-linear elliptic equations,

(2) A non-linear elliptic equation modelling the subsonic flow of a perfect compressible fluid.

2 Theoretical and Numerical Analysis of Some Mildly Non-Linear Elliptic Equations

2.1 Formulation of the continuous problem

Let Ω be a bounded domain of $\mathbb{R}^N (N \geq 2)$ with a smooth boundary Γ. We consider

- $V = H_0^1(\Omega)$,
- $L(v) =< f, v >,\ f \in V^* = H^{-1}(\Omega)$;
- $a : V \times V \to \mathbb{R}$ bilinear, continuous and V-elliptic with $\alpha > 0$ as ellipticity constant; $a(\cdot, \cdot)$ is possibly not symmetric;
- $\phi : \mathbb{R} \to \mathbb{R}$, $\phi \in C^0(\mathbb{R})$, non-decreasing with $\phi(0) = 0$.

We then consider the following non-linear elliptic equation (P) defined by :

Find $u \in V$ such that

$$\begin{cases} a(u, v) + < \phi(u), v >= L(v)\ \forall v \in V, \\ \phi(u) \in L^1(\Omega) \cap H^{-1}(\Omega). \end{cases} \tag{P}$$

It follows from the *Riesz representation Theorem* that there exists $A \in \mathscr{L}(V, V^*)$ such that $a(u, v) =< Au, v >\ \forall u, v \in V$. Therefore (P) is equivalent to

$$\begin{cases} Au + \phi(u) = f, \\ u \in V, \\ \phi(u) \in L^1(\Omega) \cap H^{-1}(\Omega). \end{cases} \tag{2.1}$$

131

Example 1. *Let us consider a function $a_0 \in L^\infty(\Omega)$ such that*

$$a_0(x) \geq \alpha > 0 \text{ a. e. in } \Omega. \tag{2.2}$$

Define $a\ (\cdot, \cdot)$ by

$$a(u, v) = \int_\Omega a_0(x) \nabla u \cdot \nabla v dx + \int_\Omega \beta \cdot \nabla u v dx \tag{2.3}$$

where β is a constant *vector in $\mathbb{R}^N$.*

From the definition of $a_0(\cdot)$ and using the fact that $\int_\Omega \beta \cdot \nabla v v dx = 0 \; \forall v \in H_0^1(\Omega)$, we clearly have

$$a(v, v) \geq \alpha \|v\|_V^2. \tag{2.4}$$

From (2.3) we obtain

$$Au = -\nabla \cdot (a_0 \nabla u) + \beta \cdot \nabla_u. \tag{2.5}$$

Hence, in this particular case, (2.1) becomes

$$\begin{cases} -\nabla \cdot (a_0 \nabla u) + \beta \cdot \nabla u + \phi(u) = f, \\ u \in V, \phi(u) \in L^1(\Omega). \end{cases} \tag{2.6}$$

REMARK 2.1. *If $N = 1$, we have $H_0^1(\Omega) \subset C^0(\overline{\Omega})$. Because of this inclusion there is no great difficulty in the study of one-dimensional problems of type (P). If $N \geq 2$ the main difficulty is precisely related to the fact that $H_0^1(\Omega)$ is not contained in $C^0(\overline{\Omega})$.*

REMARK 2.2. *The analysis given below may be extended to problems in which either $V = H^1(\Omega)$ or V is a convenient closed subspace of $H^1(\Omega)$.*

2.2 A variational inequality related to (P)

2.2.1 Definition of the variational inequality

132

Let

$$\Phi(t) = \int_0^t \phi(\tau) d\tau, \tag{2.7}$$

$$D(\Phi) = \{v \in V : \Phi(v) \in L^1(\Omega)\}. \tag{4.1}$$

The functional $j : L^2(\Omega) \to \overline{\mathbb{R}}$ is defined by

$$j(v) = \int_\Omega \Phi(v) dx \text{ if } \Phi(v) \in L^1(\Omega), \; j(v) = +\infty \text{ if } \Phi(v) \notin L^1(\Omega). \tag{4.2}$$

Instead of studying the problem (P) directly, it is natural to associate to (P) the following EVI of the second kind:

$$\begin{cases} a(u, v-u) + j(v) - j(u) \geq L(v-u) \ \forall v \in V, \\ u \in V. \end{cases} \tag{π}$$

If $a(\cdot,\cdot)$ is symmetric, a standard method to study (P) is to consider it as the *formal Euler equation* of the following minimization problem encountered in the Calculus of Variations

$$\begin{cases} J(u) \leq J(v) \ \forall v \in V, \\ u \in V, \end{cases} \tag{2.10}$$

where $J(v) = \frac{1}{2}a(v,v) + \int_\Omega \Phi(v)dx - L(v)$.

Exercise 2.1. *Prove that $D(\Phi)$ is a convex, non-empty subset of V.*

2.2.2 Properties of $j(\cdot)$.

Since $\phi : \mathbb{R} \to \mathbb{R}$ is non-decreasing and continuous with $\phi(0) = 0$, we have

$$\Phi \in C^1(\mathbb{R}), \Phi \text{ convex }, \Phi(0) = 0; \Phi(t) \geq 0 \ \forall t \in \mathbb{R}. \tag{2.12}$$

The properties of $j(\cdot)$ are given by the following

Lemma 2.1. *The functional $j(\cdot)$ is convex, proper and 1.s.c. over $L^2(\Omega)$.*

133 *Proof.* Since $j(v) \geq 0 \ \forall \ \in L^2(\Omega)$ it following that $j(\cdot)$ is proper. The convexity of $j(\cdot)$ is obvious from the fact that Φ is convex.

Let us prove that $j(\cdot)$ is l.s.c. Let $(v_n)_n$, $v_n \in L^2(\Omega)$ be such that

$$\lim_{n\to\infty} v_n = v \text{ strongly in } L^2(\Omega).$$

Then we have to prove that

$$\liminf_{n\to\infty} j(v_n) \geq j(v). \tag{2.13}$$

□

If $\liminf_{n\to\infty} j(v_n) = +\infty$ the property is proved. Therefore assume that $\liminf_{n\to\infty} j(v_n) = \ell < \infty$. Hence we can extract a subsequence $(v_{n_k})_{nk}$ such that

$$\lim_{k\to\infty} j(v_{n_k}) = \ell, \tag{2.14}$$

$$v_{n_k} \to v \text{ a. e. in } \Omega. \tag{2.15}$$

Since $\Phi \in C^1(\mathbb{R})$, (2.15) implies

$$\lim_{k\to\infty} \Phi(v_{n_k}) = \Phi(v) \text{ a.e..} \tag{2.16}$$

Moreover $\Phi(v) \geq 0$ a.e and (2.14) implies that

$$\{\Phi(v_{n_k})\}_k \text{ is bounded in } L^1(\Omega). \tag{2.17}$$

Hence by Fatou's Lemma, from (2.16) and (2.17), we have

$$\begin{cases} \Phi(v) \in L^1(\Omega), \\ \liminf_{k\to\infty} \int_\Omega \Phi(v_{n_k})dx \geq \int_\Omega \Phi(v)dx. \end{cases} \tag{2.18}$$

From (2.14) and (2.18) we obtain (2.13). This proves the lemma.

COROLLARY 2.1. *The functional $j(\cdot)$ restricted to V is convex, proper, l,s.c.*

2.2.3 Existence and uniqueness results for (π) :

THEOREM 2.1. *Under the above hypothesis on V, $a(\cdot,\cdot)$, $L(\cdot)$, $\phi(\cdot)$ the problem (π) has a unique solution in $V \cap D(\phi)$.* **134**

Proof. Since V, $a(\cdot,\cdot)$, $L(\cdot)$, $j(\cdot)$ have the properties (cf. Corollary 2.1) required to apply Theorem 4.1 of Chap. 1, Sec. 4, the EVI of the second kind (π), has a unique solution u in V. □

Let us show that $u \in D(\Phi)$. Taking $v = 0$ in (π) we obtain

$$a(u,u) + j(u) \leq L(u) \leq \|f\| \cdot \|u\|_V. \tag{4.3}$$

Since $j(u) \geq 0$, using the ellipticity of $a(\cdot,\cdot)$ we obtain

$$\|u\|_V \leq \frac{\|f\|}{\alpha}, \tag{2.20}$$

which implies

$$j(u) \leq \frac{\|f\|^2}{\alpha}. \tag{2.21}$$

This implies $u \in D(\Phi)$.

REMARK 2.3. *If $a(\cdot,\cdot)$ is symmetric, (π) is equivalent to* (2.10).

2.3 Equivalence between (P) and (π)

In this section we shall prove that (P) and (π) are equivalent. First we prove that the unique solution of (π) is also a solution of (P). In order to prove this result we need to prove that $\phi(u)$ and $u\phi(u)$ belong to $L^1(\Omega)$.

Proposition 2.1. *Let u be the solution of (π). Then $u\phi(u)$ and $\phi(u)$ belong to $L^1(\Omega)$.*

Proof. Here we use a *truncation* technique. Let n be a positive integer. Define

$$K_n\{v \in V : |v(x)| \leq n a.e.\}.$$

Since K_n is a closed, convex, non-empty subset of V, the following variational inequality

$$\begin{cases} a(u^n, v - u^n) + j(v) - j(u^n) \geq L(v - u^n)\ \forall v \in K_n, \\ u^n \in K_n \end{cases} \tag{π_n}$$

135 has a unique solution (in order to apply Theorem 4.1 of Chapter 1, we need to replace j by $j + I_{k_n}$ where I_k is the *indicator functional* of K_n).

Now we prove that $\lim_{n\to\infty} u_n = u$ *weakly* in V, where u is the solution of (π). Since $0 \in K_n$, taking $v = 0$ in (π_n) we obtain as in Theorem 2.1 of this chapter that

$$\|u_n\|_V \leq \frac{\|f\|}{\alpha}. \tag{2.22}$$

$$j(u_n) \leq \frac{\|f\|^2}{\alpha}. \tag{2.23}$$

It follows from (2.22) that there exists a subsequence $\{u_{n_k}\}$ of $(u_n)_n$ and $au^* \in V$ such that

$$\lim_{k\to\infty} u_{n_k} = u^* \text{ weakly in } V. \tag{2.24}$$

Moreover, from the compactness of the canonical injection from $H_0^1(\Omega)$ to $L^2(\Omega)$ and form (2.24), it follows that

$$\lim_{k\to\infty} u_{n_k} = u^* \text{ strongly in } L^2(\Omega). \tag{2.25}$$

Relation (2.25) implies that we can extract a subsequence, still denoted by $(u_{n_k})_{n_k}$, such that

$$\lim_{k\to\infty} u_{n_k} = u^* \text{ a. e. in } \Omega. \tag{2.26}$$

Now let $v \in V \cap L^\infty(\Omega)$; then, large k, have $v \in K_{n_k}$ and

$$a(u_{n_k}, u_{n_k}) + j(u_{n_k}) \leq a(u_{n_k}, v) + j(v) - L(v - u_{n_k}). \tag{2.27}$$

Since $\liminf_{k\to\infty} a(u_{n_k}, u_{n_k}) \geq a(u^*, u^*)$ and $\liminf_{k\to\infty} j(u_{n_k}) \geq j(u^*)$ it follows from (2.24) and (2.27) that **136**

$$\begin{cases} a(u^*, u^*) + j(u^*) \leq (u^*, v) + j(v) - L(v - u^*)\ \forall v \in L^\infty(\Omega) \cap V, \\ u^* \in V, \end{cases}$$

which can also be written as

$$\begin{cases} a(u^*, v - u^*) + j(v) - j(u^*) \geq -L(v - u^*)\ \forall v \in V \cap L^\infty(\Omega), \\ u^* \in V. \end{cases} \tag{2.28}$$

For $n > 0$, define $\tau_n : V \to K_n$ by

$$\tau_n v = \inf(n, \text{Sup}(-n, v))(\text{ see Figure 2.1}) \tag{2.29}$$

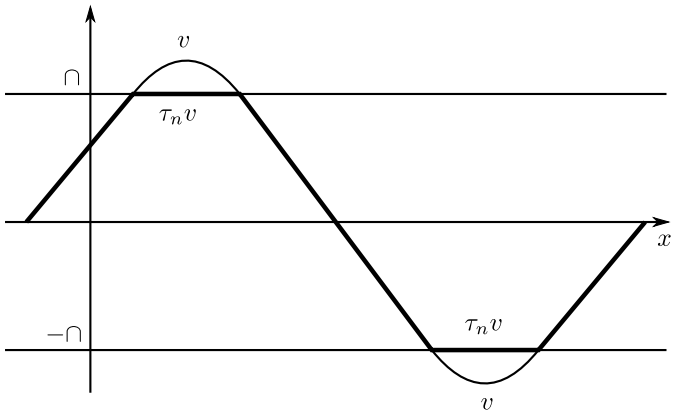

Figure 2.1:

Then the Corollary 2.1 of Chap. 2, Sec. 2.2, we have

$$\begin{cases} \lim_{n\to\infty} \tau_n v = v \text{ strongly in } V, \\ \lim_{n\to\infty} \tau_n v = v \text{ in } \Omega. \end{cases} \tag{2.30}$$

Moreover, we obviously have,

$$|\tau_n v(x)| \leq |v(x)| \text{ a.e.}, \tag{2.31}$$

$$v(x) \cdot \tau_n v(x) \geq 0 \text{ a.e.} \tag{2.32}$$

137 It follows then from (2.30)-(2.32) and from the various properties of that

$$\Phi(\tau_n v) \leq \Phi(v) \text{ a.e.}, \tag{2.33}$$

$$\lim_{n\to\infty} \Phi(\tau_n v) = \Phi(v) \text{ a.e.} \tag{2.34}$$

Since $\tau_n v \in L^\infty(\Omega) \cap V$ it follows form (2.28) that

$$\begin{cases} a(u^*, \tau_n v - u^*) + j(\tau_n v) - j(u^*) \geq L(\tau_n v - u^*) \ \forall v \in V, \\ u^* \in V. \end{cases} \tag{2.35}$$

If $v \notin D(\Phi)$, then by Fatou's lemma

$$\lim_{n\to\infty} j(\tau_n v) = +\infty.$$

If $v \in D(\Phi)$, it follows from (2.33) and (2.34) by applying Lebesgue's dominated convergence theorem that

$$\lim_{n\to\infty} j(\tau_n v) = j(v).$$

From these convergence properties and from (2.30), it follows, by taking the limit in (2.35), that

$$\begin{cases} a(u^*, v - u^*) + j(v) - j(u^*) \geq L(v - u^*) \ \forall v \in V, \\ u^* \in V. \end{cases} \tag{2.36}$$

Then u^* is a solution of (π) and from the uniqueness property we have $u^* = u$. This proves that $\lim_{n\to\infty} u_n = u$ *weakly* in V.

Let us know that $\phi(u)$, $u\phi(u) \in L^1(\Omega)$. Let $v \in K_n$. Then $u_n + t(v - u_n) \in K_n \ \forall t \in]0, 1]$. Replacing v by $u_n + t(v - u_n)$ in π_n and dividing both **138**
sides of the inequality by t we obtain

$$a(u_n, v - u_n) + \int_\Omega \frac{\Phi(u_n + t(v - u_n)) - \Phi(u_n)}{t} dx \geq L(v - u_n) \ \forall v \in K_n. \tag{2.37}$$

Since $\Phi \in C^1(R)$ and $\Phi' = \phi$ we have

$$\lim_{\substack{t\to 0 \\ t>0}} \frac{\Phi(u_n + t(v - u_n)) - \Phi(u_n)}{t} = \phi(u_n) \cdot (v - u_n) \text{ a.e.} \tag{2.38}$$

Moreover since Φ is convex, we also have $\forall t \in]0, 1]$,

$$\phi(u_n)(v - u_n) \leq \frac{\phi(u_n + t(v - u_n)) - \phi(u_n)}{t} \leq \Phi(v) - \Phi(u_n) \text{ a.e.} \tag{2.39}$$

From (2.38), (2.39) and using *Lebesgue's dominated convergence Theorem* in (2.37), we obtain

$$a(u_n, v - u_n) + \int_\Omega \phi(u_n)(v - u_n) dx \geq L(v - u_n) \quad \forall v \in K_n. \tag{2.40}$$

Then taking $v = 0$ in (2.40) we have

$$a(u_n, u_n) + \int_\Omega \phi(u_n)u_n dx \leq L(u_n),$$

which implies, using (2.2),

$$\int_\Omega \phi(u_n)u_n dx \leq \frac{\|f\|^2}{\alpha}. \tag{2.41}$$

But $\phi(v)v \geq 0\ \forall v \in V$. Hence $\phi(u_n)u_n$ is bounded in $L^1(\Omega)$. Moreover for some subsequence $(u_{n_k})_{n_k}$ of $(u_n)_n$ we have

$$\phi(u_{n_k})u_{n_k} \to \phi(u)u \text{ a. e. in } \Omega.$$

Then by Fatou's lemma we obtain $u\phi(u) \in L^1(\Omega)$ and this completes the proof of the Proposition since $u\phi(u) \in L^1(\Omega)$ implies obviously that $\phi(u) \in L^1(\Omega)$. □

139 Incidentally, when proving the convergence of $(u_n)_n$ to u, we have proved the following useful

Lemma 2.2. The solution u of (π) is characterised by

$$\begin{cases} a(u, v-u) + j(v) - j(u) \geq L(v-u)\ \forall v \in V \cap L^\infty(\Omega), \\ u \in V, \Phi(u) \in L^1(\Omega). \end{cases} \tag{2.42}$$

In view of proving that (π) *implies* (P) *we also need the following two lemmas:*

Lemma 2.3. The solution u of (π) is characterised by

$$\begin{cases} a(u, v-u) + \int_\Omega \phi(u)(v-u)dx \geq L(v-u)\ \forall v \in L^\infty(\Omega) \cap V, \\ u \in V, u\phi(u) \in L^1(\Omega). \end{cases} \tag{2.43}$$

Proof. (π) *implies* (2.43).

Let $v \in L^\infty(\Omega) \cap V$. Then $v \in D(\Phi)$ and since $D(\Phi)$ is convex we have $u + t(v-u) \in D(\Phi)\ \forall t \in]0, 1]$. Replacing v by $u + t(v-u)$ in (π) and dividing by t we obtain $\forall t \in]0, 1]$

$$a(u, v-u) + \int_\Omega \frac{\Phi(u + t(v-u)) - \Phi(u)}{t} dx \geq L(v-u),\ \forall v \in L^\infty(\Omega) \cap V. \tag{2.44}$$

Since $\Phi \in C^1$ and is convex, we have

$$\lim_{\substack{t\to 0\\ t>0}} \frac{\Phi(u + t(v-u)) - \Phi(u)}{t} = \phi(u)(v-u) \text{ a.e.,} \tag{2.45}$$

$$\phi(u)(v-u) \leq \frac{\Phi(u + t(v-u)) - \Phi(u)}{t} \leq \Phi(v) - \Phi(u). \tag{2.46}$$

□

By Proposition 2.1 we have $\phi(u), u\phi(u) \in L^1(\Omega)$. Hence $\phi(u)(v-u) \in L^1(\Omega)$ and $\Phi(v), \Phi(u) \in L^1(\Omega), \forall v \in L^\infty(\Omega) \cap V$. Then using the lebesgue dominated convergence Theorem it follows from (2.45) and (2.46) that

$$\lim_{t\to 0} \int_\Omega \frac{\Phi(u + t(v-u)) - \Phi(u)}{t} dx = \int_\Omega \phi(u)(v-u)dx.$$

Using the above relation and (2.44) we obtain (2.43). This proves that **140**
$(\pi) \Rightarrow$ (2.43).

(2) We will now prove that (2.43) $\Rightarrow (\pi)$.

Let u be a solution of (2.43). Since Φ is convex it follows that

$$-\Phi(u) = \Phi(0) - \Phi(u) \geq \phi(u)(0-u) = -\phi(u)u.$$

This implies $0 \leq \Phi(u) \leq u\phi(u)$ and $\Phi(u) \in L^1(\Omega)$. Let $v \in L^\infty(\Omega) \cap V$. Then from the inequality.

$$\phi(u)(v-u) \leq \Phi(v) - \Phi(u) \text{ a. e. in } \Omega,$$

we obtain by integration

$$\int_\Omega \phi(u)(v-u)dx \leq j(v) - ju \ \forall v \in V \cap L^\infty(\Omega),$$

which combined with (2.43) and $\Phi(u) \in L^1(\Omega)$ implies (2.42). Hence from Lemma 2.2 we obtain that (2.43) implies (π).

Lemma 2.4. Let u be the solution of (π). Then u is characterised by

$$\begin{cases} a(u,v) + \int_\Omega \phi(u)vdx = L(v) \ \forall v \in L^\infty(\Omega) \cap V, \\ u \in V, \phi(u) \in L^1(\Omega). \end{cases} \tag{2.47}$$

Proof. (1) (π) implies (2.47).

Let $v \in V \cap L^{\infty}(\Omega)$. If u is the solution of (π) then u is also the unique solution of (2.43). Let τ_n be defined by (2.29). Then $\tau_n u \in V \cap L^{\infty}(\Omega)$. Replacing v by $\tau_u u + v$ in (2.43) we obtain

$$\begin{cases} a(u,v) + \int_{\Omega} \phi(u) v dx + a(u, \tau_n u - u) + \int_{\Omega} \phi(u)(\tau_n u - u) dx \\ \qquad\qquad \geq L(v) + L(\tau_n u - u), \\ \forall v \in V \cap L^{\infty}(\Omega). \end{cases} \tag{2.48}$$

141 It follows from (2.29)–(2.32) that

$$\begin{cases} \lim\limits_{n\to\infty} a(u, \tau_n u - u) = 0, \\ \lim\limits_{n\to\infty} L'(\tau_n u - u) = 0, \end{cases} \tag{2.49}$$

$$\lim_{n\to\infty} \phi(u)(\tau_n u - u) = 0 \text{ a.e.}, \tag{2.50}$$

$$0 \leq \phi(u)(u - \tau_n u) \leq u\phi(u) \text{ ae..} \tag{2.51}$$

□

Then by the Lebesgue dominated convergence Theorem and (2.50), (2.51) we obtain

$$\lim_{n\to\infty} \phi(u)(\tau_n u - u) = 0 \text{ strongly in } L^1(\Omega). \tag{2.52}$$

Then (2.48), (2.49) and (2.52) imply

$$a(u,v) + \int_{\Omega} \phi(u) v dx \geq L(v) \ \forall v \in V \cap L^{\infty}(\Omega).$$

Since the above relation also holds for $-v$ we have

$$a(u,v) + \int_{\Omega} \phi(u) v dx = L(v) \quad \forall v \in V \cap L^{\infty}(\Omega), \tag{2.53}$$

By Proposition 2.1 we have $\phi(u) \in L^1(\Omega)$; combining this with (2.53) we obtain (2.47). This proves that (π) $\Rightarrow$ (2.47).

(2) (2.47) implies (π).
We have

$$a(u,v) + \int_\Omega \phi(u)v dx = L(v)\ \forall v \in V \cap L^\infty(\Omega).$$

then

$$a(u,\tau_n u) + \int_\Omega \phi(u)\tau_n u dx = L(\tau_n u)\ \forall n. \tag{2.54}$$

Since $\tau_n u \to u$ *strongly* in V, $\{\int_\Omega \phi(u)\tau_n u dx\}_n$ is *bounded*. But $\phi(u)\tau_n u \geq$ **142**
0 a.e. Hence we obtain that $\phi(u)\tau_n u$ is bounded in $L^1(\Omega)$. We also have $\lim_{n\to\infty} \tau_n u\phi(u) = u\phi(u)$ a.e. ; hence by Fatou's lemma we have

$$u\phi(u) \in L^1(\Omega). \tag{2.55}$$

But now we observe that

$$0 \leq \phi(u)\tau_n u \leq u\phi(u).$$

Hence by the Lebesgue dominated convergence theorem

$$\lim_{n\to\infty} \int_\Omega \phi(u)\tau_n u dx = \int_\Omega \phi(u) u dx,$$

which along with (2.54) gives

$$a(u,u) + \int_\Omega \phi(u)u dx = L(u). \tag{2.56}$$

Then by subtracting (2.56) from (2.47) we obtain

$$\begin{cases} a(u, v-u) + \int_\Omega \phi(u)(v-u)dx = L(v-u)\ \forall v \in V \cap L^\infty(\Omega), \\ u \in V, u\phi(u) \in L^1(\Omega), \end{cases} \tag{2.57}$$

and obviously (2.57) implies (2.43) . This completes the proof of the lemma.

COROLLARY 2.2. *If u is the solution of* (π) *then u is also a solution of* (P).

Proof. We recall that $V^* = H^{-1}(\Omega) \subset \mathscr{D}'(\Omega)$ and that $a(u,v) =< Au, v >$ $\forall u, v \in V$ and $L(v) =< f, v >$.

Let u be a solution of (π). Then u is characterised by (2.47) and since $\mathscr{D}(\Omega) \subset V$ we obtain

$$< Au, v > + \int_\Omega \phi(u) v dx =< f, v > \quad \forall v \in \mathscr{D}(\Omega) \tag{2.58}$$

143 From (2.58) it follows that

$$Au + \phi(u) = f \text{ in } \mathscr{D}(\Omega), \tag{2.59}$$

since Au and $f \in V^*$, we have $\phi(u) \in V^*$. Hence $\phi(u) \in L^1(\Omega) \cap H^{-1}(\Omega)$ and from (2.59) we obtain that u is a solution of (P). □

If we try to summarise what we have proved until now, we observe that the unique solution of (π) is also a solution of (P). Now we prove the reciprocal property ; that is, every solution of (P) is a solution of (π) and hence (P) has a unique solution.

In order to prove this we shall use the following density lemma :

Lemma 2.5. *$\mathscr{D}(\Omega)$ is dense in $V \cap L^\infty(\Omega)$, $V \cap L^\infty(\Omega)$ being equipped with the strong topology of V and the weak * topology of $L^\infty(\Omega)$.*

Proof. Let $v \in V \cap L^\infty(\Omega)$. Since $\overline{\mathscr{D}(\Omega)}^{H^1(\Omega)} = V$ there exists a sequence $\{v_n\}_n$, $v_n \in \mathscr{D}(\Omega)$, such that

$$\lim_{n\to\infty} v_n = v \text{ strongly in } V. \tag{2.60}$$

Let us define w_n by (see Fig. 2.2)

$$w_n = \min(v^+, v_n^+) - \min(v^-, v_n^-) \tag{2.61}$$

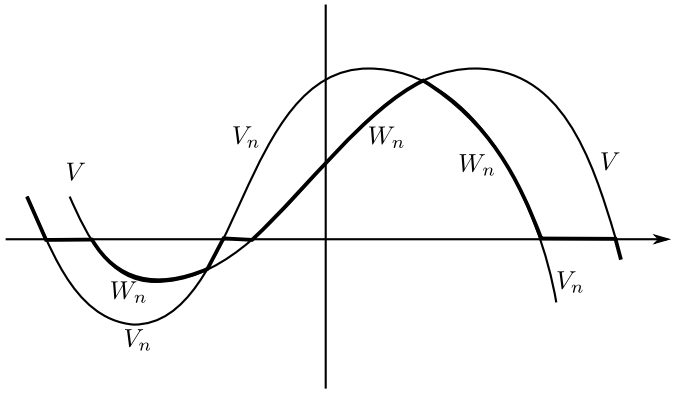

Figure 2.2:

Then **144**

$$w_n \text{ has a compact support in } \Omega, \tag{2.62}$$

$$\|w_n\|_{L^\infty(\Omega)} \leq \|v\|_{L^\infty(\Omega)} \tag{2.63}$$

and it follows from Chap. 2, Corollary 2.1, that

$$\lim_{n\to\infty} w_n = v \text{ strongly in } V. \tag{2.64}$$

□

From (2.63) and (2.64) we obtain that $\lim_{n\to\infty} w_n = v$ for the *weak * topology* of $L^\infty(\Omega)$.

Thus we have proved that

$$\mathscr{U} = \{v \in L^\infty(\Omega) \cap V : v \text{ has compact support in } \Omega\}$$

is dense in $L^\infty(\Omega) \cap V$ for the topology given in the statement of the Lemma.

Let $v \in \mathscr{U}$ and $(\rho_n)_n$ be a mollifying sequence (see Chap. 2, Lemma 2.4). Define $\tilde{v}_n$ by

$$\tilde{v}(x) = \begin{cases} v(x) & \text{if } x \in \Omega, \\ 0 & \text{if } x \notin \Omega, \end{cases} \tag{4.4}$$

$$\tilde{v}_n = \rho_n * \tilde{v}, \tag{2.66}$$

$$\text{then } \tilde{v}_n \in \mathscr{D}(\mathbb{R}^N), \lim_{n\to\infty} \tilde{v}_n = \tilde{v} \text{ strongly in } H^1(\mathbb{R}^N), \tag{2.67}$$

$$\tilde{v}_n \text{ has a compact support in } \Omega \text{ for } n \text{ large enough.} \tag{2.68}$$

Let $v_n = \tilde{v}_n|_\Omega$ then for n large enough $v_n \in \mathscr{D}(\Omega)$ and $\lim_{n\to\infty} v_n = v$ *strongly* on V.

Since $\|\tilde{v}\|_{L^\infty(\mathbb{R}^N)} = \|v\|_{L^\infty(\Omega)}$ it follows from (2.66) that

$$|v_n(x)| \le \int_{\mathbb{R}^N} \rho_n(x-y)|\tilde{v}(y)|dy \le \|v\|_{L^\infty(\Omega)} \tag{2.69}$$

145 From this it follows that

$$\|v_n\|_{L^\infty(\Omega)} \le \|v\|_{L^\infty(\Omega)} \tag{2.70}$$

Summarising the above information we have proved that $\forall v \in L^\infty(\Omega) \cap V$, there exists a sequence $(v_n)_n$, $v_n \in \mathscr{D}(\Omega)\ \forall n$, such that

$$\lim_{n\to\infty} v_n = v \text{ strongly in } V, \tag{2.71}$$

$$\|v_n\|_{L^\infty(\Omega)} \le \|v\|_{L^\infty(\Omega)}\ \forall n. \tag{2.72}$$

Hence from (2.71) and (2.72) we obtain that $v_n \to v$ in $L^\infty(\Omega)$ *weak* *. This completes the proof of the Lemma.

THEOREM 2.2. *Under the above hypothesis on V, $a(\cdot,\cdot)$, $L(\cdot)$, $\phi(\cdot)$, problems (π) and (P) are equivalent.*

Proof. We have already proved that (π) implies (P). We need only to prove that (P) implies (π).

From the definition of (P) we have

$$\begin{cases} a(u,v) + < \phi(u), v >= L(v) \ \ \forall v \in V, \\ \\ u \in V, \phi(u) \in H^{-1}(\Omega) \cap L^1(\Omega). \end{cases} \tag{2.73}$$

It follows from (2.73) that

$$a(u,v) + \int_\Omega \phi(u) v dx = L(v) \ \ \forall v \in \mathscr{D}(\Omega). \tag{2.74}$$

□

If $v \in V \cap L^\infty(\Omega)$ we know from Lemma 2.5 that there exists a sequence $(v_n)_n$, $v_n \in \mathscr{D}(\Omega)$, such that

$$\lim_{n\to\infty} v_n = v \text{ strongly in } V, \tag{2.75}$$

$$\lim_{n\to\infty} v_n = v \text{ in } L^\infty(\Omega) \text{ weak } * . \tag{2.76}$$

Since $v_n \in \mathscr{D}(\Omega)$ we have, from (2.74), **146**

$$a(u, v_n) + \int_\Omega \phi(u) v_n dx = L(v_n). \tag{2.77}$$

It follows from (2.77) that $\lim_{n\to\infty} a(u, v_n) = a(u, v)$, $\lim_{n\to\infty} L(v_n) = L(v)$ and, since $\phi(u) \in L^1(\Omega)$, (2.76) implies that

$$\lim_{n\to\infty} \int_\Omega \phi(u) v_n dx = \int_\Omega \phi(u) v dx.$$

Thus taking the limit in (2.77), we obtain

$$a(u, v) + \int_\Omega \phi(u) v dx = L(v) \ \forall v \in V \cap L^\infty(\Omega).$$

Therefore (P) implies (2.47) which implies in turn (π). This completes the proof of the Theorem.

Exercise 2.2. *Find in* $\mathbb{R}^2$*, a function* v *such that* $v \notin H^{-1}(\Omega) \cap L^1(\Omega)$, $v \notin L^p(\Omega) \, \forall_p > 1$, *where* Ω *is some bounded open set in* $\mathbb{R}^2$.

Exercise 2.3. *Prove that if* $u \geq 0$ *a.e. then* $\phi(u)v \in L^1(\Omega) \ \forall v \in V$, *where* u *is the solution of the problem* (P).

2.4 Some comments on the continuous problem

We have studied (P) and (π) with rather weak hypotheses, namely $\phi \in C^0(\mathbb{R})$ and nondecreasing, and $f \in V^*$. The proof we have given for the equivalence between (P) and (π) can be made shorter using more sophisticated tools of Convex Analysis and from the theory of Monotone Operators (see LIONS [1] and the bibliography therein). However our

proof is very elementary and some of the lemmas we have obtained will be useful in the numerical analysis of the problem (P). *Regularity results* for problems little more complicated that (P) and (π) are given in BREZIS-CRANDALL-PAZY [1] ; in particular for $f \in L^2(\Omega)$ and with convenient smoothness assumptions for A, the $H^2(\Omega)$-regularity of u is proved.

2.5 Finite element approximation of (π) and (P)

2.5.1 Definition of the approximate problem

147

Let Ω be a bounded *polygonal* domain of $\mathbb{R}^2$ and $\mathscr{C}_h$ be a triangulation of Ω satisfying (2.21)- (2.23) of Chap. 2. We approximate V by

$$V_h = \{v_h \in C^0(\overline{\Omega}) : v_h|_\Gamma = 0, v_h|_T \in P_1 \ \forall T \in \mathscr{C}_h\}.$$

Then it is natural to approximate (P) and (π) respectively by

$$\begin{cases} a(u_h, v_h) + \int_\Omega \phi(v_h) v_h dx = L(v_h) \ \forall v_h \in V_h, \\ \\ u_h \in V_h \end{cases} \tag{P_h^*}$$

and

$$\begin{cases} a(u_h, v_h - u_h) + j(v_h) - j(u_h) \geq L(v_h - u_h) \ \forall v_h \in V_h, \\ \\ u_h \in V_h \end{cases} \tag{π_h^*}$$

with

$$j(v_h) = \int_\Omega \phi(v_h) dx.$$

Obviously (P_h^*) and (π_h^*) are equivalent. From a computational point of view we cannot use in general (P_h^*) and (π_h^*) directly since they involve the computation of integrals which cannot be done exactly. For this reason we shall have to modify (π_h^*) and (P_h^*) using somewhere some *numerical integration procedures.* Actually we shall have to approximate $a(\cdot,\cdot)$, $L(\cdot)$ and $j(\cdot)$. Since the approximation of $a(\cdot,\cdot)$ and $L(\cdot)$ is

studied in CIARLET [1, Chap. 8] we shall assume that we still work with $a(\cdot,\cdot)$ and $L(\cdot)$ but we shall approximate $j(\cdot)$.

To approximate $j(\cdot)$ we shall use the *two dimensional trapezoidal method*. Hence using the notation of Figure 2.3 below we approximate $j(\cdot)$ by

$$j_h(v_h) = \sum_{T\in\mathscr{C}_h} \frac{meas\cdot(T)}{3} \sum_{i=1}^{3} \Phi(v_h(M_{iT}))\ \forall v_h \in V_h. \tag{2.78}$$

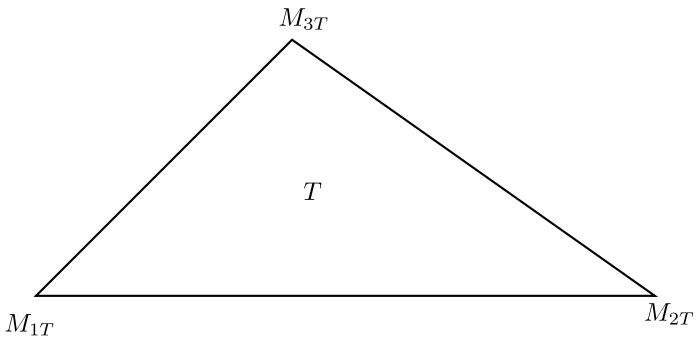

Figure 2.3:

Actually $j_h(v_h)$ may be viewed as the exact integral of some piece- **148**
wise constant functions.

Using the notation of Chap. 2, Sec. 2.5, assume that the set $\sum_h$ of the nodes of $\mathscr{C}_h$ has been ordered by $i = 1, 2, \ldots, N_h$ where $n_h = Card(\sum_h)$. Let $M_i \in \sum_h$. We define a domain Ω_i by joining, as in Figure 2.4, the centroids of the triangles, admitting M_i as a common vertex, to the midpoint of the edges admitting M_i as a common extremity (if M_i is a boundary point the modification of Figure 2.4 is trivial to do).

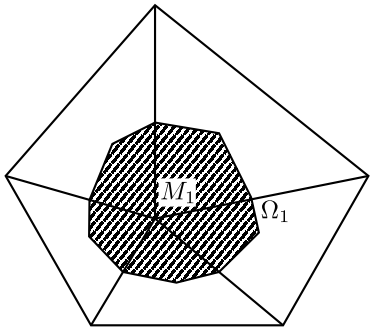

Figure 2.4:

Let us define the space of piecewise functions :

$$L_h = \{\mu_h : \mu_h = \sum^{N_h} i = 1\mu_i\chi_i, \mu_i \in \mathbb{R}, i = 1, 2, \ldots, N_h\}, \tag{2.79}$$

149 where χ_i is the characteristic function of Ω_i.

We then define $q_h : C^0(\overline{\Omega}) \cap H_0^1(\Omega) \to L_n$ by

$$q_h v = \sum_{i=1}^{N_h} v(M_i)\chi_i. \tag{2.80}$$

Then it follows from (2.79) and (2.80) that

$$j_h(v_h) = \int_\Omega \Phi(q_h v_h)dx \,\forall v_h \in V_h. \tag{2.81}$$

We also have

$$j_h(v_h) = j(q_h v_h) \,\forall v_h \in V_h. \tag{2.82}$$

Then we approximate (P) and (π) by

$$\begin{cases} a(u_h, v_h) + \int_\Omega \phi(q_h u_h) q_h v_h dx = L(v_h) \,\forall v_h \in V_h, \\ u_h \in V_h \end{cases} \tag{P_h}$$

and

$$\begin{cases} a(u_h, v_h - u_h) + j_h(v_h) - j_h(u_h) \geq L(v_h - u_h) \; \forall v_h \in V_h, \\ u_h \in V_h. \end{cases} \tag{π_h}$$

Then

THEOREM 2.3. *Problem* (P_h) *and* (π_h) *are equivalent and have a unique solution.*

Exercise 2.4. *Prove Theorem 2.3.*

2.5.2 Convergence of the approximate solutions

THEOREM 2.4. *If as* $h \to 0$ *the angles of* $\mathscr{C}_h$ *are uniformly bounded below by* $\theta_0 > 0$ *then*

$$\lim_{h \to 0} \|u_h - u\|_V = 0,$$

where u *and* u_h *are respectively the solutions of* (P) *and* (P_h).

Proof. Since $j(\cdot)$ is *not continuous* on V, the result of Chap. 1, Sec. 6 on the approximation of EVI of the second kind cannot be applied directly. However the proof of the convergence follows the same lines as in Theorem 6.2 of Chap. 1. □ **150**

(1) **A priori estimates for** u_h**.** Taking $v_h = 0$ in (π_h) we obtain

$$\|u_h\|_V \leq \frac{\|f\|}{\alpha}. \tag{2.83}$$

$$0 \leq \int_\Omega \Phi(q_h u_h) dx \leq \frac{\|f\|^2}{\alpha}. \tag{2.84}$$

(2) **Weak convergence of** u_h**.** It follows from (2.83) and from the compactness of the injection of V in $L^2(\Omega)$, that we can extract from $(u_h)_h$ a subsequence, still denoted by $(u_h)_h$, such that

$$u_h \to u^* \text{ weakly in } V, \tag{2.85}$$

$$u_h \to u^* \text{ strongly in } L^2(\Omega), \tag{2.86}$$

$$u_h \to u^* \text{ a.e. in } \Omega. \tag{2.87}$$

Admitting for the moment the following inequality

$$\|q_h v_h - v_h\|_{L^p(\Omega)} \leq \frac{2h}{3}\|\nabla v\|_{L^p(\Omega)\times L^p(\Omega)} \forall v_h \in V_h,\ \forall p \text{ with } 1 \leq p \leq \infty, \tag{2.88}$$

it follows from (2.83) and (2.86) that

$$q_h u_h \to u^* \text{ strongly in } L^2(\Omega). \tag{2.89}$$

Then, modulo another extraction of a subsequence, we have

$$q_h u_h \to u^* \text{ a.e. in } \Omega, \tag{2.90}$$

from which it follows that

$$\Phi(q_h u_h) \to \Phi(u^*) \text{ a.e. in } \Omega. \tag{2.91}$$

151 Then taking $v \in \mathscr{D}(\Omega)$, it follows from CIARLET [1], [2], STRANG-FIX [1] that under the assumptions on $\mathscr{C}_h$ of the statement of the Theorem we have

$$\|r_h v - v\|_{W^{1,\infty}(\Omega)} \leq Ch\|v\|_{W^{2,\infty}(\Omega)} \forall v \in \mathscr{D}(\Omega), \tag{2.92}$$

$$\|r_h v - v\|_{L^{\infty}(\Omega)} \leq Ch^2\|v\|_{W^{2,\infty}(\Omega)} \forall v \in \mathscr{D}(\Omega), \tag{2.93}$$

where C is a constant independent of v and h and where r_h is the usual linear interpolation operator over $\mathscr{C}_h$. Moreover (2.88) with $p = +\infty$, (2.92) and (2.93) imply that

$$\lim_{h\to 0} \|q_h r_h v - v\|_{L^{\infty}(\Omega)} = 0 \ \forall v \in \mathscr{D}(\Omega). \tag{2.94}$$

Taking $v_h = r_h v$ in (π_h) we obtain

$$\begin{aligned} a(u_h, u_h) + \int_\Omega \Phi(q_h u_h)dx &\leq a(u_h, r_h v) \\ &+ \int_\Omega \Phi(q_h r_h v)dx - L(r_h v - u_h)\ \forall v \in \mathscr{D}(\Omega). \end{aligned} \tag{2.95}$$

From (2.85), (2.89) and Lemma 2.1 we have

$$a(u^*, u^*) + \int_\Omega \Phi(u^*)dx \leq \liminf(a(u_h, u_h) + \int_\Omega \Phi(q_h u_h)dx).$$

Moreover

$$\lim_{h\to 0}\int_\Omega \Phi(q_h r_h v)dx = \int_\Omega \Phi(v)dx = j(v)\forall v \in \mathscr{D}(\Omega).$$

Then in the limit in (2.95) we obtain

$$a(u^*,u^*) + j(u^*) \le a(u^*,v) + j(v) - L(v-u^*)\,\forall v \in \mathscr{D}(\Omega). \tag{2.96}$$

From Fatou's lemma applied to (2.84) and (2.91) we obtain

$$\Phi(u^*) \in L^1(\Omega). \tag{2.97}$$

Then it follows from (2.96) and (2.97) that u^* satisfies

$$a(u^*,v-u^*) + j(v) - j(u^*) \le L(v-u^*)\,\forall v \in \mathscr{D}(\Omega), u^* \in V, \phi(u^*) \in L^1(\Omega). \tag{2.98}$$

We now take $v \in V \cap L^\infty(\Omega)$, it follows from Lemma 2.5 that there **152**
exists a sequence $(v_n)_n$, $v_n \in \mathscr{D}(\Omega)$ such that

$$\lim_{n\to\infty} v_n = v \text{ strongly in } V, \tag{2.99}$$

$$\lim_{n\to\infty} v_n = v \text{ in } L^\infty(\Omega) \text{ weak }^*. \tag{2.100}$$

We have from (2.98) that

$$a(u^*,v_n-u^*) + j(u_n) - j(u^*) \ge L(v_n-u^*)\,\forall n, u^* \in V, \Phi(u^*) \in L^1(\Omega). \tag{2.101}$$

We obviously have from (2.99)

$$\lim_{n\to\infty} a(u^*,v_n-u^*) = a(u^*,v-u^*),$$
$$\lim_{n\to\infty} L(v_n-u^*) = L(v-u).$$

Since $v_n \to v$ in the weak * topology of $L^\infty(\Omega)$ we have a constant C such that

$$\|v_n\|_{L^\infty(\Omega)} \le C\,\forall n. \tag{2.102}$$

Moreover, for some subsequence, (2.99) implies that

$$\lim_{n\to\infty} v_n = v \text{ a.e. in } \Omega. \tag{2.103}$$

From (2.103) we obtain that

$$\Phi(v_n) \to \Phi(v) \text{ a.e. in } \Omega.$$

From (2.102) and (2.103) one can easily see that the Lebesgue dominated convergence theorem can be applied to $(Phi(u_n))_n$. Hence we obtain

$$\lim_{n\to\infty} j(v_n) = \lim_{n\to\infty} \int_\Omega \Phi(v_n)dx = \int_\Omega \Phi(v)dx = j(v).$$

153 Therefore taking the limit in (2.101) we obtain

$$\begin{cases} a(u^*, v-u^*) + j(v) - j(u^*) \geq L(v-u)^* \; \forall v \in V \cap L^\infty(\Omega), \\ \\ u^* \in V, \Phi(u^*) \in L^1(\Omega). \end{cases} \tag{2.104}$$

Since from Lemma 2.2 we know that (2.104) is equivalent to (π) we have proved that $u^* = u$ where u is the solution of (π). From the uniqueness of the solution of (π) it follows that the whole sequence $(u_h)_h$ converges to u.

(3) **Strong convergence of** $(u_h)_h$: From the V-ellipticity of $a(\cdot,\cdot)$ and from the variational inequality satisfied by u_h we obtain

$$\begin{cases} \alpha\|u_h - u\|^2 + j_h(u_h) \leq a(u_h - u, u_h - u) + j_h(u_h) \\ \qquad\qquad\qquad\qquad \leq -a(u_h, u) + a(u, u) + a(u_h, u_h) \\ -a(u, u_h) + j_h(u_h) \leq -a(u_h, u) + a(u, u) + a(u_h, r_h v) \\ \qquad\qquad\qquad\qquad + j_h(r_h v) - L(r_h v - u_h) \\ -a(u, u_h) \quad \forall v \in \mathscr{D}(\Omega). \end{cases} \tag{2.105}$$

Using the various convergence results of Part (2) we obtain from (2.105) that

$$\begin{cases} j(u) \quad \leq \liminf j_h(u_h) \leq \liminf(\alpha\|u_h - u\|_V^2 + j_h(u_h)) \\ \qquad \leq \limsup(\alpha\|u_h - u\|^2 + j_h(u_h)) \\ \qquad\quad \leq a(u, v-u) + j(v) - L(v-u) \quad \forall v \in \mathscr{D}(\Omega). \end{cases} \tag{2.106}$$

Using as in Part (2) the density of $\mathscr{D}(\Omega)$ in $L^\infty(\Omega) \cap V$ (for the strong topology of V and the weak * topology of $L^\infty(\Omega)$), we obtain that (2.106) also holds for all $v \in V \cap L^\infty(\Omega)$. Then

$$\begin{cases} j(u) & \leq \liminf j_h(u_h) \leq \liminf(\alpha\|u_h - u\|_V^2 + j_h(u_h)) \\ & \leq \limsup(\alpha\|u_h - u\|_V^2 + j_h(u_h)) \\ & \leq a(u, \tau_n v - u) + j(\tau_n v) - L(\tau_n v - v) \ \ \forall v \in V. \end{cases} \tag{2.107}$$

Using the properties of $\tau_n v$, it follows then from (2.107), by taking the limit as $n \to \infty$, that (2.106) also holds for all $v \in V$. Hence by taking $v = u$ we obtain

$$\begin{cases} j(u) & \leq \liminf j_h(u_h) \leq \liminf(\alpha\|u_h - u\|^2 + j_h(u_h)) \\ & \leq \limsup(\alpha\|u_h - u\|^2 + j_h(u_h)) \leq j(u), \end{cases}$$

which implies **154**

$$\lim_{h\to 0} j_h(u_h) = j(u),$$
$$\lim_{h\to 0} \|u_h - u\|_V = 0.$$

This proves the Theorem modulo the proof of (2.88). We now prove (2.88).

Lemma 2.6. *We have*

$$\forall p, 1 \leq p \leq \infty, \|q_h v_h - v_h\|_{L^p(\Omega)} \leq \frac{2}{3} h \|\nabla v_h\|_{L^p(\Omega)\times L^p(\Omega)}$$

where q_h, v_h are as before.

Proof. We use the notation of Sec. 2.5.1 □

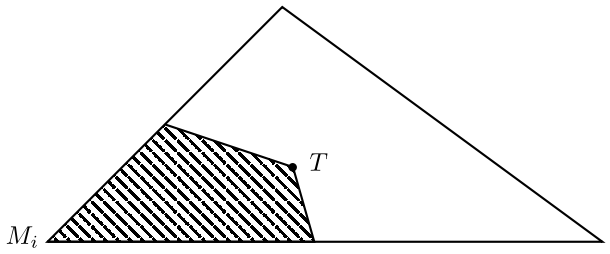

Figure 2.5:

We have (see Figure 2.5)

$$|v_h(M) - q_h v_h(M)| = |v_h(M_i) - v_h(M)| \; \forall M \in \Omega_i \cap T. \tag{2.108}$$

But since $v_h|_T \in P_1$ we have

$$v_h(M) = v_h(M_i) + \overrightarrow{M_i M} \cdot \nabla v_h \quad \forall M \in \Omega_i \cap T,$$

from which it follows, combined with (2.108), that

$$|q_h v_h(M) - v_h(M)| \leq |\overrightarrow{M_i M}| |\nabla v_h| \; \forall M \in \Omega_i \cap T.$$

155 But from the definition of h we have $|\overrightarrow{M_i M}| \leq \frac{2}{3}h \; \forall T$ so that we finally have $|q_h v_h(x) - v_h(x)| \leq \frac{2}{3}h|\nabla v_h(x)|$ a.e. in Ω, $\forall v_h \in V_h$. This implies

$$\|q_h v_h - v_h\|_{L^p(\Omega)} \leq \frac{2}{3}h\|\nabla v_h\|_{L^p(\Omega)\times L^p(\Omega)}.$$

This proves the lemma.

REMARK 2.4. *We have not considered the problem of error estimates. This problem will be discussed in GLOWINSKI [4].*

REMARK 2.5. *The numerical analysis of problem like* (P) *but with much stronger hypotheses on* $a(\cdot,\cdot)$, ϕ, f *is considered in CIARLET - SCHULTZ - VARGA [1].*

2.6 Iterative methods for solving the discrete problem

2.6.1 Introduction

In this section we briefly describe some *iterative methods* which may be useful for computing the solution of (P_h) (and (π_h)). Actually most of these methods can be extended to other non-linear problems. Many of the methods to be described here can be found in ORTEGA-RHEINBOLDT [1]. A method based on *duality* techniques will be described in Chap. 5.

2.6.2 Formulation of the discrete problem

Here we are using the notation of the continuous problem. Taking as unknowns the values of u_h at the interior nodes of $\mathscr{C}_h$, the problem (P_h) reduces to the finite dimensional non-linear problem

$$\underset{\sim}{A}\,\underset{\sim}{u} + \underset{\sim}{D}\,\phi(\underset{\sim}{u}) = \underset{\sim}{f}, \tag{2.109}$$

where $\underset{\sim}{A}$ is a $N \times N$ positive definite matrix, $\underset{\sim}{D}$ is a diagonal matrix with positive diagonal elements d_i's and where $\underset{\sim}{u} = \{u_1, \ldots, u_N\} \in \mathbb{R}^N$, $\underset{\sim}{f} \in \mathbb{R}^N$, $\phi(\underset{\sim}{u}) \in \mathbb{R}^N$ with $(\phi(\underset{\sim}{u}))_i = \phi(u_i)$. Clearly from the properties of $\underset{\sim}{A}$, $\underset{\sim}{D}$, ϕ, $\underset{\sim}{f}$ we can see that the problem (2.109) has a unique solution.

2.6.3 Gradient Methods

The basic algorithm with constant step (see CEA [1]) is given by **156**

$$\overset{0}{\underset{\sim}{u}} \in \mathbb{R}^N \text{ given }, \tag{2.110}$$

$$\overset{n+1}{\underset{\sim}{u}} = \overset{n}{\underset{\sim}{u}} - \rho \underset{\sim}{S}^{-1}(\underset{\sim}{A}\,\underset{\sim}{u}^n + \underset{\sim}{D}\,\phi(\underset{\sim}{u}^n) - \underset{\sim}{f}), \rho > 0. \tag{2.111}$$

In (2.111), $\underset{\sim}{S}$ is a *symmetric, positive definite matrix*: a canonical choice is $\underset{\sim}{S} = \mathit{Identity}$. But in most problems it will give a *slow speed of convergence*. If A is symmetric, the natural choice is $\underset{\sim}{S} = \underset{\sim}{A}$ and, if $\underset{\sim}{A} \neq \overset{*}{\underset{\sim}{A}}$, we can take $\underset{\sim}{S} = \dfrac{\underset{\sim}{A} + \overset{*}{\underset{\sim}{A}}}{2}$.

For the convergence of $\underset{\sim}{u}^n$ to $\underset{\sim}{u}$ (where $\underset{\sim}{u}$ is the solution of (2.109)) it is sufficient to have ϕ smooth enough (for example, ϕ locally Lipschitz continuous). Then $\lim_{n\to\infty} \underset{\sim}{u}^n = \underset{\sim}{u}$ if ρ is sufficiently small. Obviously the closer $\underset{\sim}{u}^0$ is to $\underset{\sim}{u}$, the faster is the convergence.

REMARK 2.6. *If* $\underset{\sim}{A} = \overset{*}{\underset{\sim}{A}}$, *then* $\underset{\sim}{A}\,\underset{\sim}{v} + \underset{\sim}{D}\,\phi(\underset{\sim}{v}) - \underset{\sim}{f}$ *is the gradient at* $\underset{\sim}{v}$ *of the functional* $j(\underset{\sim}{v}) = \dfrac{1}{2}(A\,\underset{\sim}{v}, \underset{\sim}{v}) + \sum_{i=1}^{N} d_i \Phi(v_i) - (\underset{\sim}{f}, \underset{\sim}{v})$, *where* $(\cdot,\cdot)$ *denotes the usual inner product of* $\mathbb{R}^N$ *and* $\Phi(t) = \int_0^t \phi(\tau)d\tau$.

REMARK 2.7. *In each specific case ρ has to be determined ; this can be done theoretically, experimentally or by using an automatic adjustment procedure which will not be described here.*

REMARK 2.8. *Let us define* $\underset{\sim}{g}^n$ *by*

$$\underset{\sim}{g}^n = \underset{\sim}{A}\,\underset{\sim}{u}^n + \underset{\sim}{D}\,\phi(\underset{\sim}{u}^n) - \underset{\sim}{f}\,.$$

Instead of using a constant parameter ρ we can use a family $(\rho_n)_n$ of positive parameters in (2.111). *Therefore* (2.111) *can be written as*

$$\underset{\sim}{u}^{n+1} = \underset{\sim}{u}^n - \rho_n \overset{-1}{\underset{\sim}{S}}\,\underset{\sim}{g}^n. \tag{2.112}$$

Suppose $\underset{\sim}{A} = \overset{*}{\underset{\sim}{A}}$, *then if we use* (2.110), (2.112) *with ρ_n defined by*

$$\begin{cases} J(\underset{\sim}{u}^n - \rho_n \overset{-1}{\underset{\sim}{S}}\,\underset{\sim}{g}^n) \le J(\underset{\sim}{u}^n - \rho \overset{-1}{\underset{\sim}{S}}\,\underset{\sim}{g}^n) \quad \forall \rho \in \mathbb{R}, \\ \\ \rho_n \in \mathbb{R}, \end{cases} \tag{2.113}$$

157 *the resulting algorithm is, for obvious reasons, called steepest descent method. This algorithm is convergent for $\phi \in C^0(\mathbb{R})$. We observe that at each iteration the determination of ρ_n requires the solution of a one-dimensional problem; for the solution of such one dimensional problems see HOUSEHOLDER [1], POLAK [1], BRENT [1].*

REMARK 2.9. *At each iteration of* (2.110), (2.111) *or* (2.110), (2.112) *we have to solve a linear system related to* $\underset{\sim}{S}$. *Since* $\underset{\sim}{S}$ *is symmetric and positive definite this system can be solved using Cholesky method, provided the* $\underset{\sim}{S} = \underset{\sim}{L}\,\overset{*}{\underset{\sim}{L}}$ *factorization has been done. From a practical point of view it is obvious that the factorization of* $\underset{\sim}{S}$ *will be made in the beginning once for all. Then at each iteration we just have to solve two triangular systems which is a trivial operation.*

2.6.4 Newton's method

The Newton's algorithm is given by (for sufficient conditions of convergence, see ORTEGA-RHEINBOLDT [1]) :

$$\underset{\sim}{u}^0 \in \mathbb{R}^N \text{ given }, \tag{2.114}$$

$$\underset{\sim}{u}^{n+1} = (\underset{\sim}{A} + \underset{\sim}{D}\,\phi'(\underset{\sim}{u}^n))^{-1}(\underset{\sim}{D}\,\phi'(\underset{\sim}{u}^n)\underset{\sim}{u}^n - \underset{\sim}{D}\,\phi(\underset{\sim}{u}^n) + \underset{\sim}{f}) \tag{2.115}$$

where $\phi'(\underset{\sim}{v})$ denotes the *diagonal matrix*

$$\phi'(\underset{\sim}{v}) = \begin{pmatrix} \phi'(v_1) & \ddots & 0 \\ & & \ddots \\ 0 & & \phi'(v_n) \end{pmatrix}$$

Since ϕ is nondecreasing, $\phi' \geq 0$. This implies that $\underset{\sim}{A} + \underset{\sim}{D}\,\phi'(\underset{\sim}{v})$ is positive definite $\forall\, \underset{\sim}{v} \in \mathbb{R}^N$.

REMARK 2.10. *At each iteration we have to solve a linear system. Since the matrix* $\underset{\sim}{A} + \underset{\sim}{D}\,\phi'(\underset{\sim}{u}^n)$ *depends on n, this method may not be convenient for large N.*

REMARK 2.11. *The choice of* $\underset{\sim}{u}^0$ *is very important when using Newton's method.*

2.6.5 Relaxation and over-relaxation methods

We use the following notation : **158**

$$\underset{\sim}{A} = (a_{ij})_{1\leq i,j\leq N},$$
$$\underset{\sim}{f} = \{f_1, f_2, \ldots, f_N\}.$$

Since $\underset{\sim}{A}$ is *positive definite* we have $a_{ii} > 0\, \forall i = 1, 2, \ldots, N$. Here we will describe three algorithms:

Algorithm 1.

$$\underset{\sim}{u}^0 \in \mathbb{R}^N \text{ given }, \tag{2.116}$$

then given $\underset{\sim}{u}^n$*, we compute* $\underset{\sim}{u}^{n+1}$*, component by component, using*

$$a_{ii}u_i^{-n+1} + d_i\phi(u_i^{-n+1}) = f_i - \sum_{j<i} a_{ij}u_j^{n+1} - \sum_{j>i} a_{ij}u_j^n, \tag{2.117}$$

$$u_i^{n+1} = u_i^n + \omega(u_i^{-n+1} - u_i^n), i = 1, 2, \ldots, N. \tag{2.118}$$

Since $a_{ii} > 0$, $\phi \in C^0(\mathbb{R})$ and ϕ is a nondecreasing function we can always solve (2.117) and the solution is unique.

If $\omega = 1$, we recover an *ordinary relaxation method* (see CEA [1]). If $\underset{\sim}{A} = \overset{*}{\underset{\sim}{A}}$ and since ϕ is C^0 and nondecreasing the solution $\underset{\sim}{u}^n$ of (2.116)-(2.118) converges to the solution $\underset{\sim}{u}$ of (2.109).

If, in (2.109), $\underset{\sim}{A}$ is not symmetric and $\omega \neq 1$, there are certain sufficient conditions for the convergence of $\underset{\sim}{u}^n$ to $\underset{\sim}{u}$, where $\underset{\sim}{u}^n$ is given by (2.116)-(2.118) and where $\underset{\sim}{u}$ is the solution of (2.109) (see ORTEGA-RHEINBOLDT [1], S. SCHECHTER [1], [240], [3]).

Algorithm 2. *This algorithm is the variant of* (2.116) - (2.118) *obtained by replacing* (2.117) *and* (2.118) *by*

$$\begin{cases} a_{ii}u_i^{n+1} + d_i\phi(u_i^{n+1} = (1-\omega)(a_{ii}u_i^n + d_i\phi(u_i^n)) \\ \qquad\qquad +\omega(f_i - \sum_{j<i} a_{ij}u_j^{n+1} - \sum_{j>i} a_{ij}u_j^n) \\ \text{for } i = 1, 2, \ldots, N. \end{cases} \tag{2.119}$$

159 **REMARK 2.12.** *If* $\omega = 1$ *and/or* ϕ *is linear the two algorithms coincide. In the general case the convergence of* (2.116), (2.119) *seems to be an* open question. *However from numerical experimentation's it seems that the second algorithm is more* "robust" *than the first, maybe because it is more* implicit. *Furthermore it can be used even if* ϕ *is defined only on a bounded or semi bounded interval* $]\alpha, \beta[$ *of* $\mathbb{R}$ *such that* $\phi(\alpha) = -\infty$, $\phi(\beta) = +\infty$ *; in such a case if* $\phi \in C^0(]\alpha, \beta[)$ *and* ϕ *is increasing,* (2.109) *has still a unique solution but the use of* (2.116)-(2.118) *with* $\omega > 1$ *may be dangerous.*

REMARK 2.13. *If $\phi \in C^1(\mathbb{R})$, an efficient method to compute u_i^{-n+1} in* (2.117) *and u_i^{n+1} in* (2.119) *is the* one dimensional Newton's method.

Let $g \in C^1(\mathbb{R})$. In this case Newton's algorithm to solve the equation $g(x) = 0$ is

$$x^0 \in \mathbb{R} \text{ given }, \tag{2.120}$$

$$x^{n+1} = x^n - \frac{g(x^n)}{g'(x^n)}. \tag{2.121}$$

If in the computation of u_i^{-n+1} and u_i^{n+1} we use only one iteration *of Newton's method. starting from u_i^n, then the resulting algorithms are identical and we obtain*

Algorithm 3.

$$\underset{\sim}{u}^0 \in \mathbb{R}^N \text{ given }, \tag{2.122}$$

$$u_i^{n+1} = u_i^n - \frac{\sum_{j<i} a_{ij} u_j^{n+1} + \sum_{j>i} a_{ij} u_j^n + d_i \phi(u_i^n) - f_i}{a_{ii} + d_i \phi'(u_i^n)}, i = 1, 2, \ldots N. \tag{2.123}$$

In S. SCHECHTER [1], [240], [3] sufficient conditions for the convergence of (2.122), (2.123) *are given.*

REMARK 2.14. *If $\omega > 1$ (resp. . $\omega = 1$, $\omega < 1$) the previous algorithms are* over-relaxation (resp. . relaxation, under relaxation) *algorithms.*

REMARK 2.15. *We can find in GLOWINSKI-MARROCCO [1], [2] applications of relaxation methods for solving the* nonlinear elliptic equations *modelling the* magnetic state of electrical machines. **160**

2.6.6 Alternating Direction Methods

In this section we take $\rho > 0$. Here we will give two numerical methods for solving (2.109).
First method.

$$\underset{\sim}{u}^0 \in \mathbb{R}^N \text{ given }, \tag{2.124}$$

knowing $\underset{\sim}{u}^n$ we compute $\underset{\sim}{u}^{n+\frac{1}{2}}$ by

$$\rho\underset{\sim}{u}^{n+\frac{1}{2}} + \underset{\sim}{A}\,\underset{\sim}{u}^{n+\frac{1}{2}} = \rho\underset{\sim}{u}^n - \underset{\sim}{D}\,\phi(\underset{\sim}{u}^n) + \underset{\sim}{f}, \tag{2.125}$$

then we calculate $\underset{\sim}{u}^{n+1}$ by

$$\rho\underset{\sim}{u}^{n+1} + \underset{\sim}{D}\,\phi(\underset{\sim}{u}^{n+1}) = \rho\underset{\sim}{u}^{n+\frac{1}{2}} - \underset{\sim}{A}\,\underset{\sim}{u}^{n+\frac{1}{2}}) + \underset{\sim}{f}\,. \tag{2.126}$$

For the convergence of (2.124)-(2.126) see R. B. KELLOG [1].
Second method.

$$\underset{\sim}{u}^0 \epsilon \mathbb{R}^N \text{ given}, \tag{2.127}$$

knowing $\underset{\sim}{u}^n$ we compute $\underset{\sim}{u}^{n+\frac{1}{2}}$ by

$$\rho\underset{\sim}{u}^{n+\frac{1}{2}} + \underset{\sim}{A}\,\underset{\sim}{u}^{n+\frac{1}{2}} = \rho\underset{\sim}{u}^n - \underset{\sim}{D}\,\phi(\underset{\sim}{u}^n) + \underset{\sim}{f}, \tag{2.128}$$

then we calculate $\underset{\sim}{u}^{n+1}$ by

$$\rho\underset{\sim}{u}^{n+1} + \underset{\sim}{D}\,\phi(\underset{\sim}{u}^{n+1}) = \rho\underset{\sim}{u}^n - \underset{\sim}{A}\,\underset{\sim}{u}^{n+\frac{1}{2}} + \underset{\sim}{f}\,. \tag{2.129}$$

Using the results of LIEUTAUD [1], it can be proved that, for all $\rho > 0$, $\underset{\sim}{u}^{n+\frac{1}{2}}$ and $\underset{\sim}{u}^n$ if we suppose that $\underset{\sim}{A}$ and ϕ satisfy the hypotheses given in Sec. 2.6.2.

161 **REMARK 2.16.** *At each iteration we have to solve a linear system, the matrix of which is* constant, *since we use a* constant step ρ. *This is an advantage from the computational point of view (cf. Remark 2.9).*

We also have to solve a nonlinear system of N equations, but in fact these equations are independent from each other and reduce to N nonlinear equations in one variable *which can be solved easily.*

2.6.7 Conjugate gradient methods

In this section we assume $\underset{\sim}{A} = \overset{*}{\underset{\sim}{A}}$ (if $\underset{\sim}{A} \neq \overset{*}{\underset{\sim}{A}}$ we can also use methods of conjugate gradient type). For a detailed study of these methods we refer

to POLAK [1], DANIEL [1], CONCUS-GOLUB [1]. If the functional J defined in Remark 2.6 is not quadratic (i.e. ϕ is nonlinear), several conjugate gradient methods can be used. Let us describe two of them, the convergence of which is studied in POLAK [1]. Let J be given by

$$J(\underset{\sim}{v}) = \frac{1}{2}(A\underset{\sim}{v}, \underset{\sim}{v}) + \sum_{i=1}^{N} d_i \Phi(v_i) - (\underset{\sim}{f}, \underset{\sim}{v}),$$

where $\Phi(t) = \int_0^t \phi(\tau)d\tau$, Φ being a nondecreasing continuous function on $\mathbb{R}$ with $\Phi(0) = 0$.

Let S be a $N \times N$ *symmetric, positive definite matrix. First method. (Fletcher-Reeves)*

$$\underset{\sim}{u}^0 \in \mathbb{R}^N \text{ given }, \tag{2.130}$$

$$\underset{\sim}{g}^0 = \overset{-1}{\underset{\sim}{S}}(\underset{\sim}{A}\,\underset{\sim}{u}^0 + \underset{\sim}{D}\,\phi(\underset{\sim}{u}^0) - \underset{\sim}{f}), \tag{2.131}$$

$$\underset{\sim}{w}^0 = \underset{\sim}{g}^0. \tag{2.132}$$

Then assuming that $\underset{\sim}{u}^n$ and $\underset{\sim}{w}^n$ are know we compute $\underset{\sim}{u}^{n+1}$ by

$$\underset{\sim}{u}^{n+1} = \underset{\sim}{u}^n - \rho_n \underset{\sim}{w}^n, \tag{2.133}$$

where ρ_n is the solution of the one-dimensional minimization problem

$$\begin{cases} J(\underset{\sim}{u}^n - \rho_n \underset{\sim}{w}^n) \leq J(\underset{\sim}{u}^n - \rho \underset{\sim}{w}^n)\ \forall \rho \in \mathbb{R}, \\ \\ \rho_n \in \mathbb{R}. \end{cases} \tag{2.134}$$

Then **162**

$$\underset{\sim}{g}^{n+1} = \overset{-1}{\underset{\sim}{S}}(\underset{\sim}{A}\,\underset{\sim}{u}^{n+1}) + \underset{\sim}{D}\,\phi(\underset{\sim}{u}^{n+1}) - \underset{\sim}{f}), \tag{2.135}$$

and compute $\underset{\sim}{w}^{n+1}$ by

$$\underset{\sim}{w}^{n+1} = \underset{\sim}{g}^{n+1} + \lambda_n \underset{\sim}{w}^n, \tag{2.136}$$

where

$$\lambda_n = \frac{(\underset{\sim}{S}\,\underset{\sim}{g}^{n+1}, \underset{\sim}{g}^{n+1})}{(\underset{\sim}{S}\,\underset{\sim}{g}^{n}, \underset{\sim}{g}^{n})}. \tag{2.137}$$

Second method. (Polak-Ribiere). This method is like the previous method except that (2.137) is replaced by

$$\lambda_n = \frac{(\underset{\sim}{S}\,\underset{\sim}{g}^{n+1}, \underset{\sim}{g}^{n+1} - \underset{\sim}{g}^{n})}{(\underset{\sim}{S}\,\underset{\sim}{g}^{n}, \underset{\sim}{g}^{n})}. \tag{2.138}$$

REMARK 2.17. *For the computation of ρ_n in* (2.134)*, see Remark 2.8.*

REMARK 2.18. *It follows from POLAK [1], that if ϕ is sufficiently smooth then the convergence of the above algorithms is super linear i.e. faster that the convergence of any geometric sequence.*

REMARK 2.19. *The above algorithms are very sensitive to round off errors; hence double precision may be required for some problems. Moreover it may be convenient to take periodically* $\underset{\sim}{w}^n = \underset{\sim}{g}^n$.

REMARK 2.20. *We have to solve at each iteration a linear system related to* $\underset{\sim}{S}$*; Remark 2.9 still applies to this problem.*

2.6.8 Comments

The methods of this Sec. 2.6 may be applied to more general nonlinear systems than (2.109). They can be applied of course to finite dimensional systems obtained by discretization of elliptic problems like

$$\begin{cases} -\nabla \cdot (a_0(x)\nabla u) + \beta \cdot \nabla u + \phi(x,u) = f \text{ in } \Omega, \\ \\ + \text{ suitable boundary conditions} \end{cases}$$

where, for fixed x, the function $t \to \phi(x,t)$ is continuous and nondecreasing on $\mathbb{R}$.

3 A Subsonic Flow Problem

3.1 Formulation of the continuous problem

163

Let Ω be a domain of $\mathbb{R}^N$ (in applications we have $N = 1, 2, 3$) with a sufficiently smooth boundary Γ. Then the flow of a perfect, compressible, irrotational fluid (i.e. $\nabla \times \underset{\sim}{v} = \underset{\sim}{0}$ where $\underset{\sim}{v}$ is the velocity vector of the flow) is described by

$$-\nabla \cdot (\rho(\phi)\nabla\phi) = 0 \text{ in } \Omega, \tag{3.1}$$

$$\rho(\phi) = \rho_0(1 - \frac{|\nabla\phi|^2}{\frac{\gamma+1}{\gamma-1}C_*^2})^{1/(\gamma-1)}, \tag{3.2}$$

with suitable boundary conditions. Here

- ϕ is a potential and $\nabla\phi$ is the velocity of the flow,
- $\rho(\phi)$ is the density of the flow,
- ρ_0 is the density at $\nabla\phi = 0$; in the sequel we take $\rho_0 = 1$,
- γ is the ratio of specific heats,
- C_* is the critical velocity.

The flow under consideration is subsonic if

$$|\nabla\phi| < C_* \text{ everywhere in } \Omega. \tag{3.3}$$

If $|\nabla\phi| \geq C_*$ in some part of Ω, then the flow is *transonic or supersonic* and this leads to much more complicated problems (see Chap. 6 for an introduction to the study of such flows).

REMARK 3.1. *In the case of a sunsonic flow past a convex, symmetric airfoil and assuming (see Figure 3.1) that $\vec{v}_\infty$ is parallel to the x -axis (Ω is the complement of the profile in $\mathbb{R}^2$ and $\frac{\partial\phi}{\partial n}|_\Gamma = 0$), H. BREZIS-STAMPACCHIA [1] have proved that the subsonic problem is equivalent to an EVI of the first kind in the hodograph plane (see BERS [1],*

LANDAU-LIFCHITZ [1] for the hodograph transform). This EVI is related to a linear operator and the corresponding convex set is the cone of non-negative functions.

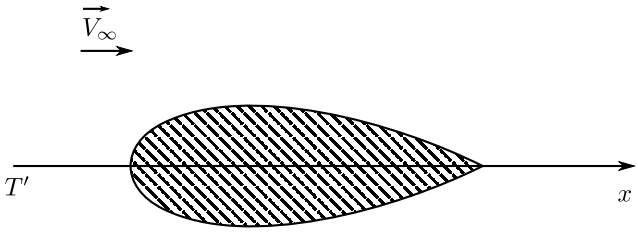

Figure 3.1:

164 *In the remainder of Sec. 3 (and also in Chap. 6) we shall only work in the physical plane since it seems more convenient for the computation of non-symmetric and/or transonic flows.*

For the reader who is interested by the mathematical aspects of the flow mentioned above, see BERS [1], BREZIS-STAMPACCHIA [1]. For the Physical and Mechanical aspects see LANDAU-LIPSCHITZ [1]. Additional references are given in Chap. 6.

3.2 Variational formulation of subsonic problems

Preliminary Remark: In the case of a non symmetric flow past an airfoil (see Figure 3.2)

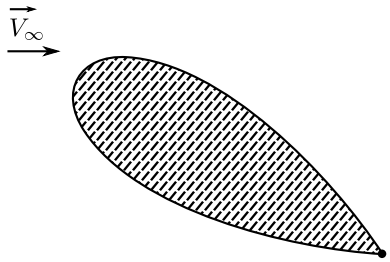

Figure 3.2:

the velocity potential has to be *discontinuous and a circulation condition* **165**
is required to ensure the uniqueness (modulo a constant) of the solution of (3.1). If the airfoil has corners (like in Figure 3.1) then the circulation condition is related to the so called *Kutta-Joukowsky condition* from which it follows that for a physical flow, the velocity field is continuous at the (like 0 in Figure 3.2). For more information about the Kutta-Joukowsky condition, see LANDAU-LIPSCHITZ [1] (see also Chap. 6).

For the sake of simplicity, we shall assume in the sequel that either Ω is simply connected, as it is the case for the nozzle of Figure 3.3, or, if Ω is multiply connected, we shall assume (like in Fig. 3.1) that the flow is physically and geometrically symmetric, since in this case the Kutta-Joukowsky condition is automatically satisfied.

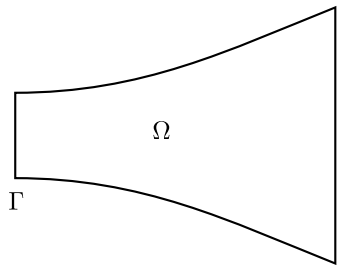

Figure 3.3:

We assume in the sequel that the boundary condition associated with (3.1), (3.2) are the following:

$$\phi = g_0 \text{ over } \Gamma_0,\ \rho\frac{\partial\phi}{\partial n}|_{\Gamma_1} = g_1 \tag{3.4}$$

where $\Gamma_0, \Gamma_1 \subset \Gamma$ with $\Gamma_0 \cap \Gamma_1 = \phi$, $\Gamma_0 \cup \Gamma_1 = \Gamma$. Then the variational formulation for the flow problem (3.1), (3.2), (3.3), (3.4) is

$$\begin{cases} \int_\Omega \rho(\phi)\nabla\phi \cdot \nabla v dx = \int_{\Gamma_1} g_1 v d\Gamma \ \forall v \in V_0, \\ \\ \phi \in V_{g_0}, \end{cases} \tag{3.5}$$

166 where

$$V_0 = \{v \in H^1(\Omega) : v|_{\Gamma_0} = 0\}, \tag{3.6}$$

$$V_{g_0} = \{v \in H^1(\Omega) : v|_{\Gamma_0} = g_0\}. \tag{3.7}$$

If g_0, g_1 are small enough, it can be proved that (3.5) has a solution such that

$$|\nabla\phi| \leq M < C_* \text{ a.e. .}$$

When solving a practical flow problem we may not know a priori, whether the flow will be purely subsonic or not. Therefore instead of solving (3.5) it may be convenient to consider (and solve) the following problem:

$$\begin{cases} \int_\Omega \rho(\phi)\nabla\phi \cdot \nabla(v-\phi)dx \geq \int_{\Gamma_1} g_1(v-\phi)d\Gamma \ \forall v \in K_\delta, \\ \\ \phi \in K_\delta, \end{cases} \tag{3.8}$$

where

$$K_\delta = \{v \in V_{g_0}, |\nabla v| \leq \delta < C_* a.e.\}. \tag{3.9}$$

The variational problem (3.8), (3.9) is an EVI of the first kind, but we have to observe that unlike the EVIs of Chap. 1 and 2, it involves a non-linear partial differential operator, namely A defined by

$$A(\phi) = -\nabla \cdot (\rho(\phi)\nabla\phi).$$

REMARK 3.2. *In practical applications we shall take δ as close as possible to C_*.*

REMARK 3.3. *Problem* (3.8), (3.9) *appears as a variant of the elasto-plastic torsion problem of Chap. 2, Sec. 3.*

3.3 Existence and uniqueness properties for the problem (3.8).

167 In this section we shall assume that Measure $(\Gamma_0) > 0$. To prove that (3.8) is well posed we will use the following

Lemma 3.1. *The function* $\xi \to -\left(1 - \frac{\xi^2}{\frac{\gamma+1}{\gamma-1}C_*^2}\right)^{\gamma/\gamma-1}$ *is convex if* $\xi \in [0, C_*]$, *concave if* $\xi \in [C_*, \sqrt{\frac{\gamma+1}{\gamma-1}}C_*]$ *and strictly convex if* $\xi \in [0, C_*[$.

Exercise 3.1. *Prove Lemma 3.1.*

We can now prove

THEOREM 3.1. *Assume that* Ω *is bounded and that* g_0, g_1 *are sufficiently smooth and small. Then* (3.8) *has a unique solution.*

Proof. Since Ω is bounded and if g_0 is sufficiently smooth and small, we observe that K_δ is a closed, convex, and nonempty bounded subset of $H^1(\Omega)$ (consisting of uniformly Lipschitz continuous functions).

Define $J(\cdot)$ by

$$J(v) = -\frac{\gamma+1}{2\gamma}C_*^2 \int_\Omega \left(1 - \frac{|\nabla v|^2}{\frac{\gamma+1}{\gamma-1}C_*^2}\right)^{\gamma/\gamma-1} dx - \int_{\Gamma_1} g_1 v d\Gamma. \tag{3.10}$$

□

It follows from Lemma 3.1 that $J(\cdot)$ is strictly convex over K_δ. It is easy to check that $J(\cdot)$ is continuous and Gateau-differentiable over K_δ with

$$(J'(v), w) = \int_\Omega \rho(v)\nabla v \cdot \nabla w dx - \int_{\Gamma_1} g_1 w d\Gamma. \tag{3.11}$$

Since K_δ is a closed, convex, nonempty subset of $H^1(\Omega)$ and that $J(\cdot)$ is continuous and strictly convex over K_δ, it follows from standard optimization theory in Hilbert space (see CEA [1], [2]) that the minimization problem

$$\begin{cases} J(u) \le J(v) \; \forall v \in K_\delta, \\ \\ u \in K_\delta, \end{cases} \tag{3.12}$$

has a unique solution.

Moreover since $J(\cdot)$ is differentiable the unique solution of (3.12) is **168**

characterised (see CEA [1], [2]) by

$$\begin{cases} (J'(u), v-u) \geq 0 \, \forall v \in K_\delta, \\ \\ u \in K_\delta; \end{cases}$$

from (3.11), this completes the proof of the Theorem.

REMARK 3.4. *Let us assume that $\Gamma_0 = \phi$. Then defining K_δ by*

$$K_\delta = \{v \in H^1(\Omega) : |\nabla v| \leq \delta < C_* a.e., v(x_0) = v_0\}$$

with $x_0 \in \overline{\Omega}$ and v_0 arbitrarily given, we can prove that if Ω is bounded and g_1 is sufficiently smooth then

$$\begin{cases} \int_\Omega \rho(\phi) \nabla\phi \cdot \nabla(v-\phi) dx \geq \int_\Gamma g_1 (v-\phi) d\Gamma \, \forall v \in K_\delta \\ \\ \phi \in K_\delta, \end{cases} \tag{3.13}$$

has a unique solution (if ϕ is a solution of (3.13) *then $\phi + C$ is the unique solution of the similar problem obtained by replacing v_0 by $v_0 + C$).*

Exercise 3.2. *Prove the statement of Remark 3.4.*

REMARK 3.5. *In all the above arguments we assumed that Ω is bounded. We refer to CIAVALDINI-POGU-TOURNMINE [1] in which one carefully studies the approximation of subsonic flow problems on an unbounded domain Ω_∞ by problems on a family $(\Omega_n)_n$ of bounded domains converging to Ω_∞ (actually they have obtained estimates for $\phi_\infty - \phi_n$).*

The above EVI's will have a practical interest if we can prove that in the cases where a purely subsonic solution exists, then for δ large enough it is the solution of (3.8); actually this property is true and follows from

169 **THEOREM 3.2.** *Assuming the same hypothesis on Ω, g_0, g_1 as in Theorem 3.1, and that* (3.4) *has a unique solution in $H^1(\Omega)$ with*

$$|\nabla\phi| \leq \delta_0 < C_* \ a.e. \tag{3.14}$$

then ϕ is a solution of (3.8), (3.9) $\forall \delta \in [\delta_0, C_*[$. *Conversely if the solution of* (3.8), (3.9) *is such that $|\nabla\phi| \leq \delta_0 < \delta$ a.e., then ϕ is a solution of* (3.1), (3.2), (3.4).

Proof. (1) Let $\phi \in H^1(\Omega)$ satisfying (3.1), (3.2), (3.4) and (3.14). If $v \in V_0$ then using Green's formula it follows from (3.1), (3.2), (3.4) that

$$\int_\Omega \rho(\phi)\nabla\phi \cdot \nabla v dx = \int_{\Gamma_1} g_1 v d\Gamma \; \forall v \in V_0. \tag{3.15}$$

It follows from (3.4), (3.15) and from the definition of V_{g_0} that

$$\int_\Omega \rho(\phi)\nabla\phi \cdot \nabla(v-\phi) dx = \int_{\Gamma_1} g_1(v-\phi) d\Gamma \; \forall v \in V_{g_0}. \tag{3.16}$$

Since $\phi \in K_\delta \subset V_{g_0} \; \forall \delta \in [\delta_0, C_*[$, it follows from (3.16) that

$$\begin{cases} \int_\Omega \rho(\phi)\nabla\phi \cdot \nabla(v-\phi) dx \geq \int_\Gamma g_1(v-\phi) d\Gamma \; \forall v \in K_\delta, \\ \\ \phi \in K_\delta, \end{cases}$$

if $\delta \in [\delta_0, C_*[$; therefore ϕ is the solution of the EVI (3.8), (3.9) $\forall \delta \in [\delta_0, C_*[$.

(2) Define $U \subset V_0$ by

$$U = \{v \in C^\infty(\overline{\Omega}) : v = 0 \text{ in a neighbourhood of } \Gamma_0\}.$$

□

Then, if we suppose that Γ is sufficiently smooth, we have

$$\overline{U}^{H^1(\Omega)} = V_0. \tag{3.17}$$

We assume that for $\delta < C_*$, (3.8) has a solution such that

$$|\nabla\phi| \leq \delta_0 < \delta \text{ a.e. }. \tag{3.18}$$

170

Then $\forall v \in U$ and for $t > 0$ sufficiently small $\phi + tv \in K_\delta$. Then replacing v by $\phi + tv$ in (3.8) and dividing by t obtain

$$\int_\Omega \rho(\phi)\nabla\phi \cdot \nabla v dx \geq \int_{\Gamma_1} g_1 v d\Gamma \; \forall v \in U,$$

which implies

$$\int_\Omega \rho(\phi)\nabla\phi \cdot \nabla v dx = \int_{\Gamma_1} g_1 v d\Gamma \; \forall v \in U. \tag{3.19}$$

Since $\mathscr{D}(\Omega) \subset U$ it follows from (3.19) that

$$\int_\Omega \rho(\phi)\nabla\phi \cdot \nabla v dx = 0 \; \forall v \in \mathscr{D}(\Omega), \tag{3.20}$$

i.e.

$$-\nabla \cdot (\rho(\phi)\nabla\phi) = 0$$

which proves (3.1).

Assuming (3.1) and Green's formula we obtain

$$\int_\Omega \rho(\phi)\nabla\phi \cdot \nabla v dx = \int_{\Gamma_1} \rho \frac{\partial \phi}{\partial n} v d\Gamma \; \forall v \in U. \tag{3.21}$$

Using (3.17) and comparing with (3.19) we obtain

$$\rho \frac{\partial \phi}{\partial n}|_{\Gamma_1} = g_1,$$

i.e. (3.4), which completes the proof of the Theorem.

REMARK 3.6. *A similar Theorem can be proved for the problem mentioned in Remark 3.4.*

3.4 Comments

The solution of subsonic flow problems via EVIs like (3.8) (3.13) is considered in CIAVALDINI-POGU-TOURNEMINE [2] (using a stream function approach) and in GLOWINSKI-MARROCCO [3].

Iterative methods for solving these EVIs may be found in the above reference and also in Chap. 5 of these notes.

Chapter 5

Decomposition–Coordination methods by augmented largrangian. Applications[1]

1 Introduction

1.1 Motivation

A large number of problems in Mathematics, Physics, Mechanics, Economics, etc... may be formulated as **171**

$$\min_{v \in V} \{F(Bv) + G(v)\} \qquad \text{(P)}$$

where

- V, H are topological vector spaces,

- $B \in \mathscr{L}(V, H)$,

- $F : H \to \overline{\mathbb{R}}$, $G : V \to \overline{\mathbb{R}}$ are convex, proper, l.s.c. functionals.

Let us give two examples taken from Chapter 2.

[1] This Chapter follows FORTIN-GLOWINSKI [1].

Example 2. *It is the* Bingham flow problem *of Chapter 2, Sec. 6; we recall that* Ω *being a bounded domain of* $\mathbb{R}^2$*, we consider the variational problem*

$$\min_{v\in H_0^1(\Omega)} \{\frac{\nu}{2}\int_\Omega |\nabla v|^2 dx + g\int_\Omega |\nabla v| dx - \int_\Omega fv dx\} \tag{1.1}$$

where ν *and* g *are* positive *constants. Then* (1.1) *is the particular problem* (P) *in which*

- $V = H_0^1(\Omega)$,
- $H = L^2(\Omega)\times L^2(\Omega)$,
- $B = \nabla$
- $F(q) = \frac{\nu}{2}\int_\Omega |q|^2 dx + g\int_\Omega |q| dx, (|q| = \sqrt{q_1^2+q_2^2}$,
- $G(v) = -\int_\Omega fv dx$.

172 *Actually we can also take*

- $F(q) = g\int_\Omega |q| dx$,
- $G(v) = \frac{\nu}{2}\int_\Omega |\nabla v|^2 dx - \int_\Omega fv dx$.

Example 3. *It is* elastic-plastic torsion problem *of Chapter 2, Sec. 3;* Ω *being still bounded in* $\mathbb{R}^2$*, we consider*

$$\min_{v\in K}\{\frac{1}{2}\int_\Omega |\nabla v|^2 dx - \int_\Omega fv dx\} \tag{1.2}$$

where

$$K = \{v\in H_0^1(\Omega), |\nabla v| \le 1 a.e.\}.$$

Problem (1.2) is the particular problem (P) in which

- $V = H_0^1(\Omega), H = L^2(\Omega)\times L^2(\Omega)$,
- $B = \nabla$,

- $F(q) = \frac{1}{2}\int_\Omega |q|^2 dx + I_{\hat{K}}(q)$,

- $G(v) = -\int_\Omega fv dx$,

where $I_{\hat{K}}$ is the *indicator functional* of the convex set

$$\hat{K} = \{q \in H, |q| \leq 1 a.e.\},$$

i.e.

$$I_{\hat{K}}(q) = \begin{cases} 0 \text{ if } q \in K \\ \\ +\infty \text{ if } q \notin K. \end{cases}$$

We can also take

- $F(q) = I_{\hat{K}}$,

- $G(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv dx$.

Orientation. Problems of type (P) have a special structure and in the sequel we shall introduce iterative methods of solution taking it into account.

1.2 Principle of the methods

The decomposition-coordination methods to follow are based on the following obvious *equivalence* result: **173**

THEOREM 1.1. (P) equivalent to

$$\min_{\{v,q\}\in W} \{F(q) + G(v)\} \qquad (\Pi)$$

where

$$W = \{\{v, q\} \in V \times H, Bv - q = 0\}.$$

We shall assume in the sequel that V and H are *Hilbert spaces* with inner products and norms respectively denoted by $((\cdot,\cdot))$, $\|.\|$ and $(\cdot,\cdot)$ and $|.|$. We define then a Lagrangian functional $\mathscr{L}$ associated to (π), by

$$\mathscr{L}(v, q, \mu) = F(q) + G(v) + (\mu, Bv - q), \qquad (1.3)$$

and for $r \geq 0$ an *Augmented Lagrangian* $\mathscr{L}_r$ by

$$\mathscr{L}_r(v, q, \mu) = \mathscr{L}(v, q, \mu) + \frac{r}{2}|Bv - q|^2. \tag{1.4}$$

REMARK 1.1. *Augmented Lagrangian methods for solving general optimization problems have been introduced by HESTENES [1], POWELL [1]. Augmented Lagrangian methods for solving problems like* (P) *via* (π) *have been introduced by GLOWINSKI-MARROCCO [237] and also [5]- [7].*

2 Properties of (P) And of The Saddle-Points of $\mathscr{L}$ And $\mathscr{L}_r$

2.1 Existence and uniqueness properties for (P).

Let define $J : V \to \overline{\mathbb{R}}$ by

$$J(v) = F(Bv) + G(v).$$

Then (P) can also be written

$$\begin{cases} J(u) \leq J(v) \ \forall v \in V, \\ u \in V. \end{cases} \tag{2.1}$$

174 Let $j : X \to \overline{\mathbb{R}}$; we define the so-called *domain* of $j(.)$ by

$$\text{dom } (j) = \{x \in X, \ j(x) \in \mathbb{R}\}$$

Then, if

$$\text{dom } (F \circ B) \cap \text{ dom } (G) \neq \phi, \tag{2.2}$$

J is *convex, proper, l.s.c.* Therefore, *sufficient* conditions for (P) to have a *unique solution* are (cf. CEA [1], [2], EKELAND-TEMAM [1]):

$$- \lim_{\|v\| \to +\infty} F(v) = +\infty,$$

$- F$ strictly convex .

REMARK 2.1. *If B is an injection from V to H, with R(B) (= range of B) closed in H, then |Bv| is a norm on V equivalent to ||v||.*

2.2 Properties of the saddle-points of $\mathscr{L}$ and $\mathscr{L}_r$

We have

THEOREM 2.1. *Let* $\{u, p, \lambda\}$ *be a saddle-point of* $\mathscr{L}$ *on* $V \times H \times H$*, then* $\{u, p, \lambda\}$ *is also a saddle-point of* $\mathscr{L}_r \; \forall r > 0$ *and conversely. Moreover* u *is solution of* (P) *and* $p = Bu$*.*

Proof. (i) Let $\{u, p, \lambda\}$ be a saddle-point of $\mathscr{L}$, then $\mathscr{L}(u, p, \lambda) \in \mathbb{R}$ and

$$\begin{cases} \mathscr{L}(u, p, \mu) \leq \mathscr{L}(u, p, \lambda) \leq \mathscr{L}(v, q, \lambda) \forall \{v, q, \mu\} \in V \times H \times H, \\ \\ \{u, p, \lambda\} \in V \times H \times H. \end{cases} \tag{2.3}$$

□

From the first inequality (2.3) and from (1.3) it follows that

$$(\mu, Bu - p) \leq (\lambda, Bu - p) \, \forall_{\mu} \in H,$$

which implies obviously that

$$Bu = p. \tag{2.4}$$

From the second inequality (2.3) and from (1.3), (2.4) it follows that **175**

$$J(u) = \mathscr{L}(u, p, \lambda) \leq (v, q, \lambda) \, \forall \{v, q\} \in V \times H \tag{2.5}$$

Taking $q = Bv$ in (2.5), it follows from (1.3) that

$$J(u) \leq \mathscr{L}(v, Bv, \lambda) = J(v) \, \forall v \in V : \tag{2.6}$$

hence u is solution of (P). Since $P = Bu$ we have

$$\mathscr{L}_r(u, p, \mu) = \mathscr{L}(u, p, \mu) = J(u) \, \forall \mu \in H; \tag{2.7}$$

it follows then from (2.3), (2.7) that

$$\mathscr{L}_r(u, p, \mu) = \mathscr{L}_r(u, p, \lambda) \leq \mathscr{L}(v, q, \lambda) \, \forall \{v, q, \mu\} \in V \times H \times H. \tag{2.8}$$

Since $\mathscr{L}_r(v, q, \mu) = \mathscr{L}(v, q, \mu) + \frac{r}{2}|Bv - q|^2$, we obtain from (2.8) that

$$\mathscr{L}_r(u, p, \mu) \leq \mathscr{L}_r(u, p, \lambda) \leq \mathscr{L}_r(v, q, \lambda) \; \forall \{v, q, \mu\} \in V \times H \times H, \quad (2.9)$$

which proves that $\{u, p, \lambda\}$ is also a saddle-point of $\mathscr{L}_r$ on $V \times H \times H$. To conclude this part of the proof we observe that from (2.3), $\{u, p\}$ is solution of

$$\begin{cases} \mathscr{L}(u, p, \lambda) \leq \mathscr{L}(v, q, \lambda) \; \forall \{v, q\} \in V \times H, \\ \{u, p\} \in V \times H, \end{cases} \quad (2.10)$$

from which it follows that $\{u, p\}$ is characterized (see CEA [1], [2], EKELAND - TEMAM [1])

$$\begin{cases} F(q) - F(p) - (\lambda, q - p) \geq 0 \; \forall q \in H, \\ p \in H, \end{cases} \quad (2.11)$$

$$\begin{cases} G(v) - G(u) + (\lambda, B(v - u)) \geq 0 \; \forall v \in V, \\ \\ u \in V. \end{cases} \quad (2.12)$$

176 (ii) Let $\{u, p, \lambda\}$ be a saddle-point of $\mathscr{L}_r$ with $r > 0$. Then as in part (i) this implies that $p = Bu$ and that u is solution of (P). Moreover, since $\{u, p, \lambda\}$ is solution of

$$\begin{cases} \mathscr{L}_r(u, p, \lambda) \leq \mathscr{L}_r(v, q, \lambda) \; \forall \{v, q\} \in V \times tH, \\ \\ \{u, p\} \in V \times H, \end{cases} \quad (2.13)$$

it is characterized by

$$\begin{cases} F(q) - F(p) + r(p - Bu, q - p) - (\lambda, q - p) \geq 0 \; \forall q \in H, \\ \\ p \in H, \end{cases} \quad (2.14)$$

$$\begin{cases} G(v) - G(u) + r(Bu - p, B(v-u)) + (\lambda, B(v-u)) \geq 0 \; \forall v \in V, \\ \\ u \in V. \end{cases} \tag{2.15}$$

But since $Bu - p = C$, (2.14), (2.15) reduce to (2.11), (2.12) and this fact implies that $\{u, p, \lambda\}$ satisfies (2.10). It follows then from (2.7) that $\{u, p, \lambda\}$ satisfies (2.3) and this completes the proof of the theorem.

3 Description of The Algorithms

It follows from Theorem 2.1 that a way of solving (P) is to solve the saddle point problem

$$\begin{cases} \mathscr{L}_r(u, p, \mu) \leq \mathscr{L}_r(u, p, \lambda) \leq (v, q, \lambda) \; \forall \{v, q, \mu\} \in V \times H \times H, \\ \\ \{u, p, \lambda\} \in V \times H \times H. \end{cases} \tag{3.1}$$

To do so we shall (See CEA [1], G.L.T [1, Ch. 2]) and algorithm of Uzawa's type and a variant of it.

3.1 First algorithm

We denote by ALG 1 the following algorithm: **177**

$$\lambda^0 \in H \text{ given }, \tag{3.2}$$

then λ^n known, we define u^n, p^n, λ^{n+1} by

$$\begin{cases} \mathscr{L}_r(u^n, p^n, \lambda^n) \leq \mathscr{L}_r(v, q, \lambda^n) \forall \{v, q\} \in V \times H, \\ \\ \{u^n, p^n\} \in V \times H, \end{cases} \tag{3.3}$$

$$\lambda^{n+1} = \lambda^n + \rho_n(Bu^n - p^n), \rho_n > 0. \tag{3.4}$$

The problem (3.3) is in fact equivalent to the following system of *two coupled variational inequalities* (of the second kind):

$$\begin{cases} G(v) - G(u^n) + (\lambda^n, B(v-u^n)) + r(Bu^n - p^n, B(v-u^n)) \geq 0 \ \forall v \in V, \\ \\ u^n \in V, \end{cases} \tag{3.5}$$

$$\begin{cases} F(q) - F(p^n) - (\lambda^n, q-p^n) + r(p^n - Bu^n, q-p^n) \geq 0 \ \forall q \in H, \\ \\ p^n \in H. \end{cases} \tag{3.6}$$

The convergence of ALG 1 will be studied in Sec. 4.

3.2 Second algorithm

The main drawback of ALG 1 is that it requires at each interaction the solution of the coupled EVIs (3.5), (3.6). To overcome thus difficulty it is natural to consider the following variant of ALG 1 (denoted ALG 2 in the following):

$$\{p^0, \lambda^1\} \in H \times H \text{ given }, \tag{3.7}$$

then $\{p^{n-1}, \lambda^n\}$ *known, we define* $\{u^n, p^n, \lambda^{n+1}\}$ *by*

$$\begin{cases} G(v) - G(u^n) + (\lambda^n, B(v-u^n)) + r(Bu^n - p^{n-1}, B(v-u^n)) \geq 0 \ \forall v \in V, \\ \\ u^n \in V, \end{cases} \tag{3.8}$$

$$\begin{cases} F(q) - F(p^n) - (\lambda^n, q-p^n) + r(p^n - Bu^n, q-p^n) \geq 0 \ \forall q \in H, \\ \\ p^n \in H, \end{cases} \tag{3.9}$$

178

$$\lambda^{n+1} = \lambda^n \rho_n (Bu^n - p^n), \rho_n > 0. \tag{3.10}$$

The convergence of ALG 2 will be studied in Sec. 5.

4 Convergence of Alg 1

4.1 General case

In this subsection V and H are possibly *infinite dimensional*; we assume that of course

$$\text{dom}\,(F \circ B) \cap \text{dom}\,(G) \neq \phi, \tag{4.1}$$

and also

$$B \text{ is an injection and } R(B) \text{ is closed in } H. \tag{4.2}$$

We assume also that

$$\lim_{|q| \to +\infty} \frac{F(q)}{|q|} = +\infty, \tag{4.3}$$

$$F = F_0 + F_1 \text{ with } F_0, F_1 \text{ convex , proper, l.s.c. ,} \tag{5.1}$$

$$\begin{cases} F_0 \text{ is Gateaux-differentiable and uniformly convex on the} \\ \\ \text{bounded sets of } H. \end{cases} \tag{4.5}$$

By definition we say that F_0 is *uniformly convex* on the bounded sets of H if the following property holds:

$$\begin{cases} \forall M > 0, \exists\, \delta_M : [0, 2M] \to \mathbb{R}, \text{continuous, strictly increasing with} \\ \delta_M(0) = 0, \text{ such that } \forall q, p \in H \text{ with } |p| \leq M, |q| \leq M \text{ we have} \\ (F_0'(q) - F_0', q - p) \geq \delta_M(|q - p|), \end{cases} \tag{4.6}$$

where $F_0' \triangledown F_0$ is the G-derivative of F_0. From the above properties, (P) has a *unique solution u* and we define $p \in H$ by $p = Bu$.

Exercise 4.1. *Prove that* (P) *is well-posed if* (4.1)–(4.5) *hold.* **179**

About the convergence of ALG 1 we have

THEOREM 4.1. *We assume that* $\mathscr{L}$ *has a saddle-point* $\{u, p, \lambda\} \in V \times H \times H$. *Then under the above assumption on B, F, G and if*

$$0 < \alpha_0 \leq \rho_n \leq \alpha_1 < 2r \tag{4.7}$$

the following convergence properties hold

$$u^n \to u \text{ strongly in } V, \tag{4.8}$$

$$P^n \to P = Bu \text{ strongly in } H, \tag{4.9}$$

$$\lambda^{n+1} - \lambda^n \to 0 \text{ strongly in } H, \tag{4.10}$$

$$\lambda^n \text{ is bounded in } H. \tag{4.11}$$

Moreover if λ^* is weak cluster point of $\{\lambda^n\}_n$ in H, then $\{u, p, \lambda^*\}$ is a saddle-point of $\mathscr{L}_r$ over $V \times H \times H$.

Proof. Since $\{u, p, \lambda\}$ is a saddle-point of $\mathscr{L}_r$ we have

$$\begin{cases} \mathscr{L}_r(u, p, \lambda) \leq \mathscr{L}_r(v, q, \lambda) \; \forall \{v, q\} \in V \times H, \\ \\ \{u, p\} \in V \times H. \end{cases} \tag{4.12}$$

Therefore $\{u, p\}$ is characterized by

$$\begin{cases} G(v) - G(u) + (\lambda, B(v-u)) + r(Bu - q, B(v-u)) \\ \qquad\qquad \geq 0 \; \forall v \in V, \\ u \in V, \end{cases} \tag{4.13}$$

$$\begin{cases} (F_0'(p), q - p) + F_1(q) - F_1(p) - (\lambda, q - p) \\ \qquad\qquad + r(p - Bu, q - p) \geq 0 \; \forall q \in H, \\ p \in H. \end{cases} \tag{4.14}$$

Moreover we have, from Theorem 2.1, $Bu = p$; therefore

$$\lambda = \lambda + \rho_n(Bu - p). \tag{4.15}$$

180 Let us define u^{-n}, p^{-n}, λ^{-n} by

$$u^{-n} = u^n - u, \; p^{-n} = p^n - p, \; \lambda^{-n} = \lambda^n - \lambda.$$

It follows then from (3.4), (4.15) that

$$\lambda^{-n+1} = \lambda^{-n} + \rho_n(Bu^{-n} - p^{-n}).$$

which implies

$$|\lambda^{-n+1}|^2 = |\lambda^{-n}|^2 + 2\rho_n(\lambda^{-n}, Bu^{-n} - p^{-n}) + \rho_n^2|Bu^{-n} - p^{-n}|^2$$

or, what will be more convenient,

$$|\lambda^{-n+1}|^2 = |\lambda^{-n}|^2 + 2\rho_n(\lambda^{-n}, Bu^{-n} - p^{-n}) + \rho_n^2|Bu^{-n} - p^{-n}|^2. \quad (4.16)$$

since $\{u^n, p^n\}$ is solution of (3.3) it is characterized by

$$\begin{cases} G(v) - G(u^n) + (\lambda^n, B(v - u^n)) + r(Bu^n - p^n, B(v - u^n)) \geq 0 \ \forall v \in V, \\ \\ u^n \in V, \end{cases} \quad (4.17)$$

$$\begin{cases} (F_0'(p^n), q - p^n) + F_1(q) - F_1(p^n) - (\lambda^n, q - p^n) \\ \qquad\qquad + r(p^n - Bu^n, q - p^n) \geq 0 \ \forall q \in H, \quad (4.18) \\ p^n \in H. \end{cases}$$

Taking $v = u$ (resp. . $v = u^n$) in (4.17) (resp. . (4.13)) and $q = p$ (resp. . $q = p^n$) in (4.18) (resp. . (4.14)) we obtain by addition

$$(\lambda^{-n}, Bu^{-n}) + r(Bu^{-n} - p^{-n}, Bu^{-n}) \leq 0, \quad (4.19)$$

$$(F_0'(p^n) - F_0'(p), p^{-n}) - (\lambda^{-n}, p^{-n}) + r(p^{-n} - Bu^{-n}, p^{-n}) \leq 0, \quad (4.20)$$

which imply, also by addition,

$$(\lambda^{-n}, Bu^{-n} - p^{-n}) + (F_0'(p^n) - F_0'(p), p^{-n}) + r|Bu^{-n} - p^{-n}|^2 \leq 0,$$

i.e.

$$-(\lambda^{-n}, Bu^{-n} - p^{-n}) \geq (F_0'(p^n) - F_0'(p), p^{-n}) + r|Bu^{-n} - p^{-n}|^2. \quad (4.21)$$

Combining (4.16) and (4.21) we obtain **181**

$$|\lambda^{-n}|^2 - |\lambda^{-n+1}|^2 \geq 2\rho_n(F_0'(p^n) - F_0'(p), p^{-n}) + \rho_n(2r - \rho_n)|Bu^{-n} - p^{-n}|^2 \geq 0. \quad (4.22)$$

Assuming that (4.7) holds it follows from (4.22) that

$$\lim_{n \to +\infty} |Bu^{-n} - p^{-n}| = 0, \quad (4.23)$$

$$\lim_{n\to+\infty}(F_0'(p^n) - F_0'(p).p^n - p) = 0, \tag{4.24}$$

$$\lambda^n \quad \text{is bounded in } H. \tag{4.25}$$

Since $p = Bu$ it follows from (4.23) that

$$\lim_{n\to+\infty} |Bu^n - p^n| = 0 \tag{4.26}$$

Since F is proper there exists $p_0 \in H$ such that $F(p_0) \in \mathbb{R}$; then from the characterisation (3.9) we have

$$F(p_0) - (\lambda^n, p_0) + r(p^n - Bu^n, p_0) \geq F(p^n) - (\lambda^n, p^n) + r(p^n - Bu^n, p^n). \tag{4.27}$$

Since λ^n and $p^n - Bu^n$are bounded, (4.27) implies

$$\beta_0 \geq F(p^n) - \beta_1|p^n|, \tag{4.28}$$

where β_0, β_1 are independent of n. It follows then from (4.3), (4.28) that

$$p^n \text{ is bounded in } H, \text{ i.e. } \exists M \text{ such that } |P^n| \leq M\ \forall n. \tag{4.29}$$

Then using the *uniform convexity* property (4.5), (4.6) of F_0 we obtain from (4.29) (assuming $M \geq |p|$) that

$$(F_0'(p^n) - F_0'(p), p^n - p) \geq \delta_M(|p^n - p|),$$

182 which implies, combined with (4.24), that

$$\lim_{n\to+\infty} \delta_M(|p^n - p|) = 0 \Leftrightarrow \lim_{n\to+\infty} |p^n - p| = 0. \tag{4.30}$$

It follows then from (4.26), (4.30) that

$$\lim_{n\to+\infty} Bu^n = P = Bu \text{ strongly in } H. \tag{4.31}$$

Since B is an injection with $R(B)$ closed in H, then (4.31) implies that

$$\lim_{n\to+\infty} u^n = u \text{ strongly in } V. \tag{4.32}$$

The convergence result (4.10) follows clearly from (4.7). (4.26). Let λ^* be a weak cluster point of $(\lambda^n)_n$ in H. Then passing to the limit in (3.5), (3.6), and using the l.s.c. property of F and G, we have

$$\begin{cases} G(v) + (\lambda^*, B(v-u)) + r(Bu - p, B(v-u)) \\ \qquad\qquad \geq \liminf G(u^n) \geq G(u)\ \forall v \in V, \\ u \in V, \end{cases}$$

$$\begin{cases} F(q) - (\lambda^*, q-p) + r(p - Bu, q-p) \\ \qquad\qquad \geq \liminf F(p^n) \geq F(p)\ \forall q \in H, \\ p \in H \end{cases}$$

i.e.

$$\begin{cases} G(v) - G(u) + (\lambda^*, B(v-u)) + r(Bu - p, B(v-u)) \geq 0\ \forall v \in V, \\ u \in V, \end{cases} \tag{4.33}$$

$$\begin{cases} F(q) - F(p) - (\lambda^*, q-p) + r(p - Bu - p, B(v-u)) \geq 0\ \forall v \in V, \\ p \in H. \end{cases} \tag{4.34}$$

As noticed before (see (2.13) - (2.15)) (4.33) is equivalent to

$$\begin{cases} \mathscr{L}_r(u, p, \lambda^*) \mathscr{L}_r(v, q, \lambda^*)\ \forall \{v, q\} \in V \times H, \\ \{u, p\} \in V \times H. \end{cases} \tag{4.35}$$

Since, from $p = Bu$, we have

$$\mathscr{L}_r(u, p, \mu) = \mathscr{L}(u, p, \mu) = J(u)\ \forall \mu \in H,$$

We obtain

$$\mathscr{L}_r(u, p, \mu) = \mathscr{L}_r(u, p, \lambda)\ \forall \mu \in H. \tag{4.36}$$

It follows clearly from (4.35), (4.36) that $\{u, p, \lambda^*\}$ is a saddle-point of $\mathscr{L}_r$; this completes the proof of the theorem. □ **183**

4.2 Finite dimensional case

If V and H are finite dimensional we have convergence of ALG 1 with weaker assumption on F, B, G than in Sec. 4.1. The reasons for this are the following:

(1) Since the constraints $Bv - q = 0$ is linear, if (p) has a solution then $\mathscr{L}$ and $\mathscr{L}_r$ have a saddle-point (see ROCKAFELLAR [1], CEA [1], [2]).

(2) $R(B)$ is *always closed*.

(3) It follows from CEA-GLOWINSKI [1] that F_0 satisfies the *uniform convexity property* (4.5), (4.6) *if* F_0 *is* C^1 *and strictly convex*.

(4) If F_0 is C^1 and strictly convex then F_0' is C^0 and *strictly monotone* i.e.

$$(F_0'(q_2) - F_0'(q_1), q_2 - q_1) > 0 \; \forall_{q_1}, q_2 \in H, q_1 \neq q_2.$$

Then if (P) has a solution, the property

$$\lim_{|q| \to +\infty} \frac{F(q)}{|q|} = +\infty$$

is not necessary. This is related to the following

Lemma 4.1. *Let H be finite dimensional and $A : H \to H$ be continuous and strictly monotone. Let $\{p^n\}_{n \geq 0}$, $p^n \in H \; \forall n$, and $p \in H$ be such that*

$$\lim_{n \to +\infty} (A(p^n) - A(p), p^n - p) = 0; \tag{4.37}$$

then

$$\lim_{n \to +\infty} p^n = p. \tag{4.38}$$

Proof. Assume that (4.38) does not hold. Then there exist $\delta > 0$ and a subsequence, denoted $(p^m)_m$, such that

$$|p^m - p| \geq \delta \; \forall m. \tag{4.39}$$

□

184 Let $S\left(p;\frac{\delta}{2}\right)=\left\{q\in H,|q-p|=\frac{\delta}{2}\right\}$. We define $z^m\in S\left(p;\frac{\delta}{2}\right)$ by

$$z^m=p+\frac{\delta}{2}\frac{p^m-p}{|p^m-p|};\tag{4.40}$$

$z^m\in]p,\ p^m[\subset H$ (see Fig. 4.1).

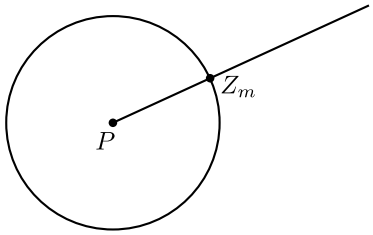

Figure 4.1:

We introduce $t^m=\dfrac{\delta}{2|p^m-p|}$; then

$$z^m=p+t^m(p^m-p)\tag{4.41}$$

and from (4.39)

$$0<t^m\le\frac{1}{2}.\tag{4.42}$$

Since A is *strictly monotone* we have

$$(A(p^m)-A,\,p^m-p)>(A(p+t(p^m-p))-A(p),\,p^m-p)\ \forall t\in]0,1[\tag{4.43}$$

Then, taking $t=t^m$ in (4.43), we obtain

$$(A(p^m)-A(p),\,p^m-p)>(A(z^m)-A(p),\,p^m-p)>0.\tag{4.44}$$

It follows then from (4.41), (4.42), (4.44) that

$$\begin{cases}(A(p^m)-A(p),\,p^m-p) & >\frac{1}{t^m}(A(z^m)-A(p),\,z^m-p)\\ & \ge 2(A(z^m)-A(p),\,z^m-p)>\\ & >(A(z^m)-A(p),\,z^m-p)>0.\end{cases}\tag{4.45}$$

Since $S(p, \frac{\delta}{2})$ is *compact* we can extract from $(z^m)_m$ a subsequence- still **185**
denoted $(z^m)_m$- such that

$$\lim_{m\to+\infty} z^m = z, \in S(p, \frac{\delta}{2}). \tag{4.46}$$

Since A is continuous it follows from (4.37), (4.46) that

$$(A(z) - A(p), z - p) = 0. \tag{4.47}$$

The *strict monotonicity* of A and (4.47) imply that $z = p$ which is impossible since $|p-z| = \frac{\delta}{2}$. Therefore (4.39) cannot hold $\Rightarrow \lim_{n\to+\infty} p^n = p$. From the above properties we can easily prove the following

THEOREM 4.2. *Assume that Vand H are finite dimensional and that* (P) *has a solution u. We suppose that*

- *B is an injection ,*
- *G is convex, proper, l.s.c.,*
- $F = F_0 + F_1$ *with* F_1 *convex, proper, l.s.c. over H and* F_0 *strictly convex and* C^1 *over H.*

Then (P) *has a unique solution and if*

$$0 < \alpha_0 \leq \rho_n \leq \alpha_1 < 2r$$

holds, we have for ALG 1 the following convergence properties.

$$\lim_{n\to+\infty} u^n = u,$$
$$\lim_{n\to+\infty} p^n = Bu,$$
$$\lim_{n\to+\infty} \lambda^{n+1} - \lambda^n = 0,$$
$$\lambda^n \text{ is bounded in } H.$$

Moreover if λ *is a cluster point of* $(\lambda^n)_n$, *then* $\{u, p, \lambda\}$ *is saddle point of* $\mathscr{L}_r$ *over* $V \times H \times H$.

4.3 Comment on the use of ALG 1. Further remarks.

186 Assume that r is *fixed* and what we use a *fixed* value ρ for ρ_n. Then from our computational experience it appears that the best convergence is obtained for $\rho = r$. About the choice of r it can be proved *theoretically* that the *larger* is r, the *faster* is the convergence; *practically* the situation is not so simple, for the following reasons:

The *larger* is r, the *worst* is the conditioning of the optimization problem (3.3) (or of the equivalent system (3.5), (3.6)). Then, since (3.3) is *numerically* (and not exactly) solved, at each iteration an *error* is made in the determination of $\{u^n, p^n\}$. The analysis of this error and the effect of it on the global behaviour of ALG 1 is a very complicated problem since we have to take into account the conditioning of (3.3), the stopping criterion of the algorithms (usually iterative) solving (3.3), round-off errors, etc ....

Fortunately it seems that the combined effect of all these factors is an algorithm which is not very sensitive to the choice of r (see GLOWINSKI - MARROCCO [6], FORTIN-GLOWINSKI [1] for more details).

Form a numerical point of view the only non-trivial part in the use of ALG 1 is the solution at each iteration of the above problem (3.3). Taking into account the particular structure of (3.3) it follows from CEA-GLOWINSKI [1], and CEA [2] that a method very well-suited to the solution of (3.3) is the *block relaxation* method described below:

All the problems (3.3) are of the following type:

$$\begin{cases} \mathscr{L}_r(u, p, \mu) \leq \mathscr{L}_r(v, q, \mu) \; \forall \{v, q\} \in V \times H, \\ \{u, p\} \in V \times H, \end{cases} \tag{4.48}$$

where μ is *given*. The minimization problem (4.48) is equivalent to the system

$$\begin{cases} G(v) - G(u) + (\mu, B(v-u)) + r(Bu - p, B(v-u)) \geq 0 \; \forall v \in V, \\ u \in V, \end{cases} \tag{4.49}$$

$$\begin{cases} F(q) - F(p) - (\mu q - p) + r(p - Bu, q - p) \geq 0 \; \forall q \in H, \\ p \in H. \end{cases} \tag{4.50}$$

187 Then a *block relaxation method* for solving (4.49), (4.50) is

$$\{u^0, p^0\} \text{ given,} \tag{4.51}$$

then $\{u^m, p^m\}$ *known, we obtain* $\{u^{m+1}, p^{m+1}\}$ *from*

$$\begin{cases} G(v) - G(u^{m+1}) + (\mu, B(v - u^{m+1})) \\ \qquad +r(Bu^{m+1} - p^m, B(v - u^{m+1})) \geq 0 \ \forall v \in V, \\ u^{m+1} \in V, \end{cases} \tag{4.52}$$

$$\begin{cases} F(q) - F(p^{m+1}) - (\mu, q - p^{m+1}) \\ \qquad +r(p^{m+1} - Bu^{m+1}, q - p^{m+1}) \geq 0 \ \forall q \in H, \\ p^{m+1} \in H. \end{cases} \tag{4.53}$$

Sufficient conditions for the convergence of (4.51)–(4.53) may be found in CEA-GLOWINSKI and CEA, loc. cit. .

In practice, when using (4.51)–(4.53), a stopping test of the following type will be used:

$$\max(\|u^{m+1} - u^m\|, |p^{m+1} - p^m|) \leq \in . \tag{4.54}$$

Another possibility is to stop after a *fixed number* of iterations. If for instance we stop after *only one* iteration of (4.51) - (4.53) an if at iteration n of ALG 1 we initialise with $\{u^{n-1}, p^{n-1}\}$ the computation of $\{u^n, p^n\}$ by (4.51)-(4.53), then we recover ALG 2.

REMARK 4.1. *Other relaxation methods can also be used; moreover it can be worthwhile to introduce* overrelaxation parameters *to increase the speed of convergence of* (4.51) - (4.53).

REMARK 4.2. *The choice* $\rho = r$ *may be motivated by the following*

188 **Proposition 4.1.** *Suppose that* $F(q) = \frac{1}{2}|q|^2$ *and that G is linear. Then* $\forall \lambda^0 \in H$ *we have for the sequence* $(u^n)_n$ *of ALG1, convergence to the solution u of* (P) *in less than three iterations if we use* $\rho_n = \rho = r$, *r given.*

Preliminary remark. *In the above situation we have* (P) *equivalent to*

$$B^t Bu = f \tag{5.2}$$

where $G(v) = ((f, v))\ \forall v \in V$. *Therefore using ALG 1 for solving* (p) *has no* practical *interest. But even in that trivial case we shall observe that the behaviour of ALG 1 is "interesting" since the convergence of* u^n *in a* finite number *of iterations* does not imply *a similar convergence for* p^n *and* λ^n.

Proof of proposition 5.1. *It follows from* (4.17), (4.18) *that in the particular case that we are considering, ALG 1 reduces to*

$$\lambda^0 \text{ given in } H, \tag{4.55}$$

$$rB^t Bu^n = rB^t p^n - B^t \lambda^n + f, \tag{4.56}$$

$$p^n = \lambda^n + r(Bu^n - p^n), \tag{4.57}$$

$$\lambda^{n+1} = \lambda^n + r(Bu^n - p^n). \tag{4.58}$$

We can easily prove that the unique saddle-point of $\mathscr{L}_r$ *over* $V \times H \times H$ *is* $\{u, Bu, Bu\}$, *i. e.* $p = Bu$, $\lambda = Bu$; *using the notation* $u^{-n} = u^n - u$, $p^{-n} = p^n - p$, $\lambda^{-n} = \lambda^n - \lambda$ *it follows from* (4.56) - (4.58) *that*

$$B^t \lambda^{-n} + rB^t(Bu^{-n} - p^{-n}) = 0 \ \ \forall n \geq 0, \tag{4.59}$$

$$\lambda^{n+1} = p^n, \Rightarrow \lambda^{-n+1} = p^{-n} \ \ \forall n \geq 0, \tag{4.60}$$

$$p^{-n} = \lambda^{-n} + r(Bu^{-n} - p^{-n}) \ \ \forall n \geq 0. \tag{4.61}$$

Multiplying (4.61) by B^t and comparing with (4.59) we obtain **189**

$$B^t\ p^{-n} = 0 \ \ \forall n \geq 0. \tag{4.62}$$

Since (4.60), (4.61) imply

$$p^{-n+1} = p^{-n} + r(Bu^{-n+1} - p^{-n+1}) \ \ \forall n \geq 0 \tag{4.63}$$

we obtain, multiplying by B^t and taking account of (4.62), that

$$B^t\ Bu^{-n+1} = 0 \ \ \forall n \geq 0 \Rightarrow Bu^{-n+1} = 0 \forall n \geq 0. \tag{4.64}$$

Since $|Bv|$ is a norm on V, (4.64) implies that $u^n = u \ \forall n \geq 1$. Hence the convergence of u^n to u requires at most two iterations. Using (4.63), (4.64) we have

$$p^{-n+1} = \frac{1}{1+r} p^{-n} \ \forall n \geq 0. \tag{4.65}$$

It follows from (4.65) that the larger is r the faster p^n converges to $p = Bu$; for more details on the convergence of p^n see FORTIN-GLOWINSKI [1].

5 Convergence of ALG 2

5.1 Orientation

We shall prove in this section that under fairly general assumptions on F and Gwe have convergence of ALG 2 if $0 < \rho_n = \rho < \frac{1+\sqrt{5}}{2} r$. We do not know if this result is optimal since in some cases (G linear, for example) the upper bound of the interval of convergence is 2r. Actually this question is rather academic since in the various experiments we have done with ALG 2, the optimal choice seems to be $\rho = r$.

5.2 General case

We study the convergence of ALG 2 with the same hypotheses of B, F, G as in Sec. 4.1. We have then

THEOREM 5.1. *We suppose that $\mathscr{L}_r$ has a saddle-point $\{u, p, \lambda\}$ over $V \times H \times H$. Then if the assumptions on B, F, G are those of Sec. 4.1 and if*

$$0 < \rho_n = \rho < \frac{1+\sqrt{5}}{2} r, \tag{5.1}$$

190 *we have the following convergence properties:*

$$u^n \to u \ \text{ strongly in } \ V, \tag{5.2}$$

$$p^n \to p \ \text{ strongly in } \ H, \tag{5.3}$$

$$\lambda^{n+1} - \lambda^n \to 0 \ \text{ strongly in } \ H, \tag{5.4}$$

$$\lambda^n \text{ is bounded in } H. \tag{5.5}$$

Moreover if λ^* is a weak cluster point of $(\lambda^n)_n$, then $\{u, p, \lambda\}$ is a saddle-point of $\mathscr{L}_r$ over $V \times H \times H$.

Proof. Let us still define u^{-n}, p^{-n}, λ^{-n} by

$$u^{-n} = u^n - u, p^{-n} = p^n - p, \lambda^{-n} = \lambda^n - \lambda.$$

□

Since $\{u, p, \lambda\}$ is a saddle-point of $\mathscr{L}_r$ over $V \times H \times H$. we have

$$G(v) - G(u) + (\lambda, B(v-u)) + r(Bu - p, B(v-u)) \geq 0 \quad \forall v \in V, \tag{5.6}$$

$$(F_0'(p), q-p) + F_1(q) - F_1(p) - (\lambda, q-p) + r(p - Bu, q-p) \geq 0 \quad \forall q \in H. \tag{5.7}$$

$$\lambda = \lambda + \rho(Bu - p). \tag{5.8}$$

Moreover, (3.8)–(3.10) imply

$$G(v) - G(u^n) + (\lambda^n, B(v-u^n)) + r(Bu^n - p^{n-1}, B(v-u^n)) \geq 0 \, \forall v \in V, \tag{5.9}$$

$$(F_0'(p^n), q-p^n) + F_1(q) - F_1(p^n) - (\lambda^n, q-p^n) + r(p^n - Bu^n, q-p^n) \geq \circ \, \forall q \in H, \tag{5.10}$$

$$\lambda^{n+1} = \lambda^n + \rho(Bu^n - p^n). \tag{5.11}$$

Taking $v = u^n$ (resp. . $v = u$) in (5.6) (resp. . (5.9)) and $q = p^n$ (resp. . $q = p$) in (5.7) (resp. . (5.10)) we obtain by addition **191**

$$r(B\overline{u}^n - \overline{p}^{n-1}, B\overline{u}^n) + (\overline{\lambda}^n, B\overline{u}^n) \leq 0, \tag{5.12}$$

$$(F_0'(p^n) - F_0'(p), \overline{p}^n) + r(\overline{p}^n - B\overline{u}^n, \overline{p}^n) - (\overline{\lambda}^n, \overline{p}^n) \leq 0. \tag{5.13}$$

By addition of (5.12), (5.13) it follows that

$$(F_0'(p^n) - F_0'(p), p^n - p) + r|B\overline{u}^n - \overline{p}^n|^2 + (\overline{\lambda}^n, B\overline{u}^n - \overline{p}^n) + r(\overline{p}^n - \overline{p}^{n-1}, B\overline{u}^n) \leq . \tag{5.14}$$

By subtracting (5.8) from (5.11) we obtain

$$|\overline{\lambda}^n|^2 - |\overline{\lambda}^{n+1}|^2 = -2\rho(B\overline{u}^n, -\overline{p}^n, \overline{\lambda}^n) - \rho^2|B\overline{u}^n - \overline{p}^n|^2. \tag{5.15}$$

It follows then from (5.14) , (5.15) that

$$\begin{aligned}|\overline{\lambda}^n|^2 - |\overline{\lambda}^{n+1}|^2 \geq\ & 2\rho(F_0'(p^n) - F_0'(p), \overline{p}^n) \\ & + \rho(2r - \rho)|B\overline{u}^n - \overline{p}^n|^2 + 2\rho r(\overline{p}^n - \overline{p}^n, B\overline{u}^n).\end{aligned} \tag{5.16}$$

Starting from

$$B\overline{u}^n = (B\overline{u}^n - B\overline{u}^{n-1}) + (B\overline{u}^{n-1} - \overline{p}^{n-1}) + \overline{p}^{n-1}$$

we obtain

$$\begin{aligned}(B\overline{u}^n, \overline{p}^n - \overline{p}^{n-1}) =\ & (B\overline{u}^n - B\overline{u}^{n-1}, \overline{p}^n - \overline{p}^{n-1}) \\ & + (B\overline{u}^{n-1}, -\overline{p}^{n-1} - \overline{p}^{n-1}) + (\overline{p}^{n-1}, \overline{p}^n - \overline{p}^{n-1}).\end{aligned} \tag{5.17}$$

Since

$$(\overline{p}^{n-1}, \overline{p}^n - \overline{p}^{n-1}) = \frac{1}{2}(|\overline{p}^n|^2 - |\overline{p}^{n-1}|^2 - |\overline{p}^n - \overline{p}^{n-1}|^2).$$

it follows from (5.17) that

$$\begin{cases} 2\rho r(B\overline{u}^n, \overline{p}^n - \overline{p}^{n-1}) = & 2\rho r(B\overline{u}^n - B\overline{u}^{n-1}, \overline{p}^n - \overline{p}^{n-1}) \\ & +2\rho r(B\overline{u}^{n-1} - \overline{p}^{n-1}, \overline{p}^n - \overline{p}^{n-1}) + \\ & \rho r(|\overline{p}^n|^2 - |\overline{p}^{n-1}|^2 - |\overline{p}^n - \overline{p}^{n-1}|^2). \end{cases} \tag{5.18}$$

Taking (5.10) ar $n-1$ instead of n, we have

$$\begin{cases} (F_0'(p^{n-1}), q - p^{n-1}) + F_1(q) - F_1(p^{n-1}) - (\lambda^{n-1}, q - p^{n-1}) + \\ +r(p^{n-1} - Bu^{n-1}, q - p^{n-1}) \geq 0. \end{cases} \tag{5.19}$$

192 Taking $q = p^{n-1}$ in (5.10) and $q = p^n$ in (5.19) we obtain by addition

$$\begin{cases} (F_0'(p^n) - F_0'(p^{n-1}), p^n - p^{n-1}) - (\overline{\lambda}^n - \overline{\lambda}^{n-1}, \\ \qquad -\overline{p}^n - \overline{p}^{n-1}) + r|\overline{p}^n - \overline{p}^{n-1}|^2 - \\ -r(B\overline{u}^n - B\overline{p}^{n-1}, \overline{p}^n - \overline{p}^{n-1}) \leq 0. \end{cases} \tag{5.20}$$

But since F_0' is monotone, it follows from (5.20) that

$$r|\overline{p}^n-\overline{p}^{n-1}|^2-(\overline{\lambda}^n-\overline{\lambda}^{n-1},\overline{p}^n-\overline{p}^{n-1})-r(B\overline{u}^n-B\overline{p}^{n-1},\overline{p}^n-\overline{p}^{n-1})\leq 0 \quad (5.21)$$

We have (from (3.10))

$$\lambda^n = \lambda^{n-1} + \rho(Bu^{n-1-p^{n-1}})$$

which implies that

$$\overline{\lambda}^n - \overline{\lambda}^{n-1} = \rho(B\overline{p}^{n-1} - \overline{p}^{n-1}). \quad (5.22)$$

It follows then from (5.21), (5.22) that

$$r|\overline{p}^n - \overline{p}^{n-1}|^2 - \rho(B\overline{u}^{n-1},\overline{p}^n - \overline{p}^{n-1}) - r(B\overline{u}^n - B\overline{u}^{n-1},\overline{p}^n - \overline{p}^{n-1}) \leq 0$$

i. e.

$$r(bu^{-n} - B\overline{u}^{n-1},\overline{p}^n - \overline{p}^{n-1}) \geq r|\overline{p}^n - \overline{p}^{n-1}|^2 - \rho(B\overline{u}^{n-1} - \overline{p}^{n-1} - \overline{p}^n - \overline{p}^{n-1}). \quad (5.23)$$

It' follows then from (5.18) , (5.23) that

$$\begin{cases} 2\rho r(B\overline{u}^n,\overline{p}^n - \overline{p}^{n-1}) & \geq \rho r(|\overline{p}^n|^2 - |\overline{p}^{n-1}|^2) + \rho r|\overline{p}^n - \overline{p}^{n-1}|^2 \\ & +2\rho(r-\rho)(B\overline{u}^{n-1} - \overline{p}^{n-1},\overline{p}^{n-1},p^n - \overline{p}^{n-1}). \end{cases} \quad (5.24)$$

Finally, combining (5.16), (5.24) we obtain

$$\begin{cases} (|\lambda^{-n}|^2 + \rho r|\overline{p}^{n-1}|^2) - (|\overline{\lambda}^{n+1}|^2 + \rho r|\overline{p}^n|^2) \geq 2\rho(F_0'(P^n) - F_0'(p),\overline{p}^n)+ \\ +\rho(2r-\rho)|B\overline{u}^n - \overline{p}^n|^2 + \rho r|\overline{p}^n - \overline{p}^{n-1}|^2 \\ \qquad +2\rho(r-\rho)(B\overline{u}^{n-1} - \overline{p}^{n-1},\overline{p}^n - \overline{p}^{n-1}). \end{cases} \quad (5.25)$$

Using the *Schwartz's inequality* it follows from (5.25) that $\forall\alpha > 0$ we **193** have

$$\begin{cases} (|\overline{\lambda}^{-n}|^2 + \rho r|\overline{p}^{n-1}|^2) - (|\lambda^{-n+1}|^2 + \rho r|\overline{p}^n|^2) \geq 2\rho(F_0'(p^n) - F_0'(p),\overline{p}^n)+ \\ +\rho(2r-\rho)|B\overline{u}^n - \overline{p}^n|^2 + \rho r|\overline{p}^n - \overline{p}^{n-1}|^2 - \rho|r-\rho| \\ \qquad (\frac{1}{\alpha}|B\overline{u}^{n-1} - \overline{p}^{n-1}|^2 + \alpha|\overline{p}^n - \overline{p}^{n-1}|^2). \end{cases} \quad (5.26)$$

It $\rho = r$ it is clear that using the same method as in the proof of Theorem 4. 1. we have (5.2) - (5.5). If $0 < \rho < r$, taking $\alpha = 1$ and observing that $|r = \rho| = r - \rho$, we observe that we have from (5.26)

$$\begin{cases} (|\overline{\lambda}^n|^2 + \rho r|\overline{p}^{n-1}|^2 + \rho(r-\rho)B\overline{u}^{n-1} - \overline{p}^{n-1}|^2) \\ \qquad -(|\overline{\lambda}^{n+1}|^2 + \rho r|\overline{p}^n|^2 + \rho(r-\rho)|B\overline{u}^n - \overline{p}^n|^2) \\ \geq 2\rho(F_0'(p^n) - F_0'(p), \overline{p}^n) + \rho r|B\overline{u}^n - \overline{p}^n|^2 + \rho^2|\overline{p}^n - \overline{p}^{n-1}|^2 \geq 0. \end{cases}$$

which implies clearly (5.2) -(5.5).

If $\rho > r$ we have $|r - \rho| = \rho - r$ and then it follows from (5.26) that (5.2) -(5.5) holds, if we have $\rho < \rho_M$ where

$$\begin{cases} \rho_M(2r - \rho_M) = \frac{1}{\alpha}\rho_M(\rho_M - r). \\ \rho_M r = \alpha\rho_M(\rho_M - r). \end{cases} \tag{5.27}$$

By elimination of α it follows from (5.27) that

$$\rho_M^2 - r\rho_M - r^2 = 0$$

i. e. (since $\rho_M > 0$)

$$\rho_M = \frac{1 + \sqrt{5}}{2} r.$$

Then using basically the same method as in the proof of Theorem 4.1 we can easily prove, from (5.2)-(5.5), that $\{u, Bu, \lambda^*\}$ is a saddle-point of $\mathscr{L}_r$ over $V \times H \times H$ if λ^* is a weak cluster point of $(\lambda^n)_n$.

5.3 Finite dimensional case

194 Using a variant of the proof of Theorem 5.1, and Lemma 4.1, we can easily prove

THEOREM 5.2. *Assume that the assumptions on V, H, F, B, G are those of the statement of Theorem 4.2. Then if*

$$0 < \rho_n = \rho < \frac{1 + \sqrt{5}}{2} r$$

the conclusions of the statement of Theorem 4.2 still hold.

5.4 Comments on the choice of ρ and r

5.4.1 Some remarks

REMARK 5.1. *If G is linear it has been proved by GABAY-MERCIER [1] that ALG 2 converges if*

$$0 < \rho_n = \rho < 2r.$$

The proof of this result is rather technical and an open question is to decide if it can be extended to the more general cases we have considered in these notes.

REMARK 5.2. *If G is is linear we observe that the step* (3.8) *of ALG 2 is a linear problem related to the self adjoint operator B^tB. Therefore in the finite dimensional case, assuming B injective, it will be convenient to factorize(by a Cholesky method, for example) the symmetric, positive definite matrix B^tB once and for all, before starting the iterations of ALG 2.*

5.4.2 On the choice of ρ and r

If r is given our computational experience seems to indicate that the best choice for ρ is $\rho = r$. The choice of r is not clear and ALG 2 appears to be *more sensitive* to the choice of r than ALG 1. By the way, ALG 1 seems to be more robust on *very stiff* problem than ALG 2; we mean that the choice of the parameter is less critical and that the *computational time* with ALG 1 may become *much shorter* than with ALG 2.

REMARK 5.3. *We have seen in Remark 4.2 that if $F(q) = \frac{1}{2}|q|^2$ and G* **195**
is linear, the sequence $\{u^n\}_n$ related to ALG 1 converges in two iterations (at most) if we use $\rho = r$. If we use ALG 2 with the same hypotheses on F, G then we have convergence of $\{u^n\}_n$ in two iterations at most, only if $\rho = r = 1$ (for any choice of $\{p^0, \lambda^1\}$). This fact also confirms the greater robustness of ALG 1.

6 Applications

6.1 Bingham flow in a cylindrical pipe

It is the problem considered in CH. 2, Sec. 6 and also in Sec. 1. 1 of this Chapter (we recall that Ω is a *bounded* domain of $\mathbb{R}^2$):

$$\min_{v\in H_0^1(\Omega)} \left\{ \frac{\nu}{2} \int_\Omega |\nabla v|^2 dx + g \int_\Omega |\nabla v| dx - \int_\Omega f v dx \right\}. \tag{6.1}$$

Then (3.1) is a particular (P) problem corresponding to

$$V = H_0^1(\Omega), H = L^2(\Omega) \times L^2(\Omega), B = \nabla, \tag{6.2}$$

$$F(q) = \frac{\nu}{2} \int_\Omega |q|^2 \, dx + g \int_\Omega |q| dx, \tag{6.3}$$

$$G(v) = -\int_\Omega f v \, dx. \tag{6.4}$$

Moreover we have $F = F_0 + F_1$ with

$$F_0(q) = \frac{\nu}{2} \int_\Omega |q|^2 dx, F_0'(q) = \nu q, \tag{6.5}$$

$$F_1(q) = g \int_\Omega |q| dx. \tag{6.6}$$

it follows then from (6.2) -(6.6) that the various assumptions required to apply Theorem 4.1 and 5.1 are satisfied. Therefore we can solve (6.1) by ALG 1 and ALG 2. Moreover since G is linear the GABAY-MERCIER[1] result holds (see Remark 5.1) and ALG 2 converges if $0 < \rho_n = \rho < 2r$. The augmented Lagrangian $\mathscr{L}_r$ to be used in this case is given by

$$\begin{cases} \mathscr{L}_r(v, q, \mu) & = \frac{\nu}{2} \int_\Omega |q|^2 dx + g \int_\Omega f v dx + \int_\Omega \mu \cdot (\nabla v - q) dx \\ & \quad + \frac{r}{2} \int_\Omega |\nabla v - q|^2 dx. \end{cases} \tag{6.7}$$

196 ***Solution of*** (6.1) ***by ALG 1.***

When applying ALG 1 to the solution of (6.1) it follows from (3.2)-(3.4), (6.7) that we have

$$\lambda^0 \in L^2(\Omega) \times L^2(\Omega), \text{arbitrarily given}, \tag{6.8}$$

then for $n \geq 0$,

$$\begin{cases} -r\Delta u^n &= f + \nabla \cdot \lambda^n - r\nabla \cdot p^n \text{on}\Omega, \\ u^n|_\Gamma &= 0, \end{cases} \tag{6.9}$$

$$\begin{cases} p^n(x) &= 0(\text{if} g \geq |\lambda^n 9x) + r\nabla u^n(x)|, \\ p^n(x) &= \dfrac{\lambda^n(x) + r\nabla u^n(x)}{\nu + r}\left(1 - \dfrac{g}{|\lambda^n(x) + r\nabla u^n(x)|}\right) \text{elsewhere}, \end{cases} \tag{6.10}$$

$$\left\{\lambda^{n+1} = \lambda^n + \rho_n(\nabla u^n - p^n). \right. \tag{6.11}$$

Solution of (6.1) **by ALG**

We have to replace (6.8) by

$$\{p^0, \lambda^1\} \text{arbitrarily given in} (L^2(\Omega))^2 \times (L^2(\Omega))^2, \tag{6.12}$$

and (6.9) by

$$\begin{cases} -r\Delta u^n = f + \nabla \cdot \lambda^n - r\nabla p^{n-1} \text{on}\Omega. \\ u^n|_\Gamma = 0. \end{cases} \tag{6.13}$$

REMARK 6.1. *In practice* (6.8)–(6.11) *and* (6.12), (6.13), (6.10), (6.11) *will be applied to* finite element *or* finite difference *approximations of* (6.1). *It follows then from* (6.9), (6.13) *that it is easy to use either ALG 1 (combined with the* block relaxation method *of Sec. 4.3) or ALG 2, once we have at our disposal an efficient program for solving approximate Dirichlet problems for* $-\Delta$.

Bibliographical comments. Numerical solutions of (6.1) by ALG 1 **197**
and ALG 2 may be found in GABAY-MERCIER [1, GLOWINSKI-MARROCCO [3]; we can also find in FORTIN [2] and iterative method of solution of (6.1), close to ALG 2 but obtained by a different approach.

6.2 Elastic-plastic torsion of a cylindrical bar

It is the problem of Chap. 2, Sec. 3, also considered in Sec. 1.1 (Ω is a bounded domain of $\mathbb{R}^2$ in the sequel):

$$\min_{v\in K}[\frac{1}{2}\int_\Omega |\nabla v|^2 dx - \int_\Omega fv dx], \tag{6.14}$$

where $K = \{v | v \in H_0^1(\Omega), |\nabla v| \leq 1 a.e.\}$; (6.14) is a particular (P) problem corresponding to

$$V = H_0^1(\Omega), H = L^2(\Omega)\times L^2(\Omega), B = \nabla, \tag{6.15}$$

$$G(V) = -\int_\Omega fv\, dx, \tag{6.16}$$

$$F = F_0 + F_1, \tag{6.17}$$

where

$$F_0(q) = \frac{1}{2}\int_\Omega |q|^2 dx \Rightarrow F_0'(q) = q, \tag{6.18}$$

$$F_1(q) = L_{\hat{K}}(q) \tag{6.19}$$

with $\hat{K} = \{q \in H, |q| \leq 1 a.e.\}$ and $L_{\hat{K}}$ the *indicator functional* of $\hat{K}$ i.e.

$$L_{\hat{K}}(q) = \begin{cases} 0 & \text{if } q \in \hat{K}, \\ +\infty & \text{if } q \notin \hat{K}. \end{cases} \tag{6.20}$$

Here too, it follows from (6.15) - (6.20) that the various assumptions required to apply Theorem 4.1 and 5.1 are satisfied. Therefore we can solve (6.14) by ALG 1 and ALG 2. Moreover from the linearity of G we have the convergence of ALG 2 if $0 < \rho_n = \rho < 2r$. In the present case $\mathscr{L}_r$ is given by

$$\mathscr{L}_r(v,q,\mu) = \frac{1}{2}\int_\Omega |q|^2 dx + I_{\hat{K}}(q) - \int_\Omega fv\, dx + \int_\Omega \mu\cdot(\nabla v - q)dx + \frac{r}{2}\int_\Omega |\nabla v - q|^2 dx. \tag{6.21}$$

198

Solution of (6.1) **by ALG 1.**

It follows from (3.2)–(3.4), (6.21) that when applying ALG 1 to (6.14) we obtain

$$\lambda^0 \text{arbitrarily given in} (L^2(\Omega))^2, \tag{6.22}$$

then for $n \geq 0$,

$$\begin{cases} -r\Delta u^n = f + \Delta \cdot \lambda^n - r\nabla \cdot p^n \text{ on } \Omega, \\ u^n|_\Gamma = 0, \end{cases} \tag{6.23}$$

$$p^n = \frac{\lambda^n + r\nabla u^n}{\underline{\sup}(1 + r, |\lambda^n + ru^n|)}, \tag{6.24}$$

$$\lambda^{n+1} = \lambda^n + \rho_n(\nabla u^n - p^n). \tag{6.25}$$

Solution of (6.1) **by ALG**

We have to replace (6.22) by (6.21) and (6.23) by (6.13). Still applies to (6.14) and numerical solutions of (6.) by ALG 1, ALG 2 may be found in GLOWINSKI-MARROCCO [3], GABAY-MERCIER [1].

6.3 A nonlinear Dirichlet problem

We follow in this section GLOWINKSKI-MARROCCO [6]; Let us consider $1 < s < +\infty$ and

$$W_0^{1,s}, (\Omega) = \overline{\mathscr{D}(\Omega)}W^{1,s}(\Omega) = \{v \in W^{1,s}(\Omega), v|_\Gamma = 0\},$$

where Ω is a *bounded* domain of $\mathbb{R}^N$.

Then we consider on Ω the following nonlinear Dirichlet problem:

$$\begin{cases} -\nabla \cdot (|\nabla u|^{s-2}\nabla u) = f, \\ u|_\Gamma = 0, \end{cases} \tag{6.26}$$

where $f \in V' = W^{-1,s'}(\Omega)(\frac{1}{s} + \frac{1}{s'} = 1 \Rightarrow s' = \frac{s}{s-1})$. it can be proved (see, 199

for instance GLOWINKSI-MARROCCO [6]) that (6.26) has a unique solution which is also the solution of

$$\min_{v\in W_0^{1,s}(\Omega)}\left[\frac{1}{s}\int_\Omega |\nabla v|^s dx - < f, v >\right]. \tag{6.27}$$

We observe that $W_0^{1,s}(\Omega)$ *is not an Hilbert space* if $s \neq 2$, therefore we cannot apply Theorems 4.1 and 5.1 to the iterative solution of (6.27). Nevertheless once (6.27) has been approximated by a convenient finite element or finite difference method it is possible to apply the above theorems (or Theorems 4.2, 5.2) to the iterative solution of the approximate problem. For the sake of simplicity we shall confine our study to the continuous problem, since it has simpler notation. We have

$$V = W_0^{1,s}(\Omega), H = (L^s(\Omega))^N, H' = (L^{s'}(\Omega))^N, B = \nabla,$$
$$F(q) = F_0(q) = \frac{1}{s}\int_\Omega |q|^s dx, F'(q) = q|q|^{s-2},$$
$$G(v) = < f, v > .$$

We observe that

$$\lim_{|q|_s\to+\infty}\frac{F(q)}{|q|_s} = +\infty,$$

where

$$|q|_s = (\int_\Omega |q|^s dx)^{1/s} = \|q\|^N_{(L^s(\Omega)}.$$

We also have $\forall p, q \in H$:

$$(F'(q) - F'(p), q - p) \geq \alpha|q - p|_s^s \text{ if } s \geq 2, \tag{6.28}$$
$$(F'(q) - F'(p), q - p) \geq \alpha\frac{|q - p|_s^2}{(|p|_s + |q|_s)^{2-s}} \text{ if } 1 < s \leq 2, \tag{6.29}$$
$$|F'(q) - F'(p)|_{s'} \geq \beta(|p|_s + |q|_s)^{s-2}|q - p|_s \text{ if } s \leq 2, \tag{6.30}$$
$$|F'(q) - F'(p)|_{s'} \geq \beta|q - p|_s^{s-1} \text{ if } 1 < s \leq 2, \tag{6.31}$$

where α, β are independent of p, q and are strictly positive.

200 **Exercise 6.1.** *Prove* (6.28) - (6.31).

We refer to GLOWINSKI - MARROCCO [6] for a detailed analysis, including error estimates, of a finite element approximation of (6.26), (6.27) (see also CIARLET [2]).

From our numerical experience it appears that solving (6.26), (6.27) if s is close to 1 (say $1 < s < 1.3$) or large (say $s > 5$) is a very difficult task if one uses standard iterative methods; to our knowledge the only very efficient methods are ALG 1 and ALG 2 (or closely related algorithms; see GLOWINSKI-MARROCCO, loc. cit., for more details). The augmented Lagrangian $\mathscr{L}_r$ to used for solving (6.26), (6.27) is defined by

$$\mathscr{L}_r(v,q,\mu) = \frac{1}{s}\int_\Omega |q|^s dx - < f, v > + \frac{r}{2}\int_\Omega |\nabla v - q|^2 dx + \int_\Omega \mu\cdot(\nabla v - q)dx. \tag{6.32}$$

Solution of (6.1) **by ALG 1.**

It follows from (3.2)–(3.4), (6.32) that when applying ALG 1 to (6.6), (6.27) we obtain

$$\lambda^0 \in (L^{s'}(\Omega))^N, \tag{6.33}$$

then for $n \geq 0$

$$\begin{cases} -r\Delta u^n = f + \nabla\cdot\lambda^n - r\nabla\cdot p^n \text{ in } \Omega, \\ u^n|_\Gamma = 0, \end{cases} \tag{6.34}$$

$$|p^n|^{s-2}p^n + rp^n = r\nabla u^n + \lambda^n, \tag{6.35}$$

$$\lambda^{n+1} = \lambda^n + \rho_n(\nabla u^n - p^n). \tag{6.36}$$

The nonlinear system (6.34) , (6.35) can be solved by the block relaxation method of Sec. 4.3 and we observe that if u^n and λ^n are known (or estimated) in (6.35) the computation of p^n is an easy task since $|p^n|$ is solution of the single variable nonlinear equation

$$|p^n|^{s-1} + r|p^n| = |r\nabla u^n + \lambda^n| \tag{6.37}$$

which can be easily solved by various methods; once $|p^n|$ is known, we obtain p^n by solving a trivial linear equation (in $(L^s(\Omega))^N$). **201**

Solution of (6.26)**,** (6.27) **by ALG 2.**

We have to replace (6.33) by

$$\{p^0, \lambda^1\} \in H \times H' \tag{6.38}$$

and (6.34) by

$$\begin{cases} -r\Delta u^n = f + \nabla \cdot \lambda^n - r\nabla \cdot p^{n-1} \\ u^n|_\Gamma = 0. \end{cases} \tag{6.39}$$

Remark 6.1 still applies to (6.26), (6.27) and since G is linear we can take $0 < \rho_n = \rho < 2r$ if we are using ALG 2. For more details and comparisons with other methods see GLOWINSKI-MARROCCO [3], [6], [8].

REMARK 6.2. *ALG 1 and ALG 2 have also been successfully applied to the iterative solution of magneto-static problems (see GLOWINSKI-MARROCCO [7]). They have also been applied by GLOWINSKI-MARROCCO [3] to the solution of the subsonic flow problem described in Ch. 4. Sec. 3; in this last case using ALG 1 and ALG 2 we obtain easy variants of* (6.33)-(6.36) *and* (6.38), (6.39), (6.35), (6.36).

6.4 Application to the solution of mildly nonlinear systems

Let $\underset{\sim}{\mathrm{A}}$ be $N \times N$ symmetric, positive definite matrix $\underset{\sim}{\mathrm{D}}$ a diagonal, positive semi-definite matrix, and $\underset{\sim}{\mathrm{f}} \in \mathbb{R}^n$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a C^0 and non-decreasing functions (we can always suppose that $\phi(0) = 0$). Using the same notation as in Chapter 4, Sec. 2., we associate to $\underset{\sim}{\mathrm{v}} = \{v_1 v_2 \dots v_N\} \in \mathbb{R}^N$ the vector $\phi(\underset{\sim}{\mathrm{v}}) \in \mathbb{R}^N$ defined by

$$(\phi(\underset{\sim}{\mathrm{v}}))_i = \phi(v_i)\ \forall i = 1, \dots .N. \tag{6.40}$$

202 Then we consider the *nonlinear system*

$$\underset{\sim}{\mathrm{A}}\ \underset{\sim}{\mathrm{u}}\ \underset{\sim}{\mathrm{D}}\ \phi(\underset{\sim}{\mathrm{u}}) = \underset{\sim}{\mathrm{f}}\,. \tag{6.41}$$

In Chapter 4, Sec. 2.6, various methods for solving (6.41) have been given, but in this section we would like to show that (6.41) can also be solved by ALG1, ALG2, once a convenient augmented Lagrangian has been introduced.

REMARK 6.3. *The methods to be described later are easily generalized to the case where* $\underset{\sim}{\mathrm{A}}$ *is* not symmetric but still positive definite.

Let us define

$$\phi(t) = \int_0^t \phi(\tau)d\tau.$$

Since ϕ is C^0 and nondecreasing we have that ϕ is C^1 and *convex.* It follows then from the symmetry of A that solving (6.41) is equivalent to solving the minimization problem

$$\begin{cases} J(\underset{\sim}{\mathrm{u}}) \leq J(\underset{\sim}{\mathrm{v}}) \; \forall\, \underset{\sim}{\mathrm{v}} \in \mathbb{R}^N, u \in \mathbb{R}^N. \\ u \in \mathbb{R}^N \end{cases} \tag{6.42}$$

In (6.42) we have

$$J(\underset{\sim}{\mathrm{v}}) = \frac{1}{2}(\underset{\sim}{\mathrm{A}}\,\underset{\sim}{\mathrm{v}}, \underset{\sim}{\mathrm{v}}) + \sum_{i=1}^{N} d_i\phi(v_i) - (\underset{\sim}{\mathrm{f}}, \underset{\sim}{\mathrm{v}}), \tag{6.43}$$

where $(\cdot,\cdot)$ denotes the usual inner-product of $\mathbb{R}^N$ and $\|\cdot\|$ the corresponding norm and where

$$\underset{\sim}{\mathrm{D}} = \begin{pmatrix} d_1 \ddots & & 0 \\ & d_i \ddots & \\ 0 & & d_N \end{pmatrix}$$

From the above properties of $\underset{\sim}{\mathrm{A}}$, $\underset{\sim\sim}{\mathrm{D}}$ and Φ it follows from e. g. CEA [1], [2] that (6.41), (6.42) has a *unique solution.*

REMARK 6.4. *If fact* (6.41) *has a unique solution if* $\underset{\sim}{\mathrm{A}}$ *is* positive definite, possibly not symmetric, *the assumption on* ϕ *and* $\underset{\sim}{\mathrm{D}}$ *remaining the same.*

The problem (6.42) is a particular problem (P) corresponding to **203**

$$V = H = \mathbb{R}^N, B = I, \tag{6.44}$$

$$G(\underset{\sim}{v}) = \sum_{i=1}^{N} d_i \Phi(v_i) - (\underset{\sim}{f}, \underset{\sim}{v}), \tag{6.45}$$

$$F(\underset{\sim}{q}) = F_0(\underset{\sim}{q}) = \frac{1}{2}(\underset{\sim}{A}\,\underset{\sim}{q}\,\underset{\sim}{q}) \Rightarrow F_0'(\underset{\sim}{q}) = \underset{\sim}{A}\,\underset{\sim}{q}\,. \tag{6.46}$$

From these properties we can solve (6.41), (6.42) by using ALG 1 and ALG 2 (we observe that unlike in the above examples G is nonlinear).

REMARK 6.5. *Instead of using G and F defined by* (6.44), (6.45), *we can use*

$$G(v) = \sum_{i=1}^{N} d_i \Phi(v_i),$$

$$F(q) = \frac{1}{2}(\underset{\sim}{A}\,\underset{\sim}{q}, \underset{\sim}{q}) - (\underset{\sim}{f}, \underset{\sim}{q}).$$

The augmented Lagrangian to be associated with (6.44)–(6.46) is

$$\mathscr{L}_r(\underset{\sim}{v}, \underset{\sim}{q}, \underset{\sim}{u}) = \frac{1}{2}(\underset{\sim}{A}\,\underset{\sim}{q}\,\underset{\sim}{q}) + \sum_{i=1}^{N} d_i \Phi(v_i) - (\underset{\sim}{f}, \underset{\sim}{v}) + \frac{r}{2}\| \underset{\sim}{v} - \underset{\sim}{q} \|^2 + (\underset{\sim}{\mu}, \underset{\sim}{v} - \underset{\sim}{q}). \tag{6.47}$$

Since the constraint $\underset{\sim}{v}, \underset{\sim}{q} = 0$ is linear we know that $\mathscr{L}_r$ has a saddle-point over $\mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N$; actually this saddle-point is unique and is equal to $\{\underset{\sim}{u}, \underset{\sim}{u}, \underset{\sim}{A}\,\underset{\sim}{u}\}$.

Solution of (6.41) **by ALG 1**

It follows from (3.2)-(3.4), (6.47) that when applying ALG 1 to (6.41), (6.42) we obtain

$$\underset{\sim}{\lambda}^0 \in \mathbb{R}^N, \tag{6.48}$$

then for $n \geq 0$,

$$r\underset{\sim}{u}^n + \underset{\sim}{D}\,\phi(\underset{\sim}{u}^n) = \underset{\sim}{f} + r\underset{\sim}{p}^n - \underset{\sim}{\lambda}^n, \tag{6.49}$$

$$(r\,\underset{\sim}{\mathrm{I}} + \underset{\sim}{\mathrm{A}})\underset{\sim}{\mathrm{p}}^n = r\underset{\sim}{\mathrm{u}}^n + \underset{\sim}{\lambda}^n, \tag{6.50}$$

$$\underset{\sim}{\lambda}^{n+1} = \underset{\sim}{\lambda}^n + \rho_n(\underset{\sim}{\mathrm{u}}^n - \underset{\sim}{\mathrm{p}}^n). \tag{6.51}$$

The nonlinear system (6.49), (6.50) can be solved by the *block relax-* **204**
ation method of Sec. 4.3 and we observe that if $\underset{\sim}{\mathrm{p}}^n$ and $\underset{\sim}{\lambda}^n$ are known (or estimated) in (6.49) the computation of $\underset{\sim}{\mathrm{u}}^n$ is easy since it is reduced to the solution of *N independent, single variable nonlinear equations* of the following type

$$r\xi + d\phi(\xi) = b \text{ (with } d \geq 0). \tag{6.52}$$

Since $r > 0$ and ϕ is C^0 and non decreasing, (6.52) has a unique solution which can be computed by various standard method (see, e.g., HOUSEHOLDER [1], BRENT[1]). Similarly if $\underset{\sim}{\mathrm{u}}^n$ and $\underset{\sim}{\lambda}^n$ are known is (6.50) we obtain $\underset{\sim}{\mathrm{p}}^n$ by solving a linear system whose matrix is r independent of n it is very convenient to prefactorize $r\,\underset{\sim}{\mathrm{I}} + \underset{\sim}{\mathrm{A}}$ (by Cholesky or Gauss methods).

Solution of (6.1) **by ALG 2.**

We have to replace (6.48) by

$$\left\{\underset{\sim}{\overset{0}{\mathrm{p}}}, \underset{\sim}{\lambda}^1\right\} \in \mathbb{R}^N \times \mathbb{R}^N \tag{6.53}$$

and (6.49) by

$$r\,\underset{\sim}{\overset{n}{\mathrm{u}}} + \underset{\sim}{\mathrm{D}}\,\phi(\underset{\sim}{\overset{n}{\mathrm{u}}}) = \underset{\sim}{\mathrm{f}} + r\,\underset{\sim}{\overset{n-1}{\mathrm{p}}} - \underset{\sim}{\lambda}^n. \tag{6.54}$$

It follows from Theorem 5.2 that we have convergence of (6.53), (6.54), (6.50), (6.51) if $0 < \rho_n = \rho < \dfrac{1+\sqrt{5}}{2} r$.

REMARK 6.6. *Suppose that $\rho_n = \rho = r$ in ALG 2; we have then*

$$\begin{cases} r\,\underset{\sim}{\overset{n}{\mathrm{u}}} + \underset{\sim}{\mathrm{D}}\,\phi(\underset{\sim}{\overset{n}{\mathrm{u}}}) = \underset{\sim}{\mathrm{f}} + r\,\underset{\sim}{\overset{n-1}{\mathrm{p}}} - \underset{\sim}{\lambda}^n, \\ r\,\underset{\sim}{\overset{n}{\mathrm{p}}} + \underset{\sim}{\mathrm{A}}\,\underset{\sim}{\overset{n}{\mathrm{p}}} = r\,\underset{\sim}{\overset{n}{\mathrm{u}}} + \underset{\sim}{\lambda}^n, \\ \underset{\sim}{\lambda}^{n+1} = \underset{\sim}{\lambda}^n + r(\underset{\sim}{\overset{n}{\mathrm{u}}} - \underset{\sim}{\overset{n}{\mathrm{p}}}). \end{cases} \tag{6.55}$$

It follows from (6.55) that

$$\underset{\sim}{\lambda}^{n+1} = \underset{\sim}{A}\,\underset{\sim}{p}^{n}, \tag{6.56}$$

205 Then from (6.55), (6.56) we obtain

$$r\,\underset{\sim}{u}^{n} + \underset{\sim}{D}\,\phi(\underset{\sim}{u}^{n}) + \underset{\sim}{A}\,\underset{\sim}{p}^{n-1} = \underset{\sim}{f} + r\,\underset{\sim}{p}^{n-1}, \tag{6.57}$$

$$r\,\underset{\sim}{p}^{n} + \underset{\sim}{A}\,\underset{\sim}{p}^{n} + \underset{\sim}{D}\,\phi(\underset{\sim}{u}^{n}) = \underset{\sim}{f} + r\,\underset{\sim}{p}^{n-1}. \tag{6.58}$$

Therefore, if $\rho_n = \rho = r$, ALG 2 reduces (with different notation) to the *Alternating Direction method* described on Ch. 4, Sec. 2.6.6.

REMARK 6.7. *From the numerical experiment done in CHAN-GLOWINSKI [1], ALG 1 combined with the* block relaxation *method of Sec. 4.3 is more robust that ALG 2; it is the case if, for instance, we solve a finite element (or finite difference) approximation of the mildly nonlinear elliptic problem*

$$\begin{cases} -\Delta u + u|u|^{s-2} = f \text{ on } \Omega, \\ u|_{\Gamma} = 0. \end{cases} \tag{6.59}$$

with $1 < s < 2$.

In CHAN - GLOWINSKI, loc. cit., we can find various numerical results and also comparisons with other methods.

6.5 Solution of Elliptic Variational Inequalities on intersections of convex sets

6.5.1 Formulation of the problem

Let V be a real Hilbert space and $a : V \times V \to \mathbb{R}$ be a bilinear form, continuous, symmetric and V- elliptic. Let K be a closed, convex, non-empty subset of V such that

$$K = \cap_{i=1}^{N} K_i, \tag{6.60}$$

where, $\forall i = 1, \ldots, N, K_i$ is a *closed convex subset* of V. We consider then the EVI problem

$$\begin{cases} a(u, v-u) \geq L(v-u) \, \forall v \in K, \\ u \in K \end{cases} \tag{6.61}$$

where $L : V \to \mathbb{R}$ is linear and continuous. Since $a(\cdot, \cdot)$ is symmetric we know from Chap. 1 that the unique solution of (6.61) is also the solution of

$$\begin{cases} J(u \geq (v) \, \forall v \in K, \\ u \in K, \end{cases} \tag{6.62}$$

where **206**

$$J(v) \frac{1}{2} a(v, v) - L(v). \tag{6.63}$$

6.5.2 Decomposition of (6.61), (6.62)

let us define (with $q = \{q_1, \ldots, q_N\}$)

$$W = \{(v, q) \in V \times V^N, v - q_i = 0 \, \forall i = 1 \cdots N\} \tag{6.64}$$

and

$$\mathscr{K} = \{(v, q) \in W, q_i \in K_i \, \forall i = 1, \ldots, N\}. \tag{6.65}$$

It is clear that (6.62) is equivalent to

$$\min_{(v,q) \in \mathscr{K}} j(v, q) \tag{6.66}$$

where

$$j(v, q) = \frac{1}{2N} \sum_{i=1}^{N} a(qi, q_i) - L(v). \tag{6.67}$$

REMARK 6.8. *We have to observe that many other decompositions are possible, as, for instance,*

$$W = \{(v, q) \in V \times V^N, v - q_1 = 0, q_{i+1} - q_i = 0 \, \forall i = 1, \ldots, N-1\}$$

with j and $\mathscr{K}$ still defined by (6.67), (6.65). We can also use

$$W = \{(v, q) \in V \times V^{N-1}, v - q_i = 0 \, \forall i = 1, \dots N - 1\}$$

with

$$\mathscr{K} = \{(v, q) \in W, v \in K_1, q_i \in K_{i+1} \, \forall i = 1, \dots N - 1\}$$

and

$$j(v, q) = \frac{1}{2N} a(v, v) - L(v) + \frac{1}{2N} \sum_{i=1}^{N-1} a(q_i, q_i).$$

207 We suppose that in the sequel we use the decomposition defined by (6.64)–(6.67); then (6.66) is particular problem (P) corresponding to

$$H = V^N, Bv = \{v, \dots, v\}, \tag{6.68}$$

$$G(v) = -L(v), \tag{6.69}$$

$$F_0 = \frac{1}{2N} \sum_{i=1}^{N} a(q_i, q_i), \tag{6.70}$$

$$F_1(q) = \sum_{i=1}^{N} I_{K_i}(q_i) \tag{6.71}$$

with

I_{K_i}: indicator function of K_i.

It is easily shown that from the properties of B, G, F, we can apply ALG 1 and ALG 2 to solve (6.62), via (6.66), provided that the following augmented Lagrangian

$$\mathscr{L}_r(v, q, \mu) = F(q) + G(v) + \frac{r}{2N} \sum_{i=1}^{N} a(v - q_i, v - q_i) + \frac{1}{N} \sum_{i=1}^{N} (\mu_i, v - q_i) \tag{6.72}$$

has a saddle-point over $V \times V^N \times V^N$. Such a saddle-point exists if H is *finite dimensional*, since the constraints $v - q_i = 0$ are *linear*.

6.5.3 Solution of (6.62) by ALG 1

It follows from (3.2) - (3.4), (6.72) that when applying ALG 1 to (6.62) we obtain

$$\lambda^0 \in V^N \text{given}, \tag{6.73}$$

then for $n \geq 0$

$$\begin{cases} ra(u^n, v) = ra(\frac{1}{N}\sum_{i=1}^{N} p_i^n, v) - (\frac{1}{N}\sum_{i=1}^{N} \lambda_i^n, v) + L(v) \; \forall v \in V, \\ u^n \in V, \end{cases} \tag{6.74}$$

$$\begin{cases} (1+r)a(p_i^n, q_i - p_i^n) \geq ra(u^n, q_i - p_i^n) + (\lambda_i^n, q_i - p_i^n \; \forall q_i \in K_i, \\ p_i^n \in K_i \end{cases} \tag{6.75}$$

for $i = 1, 2, \ldots N$; **208**

$$\lambda_i^{n+1} = \lambda_i^n + \rho_n(u^n - p_i^n) \tag{6.76}$$

$i = 1, \ldots N$.

The system (6.74), (6.72) is for λ^n given a system of coupled *EVI*s, a very convenient method to solve it is the *block overrelaxation method with projection* described in CEA-GLOWINSK [1] and in CEA [2]. This method will reduce the solution of (6.62) to a sequence of EVIs K_i, $i = 1, \ldots N$.

6.5.4 Solution of (6.62) by ALG 2

It follows from (3.7)-(3.10), (6.72) that to solve (6.62) by ALG 2 we have to use the variant of (6.73)-(6.76) obtained by replacing (6.73), (6.74) by

$$\left\{p^0, \lambda^1\right\} \in V^N \times V^N \text{ given}, \tag{6.77}$$

$$\begin{cases} ra(u^n, v) = ra\left(\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} p_i^{n-1}, v\right) - \left(\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \lambda_i^N, v\right) + L(v) \; \forall v \in V, \\ u^n \in V. \end{cases} \tag{6.78}$$

REMARK 6.9. *The two above algorithms are well- suited to the use of multiprocessor computers, since many operations may be done in parallel; this is particularly clear with algorithm* (6.77), (6.78), (6.75), (6.76).

REMARK 6.10. *Using different augmented Lagrangians, other than* $\mathscr{L}_r$ *defined by* (6.72), *we can solve* (6.62) *by algorithms better suited to* sequential computing *than to parallel computing. We leave to the reader, as exercises, the task of describing such algorithms.*

REMARK 6.11. *The two algorithms described above can be extended to EVIs where* $a(\cdot,\cdot)$ *is not symmetric. Moreover they have the advantage of reducing the solution of* (6.62) *to the solution of a sequence of simpler EVI s of the same type, to be solved over* K_i, $i = 1, \ldots N$, *instead of* K.

7 General Comments

209 As mentioned several times the methods described in this chapter may be extended to variational problem *which are not equivalent* to optimization problem. These methods have been applied by BEGIS-GLOWINSKI [1] to the solution of 4^{th} order nonlinear problems in Fluid Mechanics (see also BEGIS [1]).

From a conceptual point of view they are related to various methods, described in BENSOUSSAN-LIONS -TEMAM [1], and using *decomposition-coordination* principles.

From an historical point of view, the use of augmented Lagrangian for solving -via ALG 1 and ALG 2 -nonlinear variational problems of type (P) (see (1.1)) seems to be due to GLOWINSKI - MARROCCO [237], [5], [6]. For more details and other applications see GABAY-MERCIER [1], FORTIN-GLOWINSKI [1], [2], GLOWINSKI-MARROCCO, loc, cit., etc.. . . .

To conclude this chapter we have to mention that using some results due to OPIAL [1] we have in fact in Theorem 4.1, 5.1 (resp. . 4.2, 5.2) the *weak convergence*(resp. . the *convergence*) of the whole sequence $\{\lambda^n\}_n$ to a λ^* such that $\{u, p, \lambda^*\}$ is a saddle-point of $\mathscr{L}$ (and $\mathscr{L}_r$) over $V \times H \times H$. We refer to GLOWINSKI-LIONS -TREMOLIERES [3, Appendix 2] for a proof of the above results in a more general context.

Chapter 6

On the Computation of Transonic Flows

1 Introduction

We have considered in Chapter 4, Section 3, the non-linear elliptic equation describing the *subsonic flows* of an *inviscid compressible fluid.* **210**

In this chapter, following closely GLOWINSKI - PIRONNEAU [1], we would like to give some brief indications on the computation of transonic flows for similar fluids. Given the importance and the complexity of the problem to be described in a moment, we would like to point out that the following considerations are just an introduction to the subject and that many methods, using very different approaches, exist in the specialized literature (see the following references). Moreover, we would like to mention that from a mathematical point of view, the methods to be described in the following sections are widely heuristical. A large number of bibliographical references are given in the sequel.

2 Generalities

The theoretical and numerical studies of *transonic flows for inviscid fluids* have always been very important questions. But these problems have

become even more important in recent years in relation to the design and development of *large subsonic economical aircrafts*.

From the theoretical point of view a lot of open questions still remains, with their counterparts in the numerical methodology. The difficulties are quite considerable for the following reasons:

(1) The problems are *nonlinear*;

(2) *Shocks* may exist in the flow;

(3) One has to include an *entropy conditions*, in one way or another, to avoid non-physical solutions.

211 From the *theoretical* point of view we have to mention the work of BERS [1], C. MORAVETZ [1]. At the present moment the more commonly used numerical methods have originated from MURMAN-COLE [1] and we shall mention BAUER-GARABEDIAN-KORN [1], BAUER-GARABEDIAN - KORN JAMESON [1], JAMESON [1], [2], [3], [4] and the bibliographies therein (see also HEWITT-ILLINGWORTH and co-editors [1]).

These above numerical methods use the key idea of Murman and Cole which consists in the use of a *finite difference scheme, centered* in the *subsonic* part of the flow, *backward* (in the direction of the flow) in the *supersonic part*. The switching between these two schemes is automatically done via *a truncation operator* only active in the *supersonic* part of the flow (see JAMESON, loc. cit., for more details). A *relaxation* method is then used to solve the resulting nonlinear system (actually, *over-relaxation* is used in the subsonic part of the flow, *under-relaxation* in the supersonic part).

We shall describe a different approach -very convenient for *nozzles, and flows subsonic at infinity around airfoils* - in which the transonic flow problem is formulated as a *nonlinear least square problem*. This last problem is then viewed as an *optimal control problem* which is approximated by a *finite element* method. Since the entropy condition is formulated by a *linear inequality constraint*, a convenient method to handle it is to use *penalty* and/or *duality* methods (see CEA [1], [2]), using an *augmented Lagrangian* if penalty and duality are combined.

Then the approximate problem is solved by *iterations* of *conjugate gradient* type. Our approach is strongly motivated by the two following points of view and the corresponding methodologies:

(1) *Optimal control of distributed parameter system* (see LIONS [4], CEA [1], [2]),

(2) *Variational inequalities and their numerical solution* (see GLOWINSKI-LION-TREMOLIERES [1], [2], [3] and Chapters 1 to 5 of these notes).

3 Mathematical Model For The Transonic Flow Problem

3.1 Basic assumptions and generalities

We assume that the fluid under consideration is *inviscid* and *compressible* and that the flow of such a fluid is *is entropic* and *irrotational* (i.e. *potential)*. These assumptions are not true in general, since through a shock there is a *variation of entropy* and an *irrotational flow becomes rotational*; therefore the validity of the model to follow is assumed to be correct only in the case of a *"weak shock"*. **212**

In the case of a flow past a *sharp airfoil* we shall suppose that there is *no wake behind the trailing edge*.

3.2 Equations of the flow

Let Ω be the domain of the flow and Γ its boundary; then the flow is modelled by

$$-\nabla\left(\left(1-\frac{|\nabla\Phi|^2}{\frac{\gamma+1}{\gamma-1}C_*^2}\right)^{\frac{1}{\gamma-1}}\nabla\Phi\right)=0 \text{ in } \Omega, \tag{3.1}$$

where

- Φ is the *flow potential*, $\nabla\Phi$ the *flow velocity*,

- C_* is the *critical velocity,*

- γ is the ratio of specific heats (γ = 1. 4 for air).

We have to add to(3.1)

- Boundary conditions (of Dirichlet and/or Neumann type, for example);

- *Kutta-Joukowsky* condition in the case of the flow around a lifting body (see LANDAU - LIFCHITZ [1, Sec. 46]); some indications are also given in Sec. 5.1, Remark 5. 1.

- An *entropy condition* in order to eliminate the non-physical solutions of (3.1); this point will be discussed in Sec. 3.3

REMARK 3.1. *It can happen that on some part of the boundary,* Φ *and* $\dfrac{\partial \Phi}{\partial n}$ *have to be given simultaneously to ensure uniqueness; it is the case, for instance, for the divergent nozzle of Figure 3.1 if the velocity at the entrance is supersonic. Typical boundary conditions are* Φ *given on* Γ_1, Γ_3 *and* $\dfrac{\partial \Phi}{\partial n}$ *given on* Γ_4, Γ_1, Γ_2*; if the flow at the entrance (i.e.* Γ_1*) is subsonic we require fewer boundary conditions.*

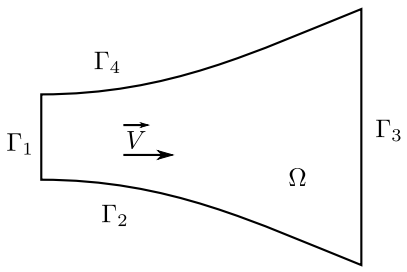

Figure 3.1:

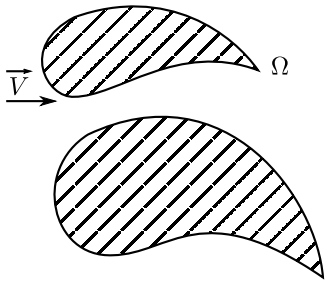

Figure 3.2:

REMARK 3.2. *In the case of the flow around a multipiece airfoil, (like* **213** *in Fig. 3.2 each piece requires a Kutta-Joukowsky condition.*

Exercise 3.1. *Verify that* (3.1) *is elliptic if* $|\nabla\phi| < C_*$ *(subsonic zone), hyperbolic if* $|\nabla\phi| > C_*$ *(supersonic zone).*

3.3 Formulation of the entropy condition

It follows from LANDAU-LIFCHITZ [1, Ch. 9] that the *entropy condition* can be formulated as follows

$$\begin{cases} \text{In the direction of the flow, once cannot have a subsonic-} \\ \text{supersonic transition through a shock.} \end{cases} \tag{3.2}$$

For *one dimensional* flow, (3.2) implies

$$\frac{d^2\Phi}{dx^2} < +\infty, \tag{3.3}$$

i. e. $\dfrac{d^2\Phi}{dx^2}$ is a *measure bounded from above; weak*(and more precise) formulations of (3.3) are : *There exists a constant M, such that either*

$$-\int_\Omega \frac{d\Phi}{dx}\frac{d\phi}{dx}dx \le M\int_\Omega \phi dx \quad \forall\phi \in \mathscr{D}_+(\Omega) \tag{3.4}$$

or

$$\int_\Omega \phi\frac{d^2\phi}{dx^2}dx \le M\int_\Omega dx \quad \forall\phi \in \mathscr{D}_+(\Omega) \tag{3.5}$$

where

$$\mathscr{D}_+(\Omega) = \{\phi \in \mathscr{D}(\Omega), \phi \geq 0\}. \tag{3.6}$$

214 In the case of a *two or three dimensional flow*, we shall suppose that (3.2) can be formulated as

$$\triangle\Phi < +\infty \tag{3.7}$$

or in a *weak from* either by

$$-\int_\Omega \nabla\Phi \cdot \nabla\phi dx \leq M \int_\Omega \phi dx \, \forall\phi \in \mathscr{D}_+(\Omega) \tag{3.8}$$

or by

$$\int_\Omega \Phi\triangle\phi dx \leq M \int_\Omega \phi dx \quad \forall\phi \in \mathscr{D}_+(\Omega). \tag{3.9}$$

The numerical results that we have obtained for *two-dimensional* flows, using discrete analogs of (3.7), seem to justify the above formulations of the entropy condition.

4 Reduction to an Optimal Control Problem

If we suppose that the density on the fluid is on if $\underset{\sim}{u} = \nabla\Phi = 0$, then the coefficient of $\nabla\Phi$ in (3.1) appears as the *density* of the fluid. We shall use the notation

$$\rho(\phi) = \left(1 - \frac{|\nabla\phi|^2}{\frac{\gamma+1}{\gamma-1}} C_*^2\right)^{\frac{1}{\gamma-1}} \tag{6.1}$$

The idea of the method to follows, is to *decouple* the density and the potential Φ. To do so we introduce a new potential ξ -the control potential -and try to recouple ξ and Φ by minimizing some *cost function* of *least square type*. We may use for instance the following formulation (for the formulation see BRISTEAU [2], BRISTEAU - GLOWINSKI - PERIAUX - PERRIER - PIRONNEAU [1], BRISTEAU - GLOWINSKI - PERIAUX - PERRIER - PIRONNEAU - POIRIER [1])

$$\min_\xi \int_\Omega \rho^\alpha(\xi)|\nabla(\Phi - \xi)|^2 dx, \xi \in X \tag{4.2}$$

where, in (4.2) Φ is a function of ξ via the *state equation*

$$\begin{cases} -\nabla \cdot (\rho(\xi)\nabla\Phi) = 0 \text{ over } \Omega \\ + \text{ boundary conditions for } \Phi \text{ on } \Gamma. \end{cases} \tag{4.3}$$

In (4.2) the parameter α is either 0 of 1 and X is a *convex* set "conve- **215**
niently" chosen. Since $\rho(\phi(x)) = 0$ iff $|\nabla\phi(x)| = \left(\frac{\gamma+1}{\gamma-1}\right)^{1/2} C_*$, and that for *air* we have $\gamma = 1.4$ which implies $\left(\frac{\gamma+1}{\gamma-1}\right)^{1/2} = \sqrt{6} \simeq 2.45$, it appears that in the transonic range (say $|\vec{v}| = |\nabla\phi| \leq 1.5C_*$) we have

$$0 < \delta < \rho(\phi(x)) \leq 1 a.e. \text{ on } \Omega. \tag{4.4}$$

it follows from (4.4) that (4.3) is an *elliptic problem* for appropriate boundary conditions. In the case of flows around lifting airfoils, *Kutta-Joukowsky conditions* are also required in order to obtain, with the other boundary conditions, a physical solution of problem (4.3) (modulo a constant if one has only Neumann conditions on the boundary).

REMARK 4.1. *If in the original problem,* Φ *and* $\dfrac{\partial\Phi}{\partial n}$ *have to be simultaneously prescribed on some part of* Γ*, the previous approach with two-potentials is very convenient since the boundary conditions can be split between* Φ *and* ξ*. However, if one wishes to use the same boundary conditions for* Φ *and* ξ *it is always possible to take into account the extra boundary conditions (assumed to be of Dirichlet type) by adding to the cost function* (4.2) *a quantity proportional to either* $\int_{\Gamma_d} |\Phi - \Phi_d|^2 d\Gamma$*, or* $\int_{\Gamma_d} |\Phi - \Phi_d|^2$ *(or to a linear combination of both), where* Γ_d *is the part of* Γ *where one requires* $\Phi|_{\Gamma_d} = \Phi_d$*. A similar idea is used in BEGIS-GLOWINSKI [2] to solve some free boundary problem.*

REMARK 4.2. *To state the entropy* conditions (3.7) *(or its weak formulations* (3.8)*, (3.9)) we have the choice between* Φ *and* ξ *(actually we can also use these two potentials simultaneously). If one uses* ξ *(resp.* Φ*) we have a* constraint on the control *(resp.* constraint on the state*).*

REMARK 4.3. *We observe that the class of flows we are considering, is physically such that*

$$\|\vec{v}\|_\infty = \|\nabla\Phi\|_\infty < +\infty.$$

216 *It follows from this remark that the convex set occurring in* (4.2) *will be taken as a convex subset of* $W^{1,\infty}(\Omega)$*. We observe also that to stay in the transonic range it may be convenient to introduce following constraints (if* $\gamma = 1.4$*):*

$$|\nabla\xi| \leq v_M < \sqrt{6}C_* \tag{4.5}$$

or

$$|\nabla\Phi| \leq v_M < \sqrt{6}C_*. \tag{4.6}$$

Actually, the computations we have done proved that for a physically well - posed travsonic problem it is not necessary to introduce (4.5) *or* (4.6).

REMARK 4.4. *If the transonic problem has a solution and if X is "large enough" the control problem will have a solution such that the* cost function will be equal to zero*; this last property will give us (for the approximate problem) indications to check the quality of the computed solution.*

5 Approximation

We assume that $\Omega \subset \mathbb{R}^2$.

5.1 Generalities

The above control problem will be approximated by a *finite element method*, since compared to *finite difference* methods it give us the possibility of handling problems posed on rather complicated geometry. Moreover the *variational foundations* of finite element formulations are very appropriate to the problem under consideration. It will be in particular easy to approximate the weak formulations of the entropy condition (3.7).

If Ω is *unbounded* it will be replaced by a *bounded domain* - still denoted by Ω- as large as possible. To approximate the above continuous problems we introduce a standard *triangulation* $\mathscr{C}_h$ of Ω (we can also use *quadrilateral finite elements* defined over a "quadrangulation"

of Ω). Then the functions ξ and Φ are approximated by piecewise polynomial functions belonging to the following subspace V_h of $H^1(\Omega)$.

$$V_h = \left\{\Phi_h \in C^0(\overline{\Omega}), \Phi_h|_T \in P_k \forall T \in \mathscr{C}_h\right\}, \tag{5.1}$$

with P_k = the space of polynomials of degree $\leq k$.

REMARK 5.1. *In the case of a* lifting body*, to take into account the* Kutta - Joukowsky condition*, one usually introduces (see Fig. 5.1) an arc γ between the leading edge of the profile and the external boundary. This arc γ supports a* constant jump (a priori unknown) *of Φ (and ξ) and this jump has to be adjusted in such a way that $\frac{\partial \Phi}{\partial n}$ (and $\frac{\partial \xi}{\partial n}$) is "continuous" when crossing γ. Since Φ is discontinuous along γ, we cannot work anymore with $H^1(\Omega)$, but introducing $\dot{\Omega} = \overset{\circ}{\overline{\Omega}-\gamma}$ one can use $H^1(\dot{\Omega})$ and define V_h over a triangulation of Ω.* **217**

In Sec. 7, results of computations for such airfoils are given; however for the sake of simplicity the numerical treatment of the Kutta-Joukowsky condition will not be discussed here and we shall assume in the sequel that one works directly over Ω.

REMARK 5.2. *If Φ and ξ have to satisfy only* Neumann boundary conditions*, the subspace of V_h to be used will be V_h itself. On the contrary if Φ and /or ξ have satisfy* Dirichlet boundary conditions *somewhere over Γ then we shall have to use subspaces of V_h* strictly *included in V_h.*

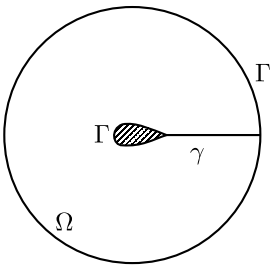

Figure 5.1:

REMARK 5.3. *We have tacitly assumed that* Ω *is a polygonal domain of* $\mathbb{R}^2$ *or has been approximated by such a domain. However in the case of a curved boundary, it is always possible to use (at some extra computational cost)* curved finite elements *(See, for instance CIARLET-RAVIART [1], STRANG - FIX [1, Ch. 3], CIARLET [1], [3]).*

REMARK 5.4. *It follows from* (5.1) *that we are using* C^0 - conforming *finite elements. Since the regularity of the solution is limited it seems that it would be unrealistic to use* $k \geq 3$. *Therefore only* Lagrange elements *will be considered. One may also use that non* C^0*- conforming element of Figure 5.2 in which, with* $\phi_{h|T} \in P_1$*, one only requires the continuity of* ϕ_h *at the mid-point of each side of the* $T \in \mathscr{C}_h$*. The number of unknowns, when using this element is much in higher than when using* (5.1) *with* $k = 1$.

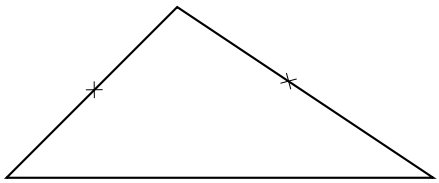

Figure 5.2:

5.2 Approximation of the state equation and of the cost function

218 To simplify the presentation we shall assume that we only have *Neumann boundary conditions* i.e.

$$\frac{\partial \Phi}{\partial n} = g \text{ on } \Gamma \tag{5.2}$$

(it is the case for the very important application of flows around airfoils, *subsonic at infinity*). It follows from the above sections that we can take the *same boundary conditions* for ξ and Φ. We shall also assume that if

$g(x) \neq 0$, $x \in \Gamma$, then the corresponding value of ρ is known and that

$$\int_\Gamma \rho g d\Gamma = 0. \tag{5.3}$$

Then the *state equation* (4.3) has the following *variational formulation*:

$$\begin{cases} \int_\Omega \rho(\xi)\nabla\Phi \cdot \nabla\phi dx = \int_\Omega \rho g \phi d\Gamma \ \forall \in H^1(\Omega), \\ \Phi \in H^1(\Omega) \end{cases} \tag{5.4}$$

which is approximated by

$$\begin{cases} \int_\Omega \rho(\xi_n)\nabla\Phi_h \cdot \nabla\phi_h dx = \int_\Gamma (\rho g)_h \phi_h d\Gamma \ \forall \phi_h \in V_h, \\ \Phi_h \in V_h, \end{cases} \tag{5.5}$$

where $(\rho g)_h$ is a convenient approximation of ρg over Γ. Since Φ and Φ_h are only defined modulo an additive constant, we shall prescribe the value of Φ and Φ_h (and ξ and ξ_h) at some point of Γ. The cost function **219**
in (4.2) is approximated by

$$\int_\Omega \rho^\alpha(\xi_h)|\nabla(\Phi_h - \xi_h)|^2 dx, \tag{5.6}$$

denoted $J_h(\xi_h)$ in the following.

REMARK 5.5. *If one uses the piecewise linear approximation (i.e. $k = 1$), then the integrals occurring in* (5.5), (5.6) *are easy to compute since $\nabla\xi_h$ being* piecewise constant, *we have a similar property for $\rho(\xi_h)$. If $k = 2$ a numerical integration procedure has to be used in* (5.5), *and also in* (5.6) *if $\alpha = 1$.*

5.3 Approximation of the entropy condition

To avoid non physical shocks (i.e. shocks for which the entropy condition is not satisfied) we have several possibilities; we shall describe two of them (for other approaches see GLOWINSKI - PIRONNEAU [1]). We still assume that we only have Neumann boundary conditions like (5.2).

5.3.1 A regularization method

We use the notation of the continuous problem; the idea is to add to the cost functional in (4.2) the following functional, with $\epsilon > 0$, either

$$\epsilon \int_\Omega |(\Delta\Phi)^+|^2 dx, \tag{5.7}$$

or

$$\epsilon \int_\Omega |(\Delta\xi)^+|^2 dx \tag{5.8}$$

(or a linear combination of both). In (5.7), (5.8), ϵ is a "small" parameter and

$$(\Delta\phi)^+ = \sup(0, \Delta\phi). \tag{5.9}$$

We can make this approach more sophisticated by using, instead of (5.7), (5.8), regularization functionals like

$$\int_\Omega \epsilon(x)|(\Delta\Phi)^+|^2 dx, (resp. \int_\Omega \epsilon(x)|(\Delta\xi)^+|^2 dx), \tag{5.10}$$

220 where $\epsilon(x)$ is a "small" non-negative *weight function*, possibly equal to zero over some part of Ω. To use the above methodology for the approximate problem it is necessary to have an approximation of $\Delta\Phi(resp.\Delta\xi)$. We shall use and approximation suggested by *mixed finite element methods for the biharmonic education* (see GLOWINSKI [6], CIARLET-RAVIART [2], GLOWINSKI- PIRONNEAU [2], BRIZZI - RAVIART [1]).

Let us assume that ψ is sufficiently smooth, then from Green's formula we have

$$\int_\Omega \Delta\psi\phi dx = \int_\Gamma \frac{\partial\psi}{\partial n}\phi d\Gamma - \int_\Omega \nabla\psi \cdot \nabla\phi dx \ \forall\phi \in H^1(\Omega). \tag{5.11}$$

Using this idea we shall define an approximation $\Delta_h\Phi_h$ of $\Delta\Phi$ as follows:

$$\begin{cases} \int_\Omega \Delta_h\Phi_h\phi_h dx = \int_\Gamma g_h\phi_h d\Gamma - \int_\Omega \nabla\phi_h \cdot \nabla\phi_h dx \ \forall\phi_h \in V_h, \\ \Delta_h\Phi_h \in V_h. \end{cases} \tag{5.12}$$

We use the same method to define $\Delta_h\xi_h$. In (5.12), g_h is an approximation of the function g of (5.2).

REMARK 5.6. *If we also have Dirichlet boundary conditions over some part of* Γ*, the same method can be used with some slight complications.*

REMARK 5.7. *To obtain* $\Delta_h\Phi_h$ *from* (5.12) *we have to solve a* linear system *whose matrix is* symmetric, positive definite, spartse, *but* not diagonal *(this matrix is an approximation of the operator I). If* $k = 1$ *one can approximate* $\int_\Omega \Delta_h\Phi_h\phi_h dx$*, using the* two- dimensional trapezoidal, numerical integration method*. Doing so we obtain* $\Delta_h\Phi_h$ *by solving a linear system with a* diagonal *matrix.*

If $k = 2$ the above regularization method is technically more complicated to use.

Once Δ has been approximated we add to the cost function J_h (cf. (5.6)) the functional

$$\int_\Omega \epsilon_h(x)|(\Delta_h\Phi_h)|^{+^2} dx(resp. \int_\Omega \epsilon_h(x)|(\Delta_h\xi_h)|^{+^2} dx) \tag{5.13}$$

with ϵ_h "small" and ≥ 0. In fact we use approximations of (5.13) obtained via a *numerical integration* procedure.

We have to mention that the optimal choice for ϵ_h is still an open **221**
question. Numerical results obtained with piecewise linear approximations ($k = 1$) are given in Sec. 7.

5.3.2 A method using 3.7

Let us describe first this method for the *continuous problem*:

We suppose that the entropy condition can be formulated by (3.7). Let $M(x)$ be a sufficiently smooth upper bound of $\Delta\Phi$ (M is estimated or guessed). Replacing (3.7) by

$$\Delta\Phi \leq M(x), \tag{5.14}$$

a *weak formulation* of (5.14) is

$$-\int_\Omega \nabla\Phi \cdot \nabla\phi dx \leq \int_\Omega M(x)dx \ \forall\phi \in \mathscr{D}_+(\Omega). \tag{5.15}$$

Instead of using M it is very convenient to introduce the solution (defined up to an arbitrary constant if we only have Neumann boundary conditions) of

$$\begin{cases} \Delta\Phi_0 & = M \text{ over } \Omega, \\ \frac{\partial\Phi_0}{\partial n} & = g \text{ over } \Gamma \end{cases} \tag{5.16}$$

then (5.16) has the following variational formulation

$$\begin{cases} \int_\Omega \nabla\Phi_0 \cdot \nabla\phi dx = -\int_\Omega M(x)\phi dx + \int_\Gamma g\phi d\Gamma \ \forall\phi \in H^1(\Omega), \\ \Phi_0 \in H^1(\omega). \end{cases} \tag{5.17}$$

It follows from (5.17) that (5.15) can also be written

$$-\int_\Omega \nabla(\Phi - \Phi_0) \cdot \nabla\phi dx \leq 0 \ \forall\phi \in \mathscr{D}_+(\Omega). \tag{5.18}$$

We observe that

$$\frac{\partial}{\partial n}(\Phi - \Phi_0) = 0. \tag{5.19}$$

222 Concerning the discrete problem, the obvious strategy seems to be the following: First we approximate Φ_0 by Φ_{oh}, a solution of

$$\begin{cases} \int_\Omega \nabla_{oh} \cdot \nabla\phi_h dx = -\int_\Omega M_h(x)\phi_h dx + \int_\Gamma g_h\phi_h d\Gamma \ \forall\phi_h \in V_h, \\ \phi_{oh} \in V_h, \end{cases} \tag{5.20}$$

where M_h is a convenient approximation of M. Then we approximate $H_0^1(\Omega)$ (and $\mathscr{D}(\Omega)$) by

$$V_{oh} = \{\phi_h \in V_h, \phi_h|_\Gamma = 0\},$$

and $\mathscr{D}_+(\Omega)$ by

$$V_{oh}^+ = \{\phi_h \in V_{oh}, \phi_h \geq 0 \text{ on } \Omega\}. \tag{5.21}$$

Finally we approximate (5.18) by

$$-\int_\Omega \nabla(\Phi_h - \Phi_{oh}) \cdot \nabla\phi dx \leq 0 \ \forall\phi_h \in V_{oh}^+. \tag{5.22}$$

In fact the above "obvious" strategy has to be modified for the following reasons:

(1) If $k = 1$, V_{oh}^+ can be generated by the canonical basis functions of V_{oh}. But it is not the case for $k = 2$.

(2) Computations using (5.22) a done with $k = 1$ have shown that the approximation of the solution is not good close to the points at which *sonic lines* touch Γ.

To overcome these difficulties we may proceed as follows: *If* $k = 1$, let $\sum_h$ be the set of the vertices of $\mathscr{C}_h$, numbered from 1 to N_h, where $N_h = \dim(V_h)$. Let β_h be the canonical basis of V_h, i.e.

$$\beta_h = \{w_i\}_{i=1}^{N_h} \tag{5.23}$$

with

$$w_i \in V_h, w_i(P_j) = \delta_{ij} \ \forall P_i \in \sum_h . \tag{5.24}$$

We observe that $w_i \geq 0$ over Ω, $\forall i$, and that the positive cone V_h^+ of V_h is **223**
generated by β_h. Then instead of using (5.22) to formulate the discrete entropy condition, one takes

$$-\int_\Omega \nabla(\Phi_h - \Phi_{oh}) \cdot \nabla\phi_h dx \leq 0 \ \forall \phi_h \in V_h^+ \tag{5.25}$$

which is equivalent to the set of the N_h following *linear inequality constraints*

$$-\int_\Omega \nabla(\Phi_h - \Phi_{oh}) \cdot \nabla w_i dx \leq 0 \ \forall i = 1, \ldots N_h. \tag{5.26}$$

If ξ_h is used, instead of (5.26) we have

$$-\int_\Omega \nabla(\xi_h - \Phi_{oh}) \cdot \nabla w_i dx \leq 0 \ \forall i = 1, \ldots N_h. \tag{5.27}$$

Computations done with $k = 1$ and using (5.27) have produced good results.

If $k = 2$, the situation is more complicated and we refer to GLOWI-N-SKI-PIRONNEAU [1] for a discussion of this case.

REMARK 5.8. *(It holds for $k = 1, 2$). If some Dirichlet boundary conditions are prescribed somewhere over Γ, the positive cone used to define the discrete entropy condition will be related to the subspace of V_h consisting of those functions vanishing at the boundary nodes corresponding to the (discrete) Dirichlet condition.*

REMARK 5.9. *The optimal choice for the bounding function M (of M_h) may not be an easy task, specially for airfoil computations. However for the approximate problem an almost natural choice is*

$$M_h(x) = C(h(x))^{-\beta}, 0 < \beta < 1, \tag{5.28}$$

where, in (5.28), *C is a positive constant and $h(x)$ is directly related to the local size of the finite element mesh. It follows from* (5.28) *that*

$$\lim_{h \to 0} M_h(x) = +\infty \ \forall x \in \overline{\Omega},$$

224 *but slower that* $(h(x))^{-1}$.

5.4 Approximation of X

If we do not take into account (4.5), (4.6) then X_h is essentially determined by the discrete entropy condition. Then if one uses the regularization method of Sec. 5.3.1 we have $X_h = V_h$. If one uses the methods described in Sec. 5. 3.2 then X_h is defined by the linear inequality constraints formulating the discrete entropy condition discussed in Sec. 5.3.2.

6 Iterative Solution of The Approximate Problems

6.1 Preliminary statements, generalities

Since we cannot discuss in detail the iterative solution of all the various approximate problems described in Sec. 5, we shall restrict our attention to the following situation:

- Ω is *bounded* and we only have Neumann boundary conditions. We assume also that Kutta-Joukowsky conditions are not required (their treatment is not specific to transonic flows).

- We do not take into account (4.5), (4.6).

- We suppose that $\alpha = 1$ in (4.2) and that an entropy condition is formulated using the method discussed in Sec. 5.3.2 with the formulation (5.27) (controls constraint).

- We finally assume that we work with piecewise linear finite elements ($k = 1$).

Since we do not take into account (4.5), (4.6), X_h is the *closed convex* set of V_h defined by (5.27). Therefore the approximate control problem is a *nonlinear* (and *non-convex*) *programming problem* in which the independent variable is ξ_h. To solve this problem, our strategy is to use *descent methods* (like gradient, conjugate gradient) taking into account the linear inequality constraints (5.27). To handle these constraints one can use separately either *penalty methods* or *dual iterative methods* using the *Kuhn - Tucker multipliers* related to the linear inequality constraints (5.27). Actually a good strategy is to combine both methods using a convenient *augmented Lagrangian* (cf. HESTENES [1], POWELL [1], etc . . .).

6.2 A saddle - point formulation of the approximate problem. Augmented lagra ngian

We use the notation of Sec. 5.3.2. Let us define over V_h the following *approximate* $L^2(\Omega)-$ *scalar product* **225**

$$(u_h, v_h)_h = \sum_{i=1}^{N_h} m_i u_h(P_i) v_h(P_i), P_i \in \sum_h \forall i, \tag{6.1}$$

where

$$m_i = \text{ measure } (\Omega_i), \Omega_i = \dot{\overline{\Omega}}_i \text{ where}$$

$$\overline{\Omega}_i = \text{ union of } T \in \mathscr{C}_h \text{ such that } P_i \text{ is a vertex of } T;$$

the correspondent *norm* is denoted by $|\cdot|_h$. Then we define $\Delta_h^* : V_h \to V_h$ by

$$(\Delta_h^* \phi_h, v_h)_h = -\int_\Omega \nabla \phi_h \cdot v_h dx \ \forall v_h \in V_h. \tag{6.2}$$

Therefore it follows from (6.2) that (5.27) can also be written

$$(\Delta_h^*(\xi_h - \Phi_{oh}), \phi_h)_h \leq 0 \ \forall \phi_h \in V_h^+, \tag{6.3}$$

which is equivalent to

$$\Delta_h^*(\xi_h - \Phi_{oh}), (P_i) \leq 0 \ \forall P_i \in \sum_h . \tag{6.4}$$

The augmented Lagrangian $\mathscr{L}_r : V_h \times V_h \to \mathbb{R}$ to be used is then defined by

$$\begin{cases} \mathscr{L}_r(\xi_h, \mu_h) = \frac{1}{2} \int_\Omega \rho(\xi_h)(\Phi_h - \xi_h)^2 dx + \frac{r}{2} |(\Delta_h^*(\xi_h - \Phi_{oh}))^+|_h^2 - \\ - \int_\Omega \nabla \mu_h \cdot \nabla(\xi_h - \Phi_{oh}) dx, \end{cases} \tag{6.5}$$

where in (6.5), $(\phi_h)^+$ *does not* denote the positive part of ϕ_h but the approximation of it defined by

$$\begin{cases} \forall \phi_h \in V_h, (\phi_h)^+ \in V_h \text{ and} \\ (\phi_h)^+(P_i) = \max(0, \phi_h(P_i)) \ \forall P_i \in \sum_h . \end{cases} \tag{6.6}$$

In (6.5), Φ_h is a function of ξ_h *through the state equation*(5.5).

226 Since the constraints are *linear inequality constraints* we have

Proposition 6.1. *If the approximate control problem has a solution then $\mathscr{L}_r$ has a saddle- point $\{\overline{\xi}_h, \lambda_h\}$ over $V_h \times V_h^+$ with $\overline{\xi}_h$ solution of the approximate control problem.*

REMARK 6.1. *The function $\lambda_h \in V_h^+$ is the* Kuhn - Tucker *multiplier of the problem. Its existence follows from the fact that we have a finite dimensional problem with linear constraints. Then the existence of a solution implies the existence of a Kuhn-Tucker multiplier.*

6.3 Iterative solution of the approximate problem via $\mathscr{L}_r$

6.3.1 Description of the algorithm

To solve the approximate problem we shall use an algorithm of Uzawa's type (see CEA [1], G.L.T. [1, Ch. 2]) which will compute a saddle - point of $\mathscr{L}_r$ over $V_h \times V_h^+$. This algorithm is the following

$$\lambda_h^0 \in V_h^+, \text{arbitrarily given}(\lambda_h^0 = 0 \text{ for example }), \tag{6.7}$$

λ_h^n known we compute $\{\Phi_h^n, \xi_h^n\} \in V_h \times V_h$ and λ_h^{n+1} by

$$\begin{cases} \mathscr{L}_r(\xi_h^n, \lambda_h^n) \leq \mathscr{L}_r(\xi_h, \lambda_h^n) \ \forall \xi_h \in V_h, \\ \xi_h^n \in V_h, \end{cases} \tag{6.8}$$

$$\xi_h^n \text{ gives } \Phi_h^n \text{ through (5.5)} \tag{6.9}$$

$$\begin{cases} \int_\Omega \nabla \lambda_h^{n+1} \cdot \nabla(\mu_h - \lambda_h^{n+1}) \geq \int_\Omega \nabla \lambda_h^n \cdot \nabla(\mu_h - \lambda_h^{n+1}) dx - \\ -\rho \int_\Omega \nabla(\xi_h^n - \Phi_{oh}) \cdot \nabla(\mu . \lambda_h^{n+1}) dx \forall \mu_h \in V_h^+, \lambda_h^{n+1} \in V_h^+. \end{cases} \tag{6.10}$$

6.3.2 Solution of (6.10)

The problem (6.10) is a *finite dimensional variational inequality* in V_h. This problem is very close to the *obstacle problem* of Chapter 2, Sec. 2; therefore it can be solved by an *overrelaxation method with projection.*

6.3.3 Solution of (6.8)

The problem (6.8) is a finite dimensional control problem. We have solved this problem using the *Polak - Ribiere* version of the non-linear conjugate gradient method (see POLAK [1, Ch. 2, pp. 53 -55]) which seems more effective (for our problem) than the *Fletcher- Reeves* version (we recall that in Chapter 4, Sec. 2.6.7 these two methods are described, when applied to the solution of a specific problem). The scalar **227**
product used in this algorithm is the scalar product induced by $H^1(\Omega)$ over V_h. Therefore a very important step in the solution of (6.8) by the above conjugate gradient algorithm is the computation of the *partial gradient* $\dfrac{\partial \mathscr{L}_r}{\partial \xi_h}(\xi_h, \lambda_h^n)$; this point is discussed in the next section.

6.3.4 Computation of $\dfrac{\partial \mathscr{L}_r}{\partial \xi_h}$

Owing to the practical importance of $\dfrac{\partial \mathscr{L}_r}{\partial \xi_h}$ we shall discuss its computation in some detail: we have

$$\begin{cases} \mathscr{L}_r(\xi_h, \mu_h) = \frac{1}{2}\int_\Omega \rho(\xi_h)|\nabla(\Phi_h - \xi_h)|^2 dx + \frac{r}{2}|(\Delta_h^*(\xi_h - \Phi_{oh}))^+|_h^2 - \\ - \int_\Omega \nabla\mu_h \cdot \nabla(\xi_h - \Phi_{oh})dx, \end{cases} \tag{6.11}$$

where, in (6.11), Φ_h and ξ_h are related by (5.5). It follows from (6.11) that

$$\begin{cases} \frac{\partial \mathscr{L}_r}{\partial \xi_h}(\xi_h, \mu_h)\cdot\delta\xi_h = \frac{1}{2}\int_\Omega \delta\rho(\xi_h)|\nabla(\Phi_h - \xi_h)|^2 dx + \int_\Omega \rho(\xi_h)\nabla(\Phi_h - \xi_h)\times \\ \times\nabla\delta(\Phi_h - \xi_h)dx + r((\Delta_h^*(\xi_h - \Phi_{oh}))^+, \Delta_h\delta\xi_h)_h - \int_\Omega \nabla\mu_h\cdot\nabla\delta\xi_h dx. \end{cases} \tag{6.12}$$

We have no difficulty to compute the last two terms of the right hand side of (6.12). About the second term, we obtain by differentiation of (5.5)

$$\int_\Omega \delta\rho(\xi_h)\nabla\Phi_h\cdot\nabla\phi_h dx = -\int_\Omega \rho(\xi_h)\nabla\delta\Phi_h\cdot\nabla\phi_h dx \;\forall\phi_h \in V_h. \tag{6.13}$$

Taking $\phi_h = \Phi_h - \xi_h$ in (6.13) we obtain

$$\begin{cases} \int_\Omega \rho(\xi_h)\nabla(\Phi_h - \xi_h)\cdot\nabla\delta(\Phi_h - \xi_h)dx = -\int_\Omega \rho(\xi_h)\nabla(\Phi_h - \xi_h)\cdot\nabla\delta\xi_h dx - \\ - \int_\Omega \delta\rho(\xi_h)\nabla\Phi_h\cdot\nabla(\Phi_h - \xi_h)dx. \end{cases} \tag{6.14}$$

It follows from (6.14) that the sum of the first two of the right hand side of (6.12) is

$$-\int_\Omega \rho(\xi_h)\nabla(\Phi_h - \xi_h)\cdot\nabla\delta\xi_h dx - \frac{1}{2}\int_\Omega \delta\rho(\xi_h)\nabla(\Phi_h - \xi_h)\cdot\nabla(\Phi_h + \xi_h)dx. \tag{6.15}$$

Since

$$\frac{1}{2}\delta\rho(\xi_h) = \frac{1}{2}\frac{d\rho}{d\xi_h}(\xi_h)\cdot\delta\xi_h = -\frac{1}{(\gamma+1)C_*^2}\left(1 - \frac{|\nabla\xi_h|^2|}{\frac{\gamma+1}{\gamma-1}C_*^2}\right)^{\frac{2-\gamma}{\gamma-1}}\nabla\xi_h\cdot\nabla\delta\xi_h, \tag{6.16}$$

228 the second term of (6.5) is easily computed and, by addition with the last two terms of the right hand side of (6.12), we obtain $\frac{\partial \mathscr{L}_r}{\partial \xi_h}(\xi_h, \mu_h) \cdot \delta\xi_h$.

REMARK 6.2. *If instead of taking* $\alpha = 1$ *in* (5.6) *one takes* $\alpha = 0$, *then the computation of* $\frac{\partial \mathscr{L}_r}{\partial \xi_h}$ *will require the use of an* adjoint state equation *(see LIONS [4], CEA [1], [2]).*

6.4 Computational considerations

When solving (6.8) by the non-linear conjugate gradient method discussed above we have to solve at each iteration the state equation (5.5). Since the bilinear form occurring in (5.5) is positive definite (once the value of Φ_h in *one* point of $\overline{\Omega}$ has been prescribed) we can use to solve (5.5) either *iterative methods* like *conjugate gradient, overrelaxation* etc., (cf., e,g., POLAK [1], CONCUS GOLUB [1], AXELSSON [1], VARGA [1], YOUNG [1]) or direct methods like Cholesky's. About the choice of r and ρ in $\mathscr{L}_r$ and (6.7) - (6.10) we can say that the *larger* is r the *more ill - conditioned* is (6.8). However the larger is r the faster will be the global convergence of (6.7) - (6.10) for a "convenient choice" of ρ(*round - off errors* being neglected). Once r has been chosen , theoretical considerations indicate that ρ has to be chosen of the same order as r.

7 A Numerical Experiment

We limit the presentation to only one example ; for more examples see GLOWINSKI - PIRONNEAU [1]. The example we have considered is the two piece airfoil of Figure 7.1; *Kutta - Joukowsky conditions* have to be imposed on both profiles. The position of the piece makes the funnel slightly convergent. The main airfoil is at 5^0 of incidence and the Mach number at infinity is $M = 0.55$. Piecewise linear finite elements ($k = 1$) were used with 2936 triangles and 1555 nodes.

The regularization method of Sec. 5.3.1 has been used and the results, showing Mach - lines, of Figure 7.1 were obtained after 50 iterations of conjugate gradient. We observe that no non-physical shocks are

present and that the Mach number on the exit of the funnel is precisely equal to one, which it should be.

The precision can be guessed by measuring $\nabla(\Phi_h - \xi_h)\|$; at $n = 0$ its value is 3.5×10^3 at $n = 50$ it is 2. On each triangle it varies from 10^{-5}
229 in the subsonic region to 10^{-2} in the supersonic.

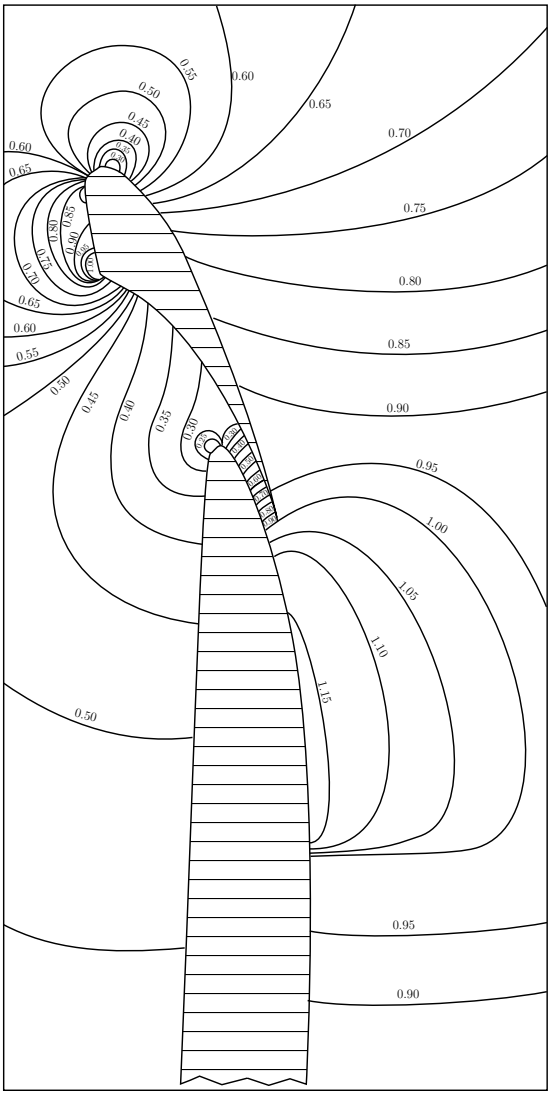
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00
1.05
1.10
1.15
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

Figure 7.1:

8 Comments Conclusion

230 The methods that we have described above seem effective for computing transonic flows on complicated domains since finite elements approximations are used. Moreover the non-linear programming approach that we have used (based on an optimal control formulation) gives much flexibility for taking into account the *entropy condition* and for the choice of the iterative methods for solving the approximate problem. We have to observe that regularization method of Sec. 5.3.1 is actually a method for computing those solutions such that

$$(\Delta\Phi)^+ \in L^2(\Omega). \tag{8.1}$$

If one wishes to approximate (3.7), i.e.

$$(\Delta\Phi)^+ \in L^\infty(\Omega), \tag{8.2}$$

one may use a regularization functional like

$$\int_\Omega \in (x)|(\Delta\phi)^+|^p dx,\ p \text{ “large”}. \tag{8.3}$$

All these methods can be extended to 3-dimensional computations, but one of the main difficulties is then the treatment of the Kutta - Joukowsky condition. For more details and numerical experiments, and other methods for treating the entropy condition we refer to GLOWINSKI- PIRONNEAU [1], BRISTEAU [2], BRISTEAU -GLOWINSKI-PERIAUX-PERRIER-PIRONNEAU-POIRIER [1], [2], BRISTEAU - GLOWINSKI- PERIAUX-PERRIER-PIRONNEAU [1].

In CEA-GEYMONAT [1] one may find results on the solution of nonlinear boundary value problems via optimal control.

Let us mention to conclude that various methods for treating shocks in fluid mechanic problems can be found in LASCAUX [1] and the bibliography therein.

Bibliography

[1] AGMON S., DOUGLIS A., NIRENBERG L. **231**

1. Estimates near the boundary for solution of elliptic partial differential equations satisfying general boundary conditions (I). Comm. Pure Applied Math., 12, (1959), pp. 623-727.

[2] AXELSSON 0.

1. A class of iterative methods for finite element equations, Comp. Meth. Appl. Mech. Eng., 19 (1976), pp. 123-138.

[3] BAIOCCHI, C.

1. Sur un probleme a frontiere libre traduisant le filtrage de liquides a travers des milieux poreux. C.R. Acade. Sc. Paris, 273 A, (1971), pp. 1215 - 1217.

[4] BAUER F., GARABEDIAN P., KORN D.

1. Superitical wing section (I). Lecture Notes in Economics and Math. Systems, Vol. 66, Springer - Verlag, Berlin, 1972.

[5] BAUER F., GARABEDIAN P., KORN D., JAMESON A.

1. Supercritical wing section (II). Lecture Notes in economics and Math. Systems Vol. 108, Springer- Verlag, Berlin, 1975.

[6] BEGIS D.

1. Analyse numerique de l' ecoulement d'un fluide de Bindham. These de 3eme cycle, Universite Pierre et Marie Curie, Paris, 1972.
2. Etude numerique de l' ecoulement d' un fluide visco-plasique de Bingham par une methode de lagrangien augmente. Laboria Report 355, 1979.

[7] BEGIS D., GLOWINSKI R.

1. Chapter 7 of Numerical solution of boundary value problems by augmented lagrangians, M. Fortin, R. Gloinski ed. (in preparation).
2. Application de la methode des elements finis a la resolution de problemes de domaine optimal. Methods de resolution des problems approches. Applied Math. Optimization, 2, (1975), 2, pp. 130-169.

[8] BENSOUSSAN A., LIONS J.L.

1. Sur l' approximation numerique d' inequations quasi - variationnelles statonnaires, in Computing Methods in Applied Sciences and Engineering, Part 2, R. Glowinski, J.L. Lions eds., Lecture Notes in Computer Sciences, Vol. 11, Springer - Verlag, 1974.

[9] BENSOUSSAN A., LIONS J.L., TEMAM R.

1. Sur les methods de decomposition, de decentralisation, de coordination et applications. In les Sur Les methods numeriques en sciences physiques et eonomiques, J.L. Lions, G. I. Marchouk eds., Dunod- Bordas, Paris, 1974, pp. 133-257.

[10] BERGER A.E

1. The truncation method for the solution of a class of variational inequalities Revue Francaise Automatique, Informaique, Rech. Operationnelle, Vol. de, N^0 (1956), pp. 29-42.

[11] BERS L. **232**

1. Mathematical aspects of subsonic and transonic gas dynamics, Chapman and Hall, London, 1958.

[12] BOURGAT J.F., DUVAUT G.

1. Numerical Analysis of flows with or without wake past a symmetric two - dimensional profile, with or without incidence. Int. J. Num. Meth. Eng., 11, (1977), pp. 975 - 993.

[13] BRENT R.

1. Algorithms for minimization without derivative, Prentice Hall, N. J., 1973.

[14] BREZIS H

1. A new method in the study of subsonic flows. In Partial Differential Equations and Related Topics, J. Goldstein ed., Lecture Notes in Math., Vol. 446, Springer - Verlag, Berlin, 1975, pp. 50-60.
2. Multiplicateur de Lagrange en torsion elasto - platique. Arch. Rat. Mech. Anal. 49, (1972), pp. 32-40.
3. Problems unilateraux. J. de Math. Pures et Appliquess 9, Serie 72, (1971), pp 1-168.
4. Monotonicity in Hilbert spaces and some applications to nonlinear partial differential equations. In Contributions to Nonlinear Functional Analysis, E. Zarantonello ed., Acad. Press, New-York, 1971, pp. 101-156.c

5. Operateurs maximaux monotones et semmi - groups de contraction dans les espaces de Hilbert, North - Holland, Amsterdam, 1973.

[15] BREZIS H., SIBONY M.

1. Equivalence de deux inequations variationnelles et application Arch. Ract. Mech. Anal. 41, (1971), pp. 254-265.

[16] BREZIS H., STAMPACCHIA G.

1. The hodograph method in fluid dynamics in the light of variationsl inequalities, Arch, Rat. Mech. Anal. 61, (1976), pp. 1-18.
2. Sur la regularite de la solution d' inequations elliptiques. Bull. Soc. Math. France, 96, (1968), pp. 153-180.

[17] BREZIS H., CRANDALL M., PAZY A.

1. Perturbation of nonlinear monotone sets in Banach spaces. Comm. Pure Applied Math., Vol. 23, (1970), pp. 123-144.

[18] BREZZI F., HAGER W.W., RAVIART P.A.

1. Error estimates for the finite element solution of variational inequalities, Part I- Primal Theory. Numer. Math. 28, (1977), pp. 431 -443.

[19] BREZZI F., RAVIART P.A.

1. Mixed finite element methods for 4th order elliptic equations. In Tomos in Numerical Analysis (III), J. J.H. Miller ed., Acad, Press, London, 1976, pp. 33-36.

233 [20] BREZZI F. SACCHI G.

1. A finite element approximation of variational inequalities related to hydraulics. Calcolo, 13, (1976), pp. 259-273.

[21] BRISTEAU M.O.

1. Application de la methode des elements finis a la resoution d'inequations variationnnelles d'evolution de type Bingham, These de 3ene cycle, Universite Pierre et Marie Curie, Paris, 1975.
2. Applications of optimal control theory to transonic flow computations by finite element methods. In Computing methods in Applied Sciences and Engineering, 1977, II, R. Glowinski, J.L. Lions ed., Lecture Notes in Physics Vol. 91, Springer - Verlag, Berlin, 1979, pp. 103 - 124.

[22] BRISTEAU M.O., GLOWINSKI R.

1. Finite element analysis of the unsteady flow of a viscous - plastic fluid in a cylindrical pipe. In Finite Element methods in Flow Problems, J. T. Oden, O.C. Zienkiewicz, R. H. Galagher, C. Taylor eds., University of Alabama Press, Huntsville, Alabama, 1974, pp. 471-488.

[23] BRISTEAU M.O., GLOWINSKI R., PERIAUX J., PERRIER P., PIRONNEAU O.

1. On the numerical solution of nonlinear problems in fluid dynamics by least squares and finite element methods. (I) Leas square formulation and conjugate gradient solution of the continuous problem. Comp. Meth. Applied Mech. Eng., 17/18, (1979), pp. 619-657.

[24] BRISTEAU M.O., GLOWINSKI R., PERIAUX J., PERRIER P., PIRONNEAU 0., POIRIER G.

1. Applications of optimal control and finite element methods to the calculation of transonic flows and incompressible viscous flows. Laboria Report 294, (1978), and Numerical methods in applied fluid dynamics, B. hunt ed., Acad. Press, London (to appear).

2. On the numerical solution of nonlinear problem in fluid dynamics by least square and finite element methods. (II) Application to the computaion of transonic potential flows of compressible inviscid fluids (to appear).

[25] CARTAN H.

1. Calcul Differentiel. Hermann, Paris, 1967.

[26] CEA J., GEYMONAT G.

1. Une methode de linearisation via 1' optimistion, Instituto di Alta Mat. Sym. Mat.m 10, (1972), Bologna, pp. 431 - 451.

[27] CEA J.

1. Optimization : Theorie et Algorithmes. Dunod, Paris, 1971.
2. Optimization - Theory and Algorithms. Tata Institute of Fundamental Research, Bombay, published by Springer-Verlag, Berlin, 1978.

[28] CEA J., GLOWINSKI R.

1. Sur de methods d' optimisation par relaxation. Revaue Francaise Automatique, Informat. Rech. Operationnelle, R-3, (1973), pp 53-32.
2. Methods numeriques pour 1'ecoulement laminaire d' un fluide rigide viscoplatque incompressible. Interm. J. Computer March., Sect. B, Vol. 3, (1972), pp. 225-255.

234 [29] CEA J., GLOWINSKI R., NEDELEC J.C.

1. Applications des methods d'optimisation, de differences et d' elements finis, a 1' analyse numerique de la torision elasto - plastique d' une barre cylindrique. In Approximations et Methods Iteratives de Resolution d' Inequations Variationnelles et de Problems Non Lineaires. Cahier de 1' IRIA, N^0 12, (1974), PP. 7-138.

[30] CHAN T.F., GLOWINSKI R.

1. Finite element approximation and iterative solution of class of midly nonlinear elliptic equations. Stanford University Report STAN-CS - 78-674, Comp. Science Department, Stanford University, 1978.

[31] CIARLET P.G.

1. The Finite Method, Lecture Notes, Tata Insitute of Fundamental Research, Bombay, 1975.
2. The Finite Element Method for Elliptic Problems. North Holland, Amsterdam, 1978.
3. Numerical Analysis of the Finite Element Method. Seminaire de Math. Superieures Preses de 1' Universite de Montreal, 1976.

[32] CIARLET P.G., RAVIART P. A.

1. Interpolation theory over curved elements with applications to finite element methods. Comp. Meth. Appl. Mech. Eng., 1, (1972), pp. 217 - 249.
2. A mixed finite element method for the biharmonic equation. In Mathematical aspects of finite element in partial differential equation, C. de Boor, ed., Acad. Press, N. Y., 1974, pp. 125 - 145.

[33] CIARLET P.G., SCHULTZ M.H., VARGA R.S.

1. Numerical methods of high order accuracy for nonlinear boundary value problems. V. Monotone operator theory, Numer. Math., 13, (1969), pp, 51-77.

[34] CIARLET P. G., WAGSHAL C.

1. Multipoint Taylor Formulas and applications to the Finite Element Method. Numerishe Math., Vol. 17, (1971), pp. 84-100.

[35] CIAVALDINI J.F., TOUNEMINE G.

1. A finite element method to compute stationary steady flows in the hodograph plane. J. of Indian Math. Soc. 41, (1977), pp. 69-82.

[36] CIVALDINI J.F., POGU M., TOURNEMINE G.

1. Approximation des ecoulements compressibles autour d'un profil regulier place en atmoshphere infine: estimation asymptotique lorsque 1' on borne le domine exteriur au profile. Rapport de l'Universite de Rennes I et de 1' Institut National des Sciences Appliquees, March 1979.
2. Une methode variationnelle non lineaire pour 1' etude dans plan physique d' ecoulements compressibles subcritiques en atmoshphere infinie. Compt. Rend. Acad. Sciences Paris, t. 281 A, (1975), pp. 1105 - 1108.

[37] COMINCIOLI V.

1. On some oblique derivative problems arising in the fluid flow in porous media. A theoretical and numerical approach. Applied Math. and Optimization, Vol. 1, N^0 4, (1975), p. 313 -336

235 [38] CONCUS P., GOLUB G.H.

1. A generalized conjugate gradient method for non symmetric systems of linear equations. In Computing methods in Applied Sciences and Engineering, R. Glowinski, J.L. Lions ed., Lecture Notes in Economics and Math, Systems, Vol. 134, Springer - Verlag, Berlin, 1976, pp. 56-65.

[39] CROUZEIX M.

1. Sur l' approximation des equations differentielles operationnelles par des methodes de Runge-Kutta. These d'Etat, Universite Pierre et Marie Curie, Paris, 1975.

[40] CRYER C. W.

1. The method of Christoferson for solving free boundary problems for infinite journal bearings by means of finite differences. Math. Comp. 25, (1971), pp. 435-443.

[41] DANIEL J.

1. The approximate minimization of functionals. Prentice Hall, N.J., 1970

[42] DUVAUT G., LIONS J.L.

1. Les inequations en Meanique et en Physique. Dunod, Paris, 1972.

[43] EKELAND I., TEMAM R.

1. Analyse Convexe et Problemes Variationnels. Dunod - Gauthier - Villars, Paris, 1974.

[44] FALK R.S.

1. Approximate solutions of some Variational Inequalities with Order of Convergence Estimates. Ph. D. Thesis, Cornell University, 1971.
2. Error estimates for the approximation of a class of Variational Inequalities, Math. Comp., 28, (1974), pp. 963-971.
3. Approximation of an Elliptic Boundary Value Problem with Unilateral Constraints Rev. Francaise Automat. Informat. Rech. Operationnelle. R2, (1975), pp. 5-12.

[45] FALK R.S, MERCIER B.

1. Error esimate for elasto - plastic problems. Rev. Francaise Automat. Informat Rech. Operationnelle, 11, (1977), pp. 135-144.

[46] FORTIN M.

1. Calcul numerique des ecoulements des fluides de Bingham et des fluides newtonies incompressibles par des methodes d' elements finis. These d'Ettat, Universite Pierre et Marie Curie, Paris, 1972
2. Minimization of some non-differentiable functionals by the augment lagrangian method of Hestenes and Powell. Appl. Math. Opt., 2, (1976), pp. 236-250.

[47] FORTIN M., GLOWINSKI R.

1. Chapter 3 of Resolution numerique de problems oux limites par des methodes de lagrangien augmente, M. Fortin, R. Glowinski ed.(in preparation).
2. (Ed.) Resolution numerique de problemes our limites par des mithodes de lagrangien augmente, (in preparation).

236 [48] GABAY D., MERCIER B.

1. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comp. and Math. with Applications, Vol. 2, (1976), N^0 1, pp. 17 - 40.

[49] GERMAIN P.

1. Mecanique des MIleux Continus, Vol. 1, Masson, Paris, 1973.

[50] GIRAULT V., RAVIART P.A.

1. Finite Element Approximation of the Navier- Stokes Equations. Lecture Notes in Math., Vol. 749, Springer - Verlag, Berlin, 1979.

[51] GLOWINSKI R.

1. Introduction to the Approximation of Elliptic Variational Inequalities. Report 76006, Laboratoire d' Analyse Numerique, Universite Pierre et Marie Curie, 1976.
2. Analyse Numereique d' Inequations Variationnelles d'orde quatre. Reporf 75002, Laboratoire d' Analyse Numerique, Universite Pierre et Marie Curie, 1975.
3. Sur 1' approximation d'une inequation variationnelle elliptique de type binghan, Revue Francaise d' Automatique, Informatique, Recherche Operaionnelle, Vol. 10, 12, (1976), pp. 13-30.
4. Sur 1' ecoulement d'un fluide de Bingham dans une conduite concuite cylidrique. J. de Mecanique, Vol. 1, N^0 4, (1974), pp. 601 - 621.
5. Numerical Analysis of Nonlinear Boundary Value Problems. (I) Methods of Convexity and and Monotonicity (in preparation).
6. Approximations externes par elements finis d'ordre un et deux du problem de Direchlet pour Δ^2. In Topics in Numerical Analysis (I), J. J. H. Miller Ed., Acad. Press, London, 1973, pp. 123 - 171.

[52] GLOWINSKI R., LANCHON H.

1. Torsion elasto - plastique d' une barre cylindrique de section multiconnexe Journal de Mecanique, 12, (1971), 1, pp. 151 -171.

[53] GLOWINSKI R., LIONS J.L., TREMOLIERES R.

1. Analyse Numerique des Inequations Vaviationnelles, Vol. 1, Theorie Generale et Premieres Applications. Dunod - Bordas, Paris, 1976.
2. Analyse Numerique des Inequations Variationnelles, Vol. 2, Applications aux Phenomenes Stationnaires et d' Evolution. Dunod - Bordas, Paris, 1976.
3. Numerical Analysis of Variational Inequalities, North-Holland, Amsterdam (to appear).

[54] GLOWINSKI R., MARROCCO A.

1. Analyse Numerique du champ magnetique dans un alternatieur tetrapolaire par la methode des elements finis et surrelaxation ponctuelle non lineaire . Comp. Meth. Applied. Mech. Eng., 3, (1974), pp. 55-85
2. Etude numerique du champ magnetique dans un alternateur tetrapolaire par la methoks des elements finis. In Computing Methods in Applied Sciences and Engineering, Part 1, R. Glowinski, J.L. Lions ed., Lecture Notes in Comp. Sciences, Vol. 10, Springer-verlag, Berlin, 1974, pp. 392-409.
3. Chapter 5 of Resolution Numerque de Problems aux Limites par des Methodes de Lagrangien Augmente, M. Fortin, R. Glowinski eds., (in preparation).

237

4. Sur 1' approximation par elemets finis d'ordre un et la resolution per penalisation-dualite d'nue classe de problems de Dirichlet non lineaire. Compt. Rend. Acad.Sc., Paris, t. 278 A, (1974), pp (1964-1652)
5. On the solution of a class nonlinear Dirichlet problems by a penalty- duality method and dinite elemet of order one. In Optimization Techmiques : IFIP Technical Conference, G.I.Marchouk ed. Lecture Notes in Computer Sciences, Vol.27, Springer-Verlag, Berlin, 1975, pp 327-333.
6. Sur 1' approximation per elemets finis d'ordre un et la resolution par penalisation-dualite, d'une classe de problems de

Dirichlet non lineaire. Revue Francaise d' Automatique, Informatique, Recherche Operationnelle, Analyse Numerique, R-2 (1975), pp. 41-76.

7. Numerical solution of two dimensional magneto-static problems by augmented lagrangian methods. Comp. Meth. Appl. Mech. Eng. ; 12, (1977), pp. 33-46
8. Sur 1' approximation par elements finis order un et la resolution par penalistion dualite d'une classe de problems de Dirichlet non lineaires, Rapport Laboria 115, 1975 (extended version of [6]).

[55] GLOWINSKI R., PIRONNEAU 0.

1. On the computation of transonic flows. In Functional Analysis and Numerical Anaysis, Japan-France Seminar, Tokyo and Kyoto 1976, H.Fujita ed. Japan socity for the Promotion of Science, Tokyo,1978, pp. 143-173.
2. Numerical methods for the first biharmonic equation and for the two-dimensional Stockes problems. SIAM Review, 21, (1979), 2, pp. 167-212.

[56] HESTENES M.

1. Multiplier and gradient methods. J. Opt.Theoty Appl., 4, (1969), pp. 303-320. HEWITT B.L., ILLINGWORTH C.R., LOCK R.C., MANGLER K.W., Mc DONNEL J.H., RICHARDS C., WALKDEN F.
2. Computational Methods and problems in Aeronautical Fluid Dynamics. Acad. Press, London, 1976.

[57] HOUSEHOLDER A. S.

1. The numerical treatment of a single nonlinear equation. Mc Graw-Hill, N.Y., 1970.

[58] JAMESON A.

1. The numerical treatment of a single nonlinear equation. Mc Graw-Hill, N.Y., 1970.

[59] JAMESON A.

1. Transonic flow calulations. In Numerical Methods in Fluid Dynamics, H.J.Wirz, J.J Smolderen ed. Mc Graw Hill, N.Y., 1978, pp. 1-87.
2. Three dimensional flows around airfoils with shocks. In Computing Methods in Applied Sciences and Engineering (II), R. Glowinski, J.L. Lions ed., Lecture Notes in Comp. Sicence, Vol.11, Springer -Verlag, Berlin, 1974, pp. 185-212.
3. Interative solution of transionc flows over airfoils and wings, including flows at Mach 1.Comm. Pure Appl. Math., Vol.27, (1974), pp. 283-309.
4. Numerical solution of nonlinear partial differential equations of mixed type. In Numerical solution of Partial Differentail Equations-III, Synspads, 1975, B.Habbard ed., Acad.Press, N.Y., 1976, pp. 275-320.

238 [60] JOHSON C.

1. A convergence estimate for an approximation of a parabolic variational inequality. SIAM J. Num. Anal., 13, (1976), 4, pp. 599-606.

[61] KELLOG R.B

1. A nonlinear alternating direction method. Math. of Comp. 23, (1969), 105, pp. 23-27.

[62] KOITER W.T

1. General Theorems for Flastic Plastic solids. Progress in solid mechanics, pp. 165-221. North-Holland, 1960.

[63] LANCHON H.

1. Torsion elasto-plastique d'un arbre cylindrique de section simplement ou multiplement connexe. These d' Elat, Universite Paris VI, 1972.

[64] LANDHU L., LIFCHITZ E.

1. Mecanique des Fluides, Mir,Moscow, 1953.

[65] LASCAUX P.

1. Numerical methods for time dependent equations.Applications to fluid flow problems, lecture Notes, Tata Institute of Fundamental Research, Bombay, 1976.

[66] LIEUTAUD J.

1. Approximation d'operateurs par des methods de decomposition.These d'Etat, Universite Paris VI, 1968.

[67] LIONS J.L

1. Quelques methods se resolution des problems aux limits non lineaire. Dunod, Gauthier-Villars, Paris, 1969.
2. Problems aux limites dans les equations aux derivees partielles. Seminaire de Math. Superieures de 1' Universite de Montreal. Presses de 1' Universite Montreal, 1962.
3. Equations differentielles operationnelles et problems aux limites. Springer Verlag, 1961.
4. Controle Optimal des Systemes gouvernes par des equations aux derivees partielles. Dunod, Paris, 1968.

[68] LIONS J.L., MANGENES E.

1. Problems aux limites nno homogene, Vol, 1. Dunod, Paris, 1968.

[69] LIONS J.L., STAMPACCHIA G.

1. Variational Inequalities. Comm. Pure Applied Math., XX, (1967), pp 493-519.

[70] MORAVETZ C.S.

1. Mixed equations and transonic flows. Rendi Conti di Mat., 25, (1960), pp. 1-28

239 [71] MOSOLOV P.P., MIASNIKOV V.P

1. Variational Methods in the Thoey of the Fluidity of a Viscous-Plastic Medium. J. Mech. and Appl. Math. (P.M.M), Vol.29,(1956), 3,pp. 468-492.
2. On stagnant flow regions of a Viscous-Plastic Medium in Pipes. J. Mech. and Appl. Math (P.M.M.), Vol. 30, (1966), pp. 705-719.
3. On qualitative singularities of the flow of a Viscous-Plastic Medium in Pipes. J. Mech. and Appl. Math. (P. M. M.), Vol. 31, (1967), pp. 581-585.

[72] MURMAN E.J., COLE J.D.

1. Calcution of plane steady transonic flows. AIAA Journal, Vol. 9, (1971), pp. 114-121.

[73] NECAS J.

1. Les Methods Directes en Theorie des Equations Elliptiques. Masson, Paris, 1967.

[74] ODEN J.T., REDDY J.N

1. Mathematical Theory of Finite Elements. Wiley, N.Y., 1976.

[75] OPIAL Z.

1. Weak convergence of the successive approximations for non expansive mappings in Banach spaces. Bull. A.M.S., 73, (1967), pp. 591-597.

[76] ORTEGA J., RHEINBOLDT W.C

1. Iterative solution of nonlinear equations in several variables. Acad. Press, N.Y., 1970.

[77] POLAK E.

1. Computational Methods in Optimization. Acad. Press. N.Y., 1971.

[78] POWELL M.J.D

1. A method for nonlinear constraints in minimization problems. Optimization, R. Fletcher ed., Acad. Press, London, 1969, Ch. 19.

[79] PRAGER W.

1. Introduction to Mechanics of Continua. Ginn and Company, Boston, 1961.

[80] RAVIART P.A

1. The use of numerical integration in finite element methods for solving parabolic equations. In Topics in Numerical Analysis J.J.H Miller ed., Acad. Press, London, 1973, pp., 233-264.
2. Multistep methods and parabolic equations.In Functional Analysis and Numerical Analysis, Japan-France Seminar, Tokyo and Kyoto, 1976, H. Fujita ed., Japan Society for the Promotion of Science, 1978, pp. 429-454.

[81] ROCKAFELLAR T.R.

1. Convex Analysis. Prinection University Press, Princeton, N.J., 1970.

[82] SCHECHTER S.

1. Iteration methods for nonlinear problems. Trans. Amer, Math. Soc., 104, (1962), pp. 601-612.
2. Minimization of Convex Functions by Relaxation ; Ch. 7 of Integer and Nonlinear Programming, J. Abadie ed., North - Holland, Amsterdam, 1970, pp. 177-189.
3. Relaxation methods for convex problems, SIAM J. Num. Anal., 5, (1968), pp. 601 -612.

[83] STAMPACCHIA G.

1. Equations Elliptiques du Second Ordre a Coefficients Discontinus. Seminaire de Math. Superieures de 1' Universite de Montreal, Presses de 1' Universite de Montreal, 1965.

[84] STRANG G.

1. The finte element methof, linear and nonlinear applications. *Proceedings of the Int. Congress of Math., Vol. 2,* pp 429-436.

[85] STRANG G., FIX G.

1. An analysis of the finite element method, Prentice Hall, N.J., 1973.

[86] STRANG G., MOSCO U.,

1. One sided approximation and variational inequalities. Bull. American Math. Soc., 80, (1974), pp. 308 - 312.

[87] STRAUSS M.J.

1. Variations of Korn's and Sobolev's Inequalities. In Proceedings of Symposia in Pure Math., Vol. 23, A.M.S., Providence, R. I., 1973, pp. 207 - 214.

[88] TEMAM R.

1. Navier - Stokes equations. North - Holland, Amsterdam, 1977.

[89] TREMOLIERES R.

1. Inequations Variationnelles : Existence, Approximations, Resolution. These d'Etat. Universite Pierre et Marie Curie, Paris, 1972.

[90] VARGA R.S.

1. Matrix Iterative Analysis, Prentice-Hall, N.J., 1962.

[91] YOSIDA K.

1. Fuctional Analysis, Springer-Verlag, Berin, 1965.

[92] YOUNG D.M.

1. Itarative solution of Large Linear Systems. Acad. Press, New-York, 1971.