Lectures on
Topics In Finite Element Solution of Elliptic Problems

By

Bertrand Mercier

Tata Institute of Fundamental Research
Bombay
1979

Lectures on
Topics In Finite Element Solution of Elliptic Problems

By

Bertrand Mercier

Notes By

G. Vijayasundaram

Published for the

Tata Institute of Fundamental Research, Bombay

Springer-Verlag
Berlin Heidelberg New York
1979

Author

Bertrand Mercier
Ecole Polytechnique
Centre de Mathematiques Appliquees
91128 Palaiseau (France)

ISBN 3–540–09543–8 Springer-Verlag Berlin. Heidelberg. New York
ISBN 0–387–09543–8 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

Preface

THESE NOTES SUMMARISE a course on the finite element solution of Elliptic problems, which took place in August 1978, in Bangalore.

I would like to thank Professor Ramanathan without whom this course would not have been possible, and Dr. K. Balagangadharan who welcomed me in Bangalore.

Mr. Vijayasundaram wrote these notes and gave them a much better form that what I would have been able to.

Finally, I am grateful to all the people I met in Bangalore since they helped me to discover the smile of India and the depth of Indian civilization.

Bertrand Mercier
Paris, June 7, 1979.

Contents

Preface **v**

1 Sobolev Spaces **1**

1.1 Notations 1
1.2 Distributions 2
1.3 Sobolev Space 3
1.4 Negative Properties of $H^1(\Omega)$ 5
1.5 Trace Theorem 6
1.6 Dual Spaces of $H^1(\Omega)$ and $H^1_\circ(\Omega)$ 7
1.7 Positive Properties of $H^1_\circ(\Omega)$ and $H^1(\Omega)$. 8

2 Abstract Variational Problems and Examples **11**

2.1 Dirichlet Problem 11
2.2 Abstract Variational Problem. 13
2.3 Neumann's problem. 16
2.4 Mixed Problem. 19
2.5 Elasticity Problem. 23
2.6 Stokes Problem. 27

3 Conforming Finite Element Methods **31**

3.1 Approximate Problem. 31
3.2 Internal Approximation of $H^1(\Omega)$. 33
3.3 Finite Elements of Higher Degree. 37
3.4 Internal Approximation of $H^2(\Omega)$. 43

4 Computation of the Solution of the Approximate Problem 49
4.1 Introduction 49
4.2 Steepest Descent Method 50
4.3 Conjugate Gradient method 52
4.4 Computer Representation of a Triangulation 55
4.5 Computation of the Gradient. 56
4.6 Solution by Direct Methods 58

5 Review of the Error Estimates for the... 65

6 Problems with an Incompressibility Constraint 71
6.1 Introduction 71
6.2 Approximation Via Finite Elements of Degree 1 72
6.3 The Fraeijs De Veubeke - Sander Element 75
6.4 Approximation of the Stokes Problem... 77
6.5 Penalty Methods. 78
6.6 The Navier-Stokes Equations. 78
6.7 Existence and Uniqueness of Solutions of... 79
6.8 Error Estimates for Conforming Method 82

7 Mixed Finite Element Methods 85
7.1 The Abstract Continuous Problem 85
7.2 The Approximate Problem 91
7.3 Application to the Stokes Problem 93
7.4 Dual Error Estimates for $u - u_h$ 96
7.5 Nonconforming Finite Element Method for... 98
7.6 Approximate Brezzi Condition 104
7.7 Dual Error Estimate for the Multiplier 107
7.8 Application to Biharmonic Problem 108
7.9 General Numerical Methods for the... 110
7.10 Equilibrium Elements for the Dirichlet Problem 113
7.11 Equilibrium Elements for the Plate Problem 118

8 Spectral Approximation for Conforming Finite... 129
8.1 The Eigen Value Problem 129
8.2 The Operator $\top$ 130

8.3 Example 130
8.4 Approximate Problem 131
8.5 Convergence and Error Estimate for the Eigen Space. . . 132
8.6 Error Estimates for the Eigen Values. 134
8.7 Improvement of the Error Estimate for the... 136

9 Nonlinear Problems 141
9.2 Generalization 146
9.3 Contractive Operators. 147
9.4 Application to Unconstrained Problem 154
9.5 Application to Problems with Constraint. 156

Bibliography 163

Chapter 1

Sobolev Spaces

IN THIS CHAPTER the notion of Sobolev space $H^1(\Omega)$ is introduced. **1**
We state the Sobolev imbedding theorem, Rellich theorem, and Trace theorem for $H^1(\Omega)$, without proof. For the proof of the theorems the reader is referred to ADAMS [1].

1.1 Notations

Let $\Omega \subset \mathbb{R}^n (n = 1, 2 \textit{ or } 3)$ be an open set. Let Γ denote the boundary of Ω, it is assumed to be bounded and smooth. Let

$$L^2(\Omega) = \left\{ f : \int_\Omega |f|^2 \, dx < \infty \right\} \quad \text{and}$$

$$(f, g) = \int_\Omega fg \, dx$$

Then $L^2(\Omega)$ is a Hilbert space with $(\cdot\,,\cdot)$ as the scalar product.

1.2 Distributions

Let $\mathscr{D}(\Omega)$ denote the space of infinitely differentiable functions with compact support in Ω. $\mathscr{D}(\Omega)$ is a nonempty set. If

$$f(x) = \begin{cases} \exp\left(\frac{1}{|x|^2-1}\right) & \text{if} \quad |x| < 1 \\ 0 & \text{if} \quad |x| \geq 1 \end{cases}$$

then $f(x) \epsilon \mathscr{D}(\Omega), \Omega = \mathbb{R}$.

The topology chosen for $\mathscr{D}(\Omega)$ is such that a sequence of elements ϕ_n in $\mathscr{D}(\Omega)$ converges to an element ϕ belonging to $\mathscr{D}(\Omega)$ in $\mathscr{D}(\Omega)$ if there exists a compact set K such that

$$\text{supp } \phi_n, \text{supp } \phi \subset K$$

2 $D^\alpha \phi_n \to D^\alpha \phi$ uniformly for each multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$ where $D^\alpha \phi$ stands for

$$\frac{\partial^{\alpha_1 + \ldots + \alpha_n} \phi}{\partial^{\alpha_1} x_1 \cdots \partial^{\alpha_n} x_n}.$$

A continuous linear functional on $\mathscr{D}(\Omega)$ is said to be a *distribution*. The space of distributions is denoted by $\mathscr{D}'(\Omega)$. We use $<\cdot, \cdot>$ for the duality bracket between $\mathscr{D}'(\Omega)$ and $\mathscr{D}(\Omega)$.

EXAMPLE 1. (a) A square integrable function defines a distribution: If $f \epsilon L^2(\Omega)$ then

$$\langle f, \phi \rangle = \int\limits_\Omega f \phi \, dx \quad \text{for all } \phi \epsilon \mathscr{D}(\Omega)$$

can be seen to be a distribution. We identify $L^2(\Omega)$ as a space of distribution, i.e.

$$L^2(\Omega) \subset \mathscr{D}'(\Omega).$$

(b) The dirac mass δ, concentrated at the origin, defined by

$$\langle \delta, \phi \rangle = \phi(0) \quad \text{for all } \phi \epsilon \mathscr{D}(\Omega)$$

defines a distribution.

DEFINITION. Derivation of a Distribution

If f is a smooth function and $\phi \epsilon \mathscr{D}(\Omega)$ then using integration by parts we obtain

$$\int\limits_{\Omega} \frac{\partial f}{\partial x_i} \phi \, dx = -\int\limits_{\Omega} f \frac{\partial \phi}{\partial x_i} \, dx.$$

This gives a motivation for defining the derivative of a distribution. **3**

If $T \epsilon \mathscr{D}'(\Omega)$ and α is a multi index then $D^\alpha T \epsilon \mathscr{D}'(\Omega)$ is defined by

$$\langle D^\alpha T, \phi \rangle = (-1)^{|\alpha|} \langle T, D^\alpha \phi \rangle \ \forall \phi \epsilon \mathscr{D}(\Omega).$$

If $T_n, T \epsilon \mathscr{D}'(\Omega)$ then we say $T_n \to T$ in $\mathscr{D}'(\Omega)$ if

$$\langle T_n, \phi \rangle \to \langle T, \phi \rangle \quad \textit{for all } \phi \epsilon \mathscr{D}(\Omega).$$

The derivative mapping $D^\alpha : \mathscr{D}' \to \mathscr{D}'$ is continuous since if $T_n \to T$ in $\mathscr{D}'$ then

$$\begin{aligned} \langle D^\alpha T_n, \phi \rangle &= (-1)^{|\alpha|} \langle T_n, D^\alpha \phi \rangle \\ &\to (-1)^{|\alpha|} \langle T, D^\alpha \phi \rangle \\ &= \langle D^\alpha T, \phi \rangle \quad \textit{for all } \phi \epsilon \mathscr{D}(\Omega). \end{aligned}$$

1.3 Sobolev Space

The Sobolev space $H^1(\Omega)$ is defined by

$$H^1(\Omega) = \left\{ v \epsilon L^2(\Omega) : \frac{\partial v}{\partial x_i} \epsilon L^2(\Omega), \ 1 \leq i \leq n \right\}$$

where the derivatives are taken in the sense of distribution.

$$f \epsilon L^2(\Omega) \text{ need not imply } \frac{\partial f}{\partial x_i} \epsilon L^2(\Omega).$$

EXAMPLE 2. Let $\Omega = [-l, l]$

$$f(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x \geq 0. \end{cases}$$

Then $f \epsilon L^2[-l, l]$; but $df/dx = \delta$ is not given by a locally integrable function and hence not by an L^2 function.

We define an inner product $(\cdot\,,\cdot\,)_1$ in $H^1(\Omega)$ as follows: **4**

$$(u, v)_1 = (u, v) + \sum_{i=1}^{n} \left(\frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_i} \right) \quad \text{for all } u, v \epsilon H^1(\Omega).$$

Let $\|\cdot\|_1$ be the norm associated with this inner product. Then

LEMMA 1. *$H^1(\Omega)$ with $\|\cdot\|_1$ is a Hilbert spaces.*

Proof. Let u_j be a Cauchy sequence in $H^1(\Omega)$. This imply

$$\{u_j\}, \left\{ \frac{\partial u_j}{\partial x_i} \right\} i = 1, 2, \ldots, n$$

are Cauchy in L^2. Hence there exists $v, v_i \epsilon L^2(\Omega) 1 \le i \le n$ such that

$$u_j \to v \text{ in } L^2(\Omega),$$
$$\frac{\partial u_j}{\partial x_i} \to v_i \text{ in } L^2(\Omega), 1 \le i \le n,$$

For any $\phi \epsilon \mathscr{D}(\Omega)$,

$$\left\langle \frac{\partial u_j}{\partial x_i}, \phi \right\rangle = -\left\langle u_j, \frac{\partial \phi}{\partial x_i} \right\rangle \to -\left\langle u, \frac{\partial \phi}{\partial x_i} \right\rangle = \left\langle \frac{\partial u}{\partial x_i}, \phi \right\rangle.$$

But

$$\left\langle < \frac{\partial u_j}{\partial x_i}, \phi \right\rangle \to \langle v_i, \phi \rangle.$$

Hence

$$v_i = \frac{\partial u}{\partial x_i}.$$

Thus

$$u_j \to u \quad \text{in} \quad L^2(\Omega)$$
$$\frac{\partial u_j}{\partial x_i} \to \frac{\partial u}{\partial x_i} \quad \text{in} \quad L^2(\Omega).$$

5 This proves $u_j \to u$ in $H^1(\Omega)$. □

1.4 Negative Properties of $H^1(\Omega)$

(a) The functions in $H^1(\Omega)$ need not be continuous except in the case $n = 1$.

EXAMPLE 3. Let

$$\begin{aligned} \Omega &= \left\{(x, y)\epsilon\mathbb{R}^2 : x^2 + y^2 < r_\circ^2\right\}, r_\circ < 1. \\ f(r) &= (\log 1/r)^k, k < 1/2 \quad \text{where} \\ r &= (x^2 + y^2)^{1/2}. \end{aligned}$$

Then $f\epsilon H^1(\Omega)$ but f is not continuous at the origin.

In the case $n = 1$, if $u\epsilon H^1(\Omega), \Omega \subset \mathbb{R}^1$ then u can be shown to be continuous using the formula

$$u(y) - u(x) = \int_x^y \frac{du}{dx}(s)\, ds$$

where du/ds denotes the distributional derivative of u.

(b) $\mathscr{D}(\Omega)$ is not dense in $H^1(\Omega)$. To see this let $u\epsilon(\mathscr{D}(\Omega))^\perp$ in H^1 and $\phi\epsilon\mathscr{D}(\Omega)$. We have

$$(u, \phi)_1 = (u, \phi) + \sum_{i=1}^n \left(\frac{\partial u}{\partial x_i}, \frac{\partial \phi}{\partial x_i}\right) = 0$$

$$i.e. \qquad \langle u, \phi\rangle + \sum_{i=1}^n \left\langle -\frac{\partial^2 u}{\partial x_i^2}, \phi\right\rangle = 0$$

Thus

$$\langle -\Delta u + u, \phi\rangle = 0 \quad \text{for all} \quad \phi\epsilon\mathscr{D}(\Omega).$$

Hence

$$-\Delta u + u = 0 \quad \text{in} \quad \mathscr{D}'(\Omega).$$

Let $\Omega = \{x \epsilon \mathbb{R}^n : |x| < 1\}$, 6

$$u(x) = e^{r.x} \quad \text{where} \quad r \epsilon \mathbb{R}^n,$$
$$\Delta u(x) = |r|^2 e^{r.x} = |r|^2 u.$$
$$= u \quad \text{if} \quad |r| = 1.$$

Thus when $n = 1, u$ with $r = \pm 1$ belongs to $\mathscr{D}(\Omega))^{\perp}$, when $n > 1$ there are infinitely many $r' s(r \epsilon S^{n-1})$ such that $u \epsilon (\mathscr{D}(\Omega))^{\perp}$. Moreover these functions for different r's are linearly independent. Therefore

$$\text{dimension} \quad (\mathscr{D}(\Omega))^{\perp} \geq 2 \quad \text{if} \quad n = 1$$
$$\text{dimension} \quad (\mathscr{D}(\Omega))^{\perp} = \infty \quad \text{if} \quad n > 1.$$

This proves the claim (b).

We shall define $H^1_\circ(\Omega)$ as the closure of $\mathscr{D}(\Omega)$ in $H^1(\Omega)$. We have the following inclusions

$$\mathscr{D}(\Omega) \underset{dense}{\subset} H^1_\circ(\Omega) \subset H^1(\Omega) \underset{dense}{\subset} L^2(\Omega).$$

1.5 Trace Theorem

Let Ω be a bounded open subset of $\mathbb{R}^n$ with a Lipschitz continuous boundary Γ : *i.e.* there exists finite number of local charts $a_j, 1 \leq j \leq J$ from $\{y' \epsilon \mathbb{R}^{n-1} : |y'| < \alpha\}$ into $\mathbb{R}^n$ and a number $\beta > 0$ such that

$$\Gamma = \bigcup_{j=1}^{J} \left\{(y', y_n) : y_n = a_j(y'), |y'| < \alpha\right\},$$
$$\left\{(y', y_n) : a_j(y') < y_n < a_j(y') + \beta, |y'| < \alpha\right\} \subset \Omega, 1 \leq j \leq J,$$
$$\left\{(y', y_n) : a_j(y') - \beta < y_n < a_j(y'), |y'| < \alpha\right\} \subset C\overline{\Omega}, 1 \leq j \leq J.$$

7 It can be proved that $C^\infty(\overline{\Omega})$ is dense in $H^1(\Omega)$. If $f \epsilon C^\infty(\overline{\Omega})$ we define the trace of f, namely γf, by

$$\gamma f = f|_\Gamma. \quad \text{Note} \quad \gamma f \epsilon L^2(\Gamma) \quad \text{if } f \epsilon C^\infty(\overline{\Omega})$$

$\gamma : C^\infty(\overline{\Omega}) \to L^2(\Gamma)$ is continuous and linear with norm $\| \gamma u \|_{L^2(\Gamma)} \leq C \| u \|_1$. Hence this can be extended as continuous linear map from $H^1(\Omega)$ to $L^2(\Gamma)$.

$H^1_\circ(\Omega)$ is characterised by

THEOREM 2.

$$H^1_\circ(\Omega) = \{v \epsilon H^1(\Omega) : \gamma V = 0\}$$

1.6 Dual Spaces of $H^1(\Omega)$ and $H^1_\circ(\Omega)$

The mapping

$$I : H^1(\Omega) \to (L^2(\Omega))^{n+1} \quad \text{defined by}$$

$$I(v) = \left(v, \frac{\partial v}{\partial x_1}, \ldots, \frac{\partial v}{\partial x_n}\right)$$

is easily seen to be an isometric isomorphism of $H^1(\Omega))$ into subspace of $(L^2(\Omega))^{n+1}$. If $f \epsilon (H^1(\Omega))'$ then $F : I(H^1(\Omega)) \to \mathbb{R}$ with $F(Iu) = f(u)$ is a continuous linear functional on $I(H^1(\Omega))$. Hence by Hahn Banach theorem F can be extended to $(L^2(\Omega))^{n+1}$. Therefore, there exists $(v, v_1, \ldots, v_n) \epsilon (L^2(\Omega))^{n+1}$ such that

$$f(u) = F(Iu) = (v, u) + \sum_{i=1}^{n} (v_i, \partial u / \partial x_i).$$

This representation is not unique since F cannot be extended uniquely to $(L^2(\Omega))^{n+1}$ For all $\phi \epsilon \mathscr{D}(\Omega)$ we have

$$f(\phi) = \langle v, u \rangle - \sum_{i=1}^{n} \left\langle \frac{\partial v_i}{\partial x_i}, \phi \right\rangle,$$

Thus **8**

$$f|_{\mathscr{D}(\Omega)} = v - \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i}.$$

Conversely if $T \epsilon \mathscr{D}'(\Omega)$ is given by

$$T = v - \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i},$$

where $v, v_i \epsilon L^2(\Omega), 1 \leq i \leq n$ then T can be extended as a continuous linear functional on $H^1(\Omega)$ by the prescription

$$T(u) = (v, u) + \sum_{i=1}^{n} \left(v_i, \frac{\partial u}{\partial x_i} \right) \quad \text{for all } u \epsilon H^1(\Omega).$$

The extension of T to $H^1(\Omega)$ need not be unique. But we will prove that the extension of T to $H^1_\circ(\Omega)$ is unique. Let $\tilde{T} \epsilon (H^1_\circ(\Omega))'$ be such that $\tilde{T}|_{\mathscr{D}(\Omega)} = T$.

Let $u \epsilon H^1_\circ(\Omega)$. Then there exists $u_m \epsilon \mathscr{D}(\Omega)$ such that $u_m \to u$ in H^1. Now

$$\begin{aligned}
\tilde{T}(u) &= \tilde{T}(\lim_{\text{in } H^1} u_m) = \lim_m \tilde{T}(u_m) \\
&= \lim_m T(u_m) \\
&= \lim_m \left[(v, u_m) + \sum_{i=1}^{n} \left(v_i, \frac{\partial u_m}{\partial x_i} \right) \right] \\
&= (v, u) + \sum_{i=1}^{n} \left(v_i, \frac{\partial u}{\partial x_i} \right)
\end{aligned}$$

Thus

$$\tilde{T} = T \text{ on } H^1_\circ(\Omega).$$

9 Hence we identify $(H^1_\circ(\Omega))'$ with a space of distribution and we denote it by $H^{-1}(\Omega)$. That is

$$H^{-1}(\Omega) = \left\{ v - \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i} : (v, v_1, \ldots, v_n) \epsilon (L^2(\Omega))^{n+1} \right\} \subset \mathscr{D}'(\Omega).$$

EXERCISE 1. Show that $\partial/\partial x_i : L^2(\Omega) \to H^{-1}(\Omega)$ is continuous.

1.7 Positive Properties of $H^1_\circ(\Omega)$ and $H^1(\Omega)$.

THEOREM 3. (Poincare's Inequality). *Let Ω be an open bounded subset of $\mathbb{R}^n$. Then there exists a constant $C(\Omega)$ such that*

$$\int_\Omega v^2\, dx \leq C(\Omega) \int_\Omega |\nabla v|^2\, dx \quad \textit{for all } v \epsilon H^1_\circ(\Omega).$$

Proof. We shall prove the inequality for the functions in $\mathscr{D}(\Omega)$ and use the density of $\mathscr{D}(\Omega)$ in $H^1_\circ(\Omega)$.

Since Ω is bounded, we have

$$\Omega \subset [a_1, b_1] \times \ldots \times [a_n, b_n].$$

For any $u(x) \epsilon \mathscr{D}(\Omega)$, we have

$$u(x) = \int_{a_i}^{x_i} \frac{\partial u}{\partial x_i}(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n)\, dt.$$

Thus

$$|u(x)| \leq (b_i - a_i)^{1/2} \left(\int_{a_i}^{b_i} \left| \frac{\partial u}{\partial x_i} \right|^2 dt \right)^{1/2}$$

Squaring both sides and integrating we obtain

$$\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} |u(x)|^2\, dx \leq (b_i - a_i)^2 \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \left| \frac{\partial u}{\partial x_i} \right|^2 dx.$$

Thus **10**

$$\int_\Omega |u(x)|^2\, dx \leq C(\Omega) \int_\Omega |\nabla u|^2\, dx$$

where

$$C(\Omega) = (b_1 - a_1)^2 + \cdots + (b_n - a_n)^2.$$

Let $v \epsilon H^1_\circ(\Omega)$. Then there exists $u_n \epsilon \mathscr{D}(\Omega)$ such that $u_n \to v$ in H^1, which implies $u_n \to v$ in L^2 and $\dfrac{\partial u_n}{\partial x_i} \to \dfrac{\partial v}{\partial x_i}$ in L^2. Using this and the inequality for smooth functions we arrive at the result. □

REMARK 1. *The theorem is not true for functions in $H^1(\Omega)$. For example a nonzero constant function belongs to $H^1(\Omega)$ but does not satisfy the above inequality.*

We state Sobolev imbedding theorem and Rellich's theorem, which have many important applications.

SOBOLEV IMBEDDING THEOREM 4. *If Ω is an open bounded set having a Lipschitz continuous boundary then we have the imbedding*

$$H^1(\Omega) \hookrightarrow L^p(\Omega),$$

$p < q$ *or* $p \leq q$ *according as* $n = 2$ *or* $n > 2$ *where*

$$1/q = 1/2 - 1/n.$$

RELLICH'S THEOREM 5. *The above imbedding is compact for* $p < q$.

Chapter 2

Abstract Variational Problems and Examples

IN SECTION 1 OF this chapter, we give a variational formulation of the Dirichlet problem. In section 2 we prove the existence and uniqueness results for the abstract variational problem. In the remaining sections, we deal with the Neumann problem, Elasticity problem, Stokes problem and Mixed problem and their variational formulations. **11**

2.1 Dirichlet Problem

The Dirichlet problem is to find u such that

$$-\Delta u = f \quad \text{in} \quad \Omega \tag{2.1}$$

$$u = 0 \quad \text{on} \quad \Gamma \tag{2.2}$$

where $\Omega \subset \mathbb{R}^n$ is a bounded open set with smooth boundary Γ and f is a given function.

Multiplying equation (2.1) by a smooth function v which vanishes on Γ and integrating, we obtain

$$\int_{\Omega} -\Delta u.v\,dx = \int_{\Omega} f.v\,dx \tag{2.3}$$

Formally, using integration by parts and the fact that $v = 0$ on Γ, we see that

$$\int_{\Omega} \nabla v.\nabla u\,dx = \int_{\Gamma} \frac{\partial u}{\partial n} v\,d\Gamma - \int_{\Omega} \Delta u\ v\,dx = \int_{\Omega} -\Delta u.v\,dx \tag{2.4}$$

Equations (2.3) and (2.4) give

$$\int_{\Omega} \nabla u.v\,dx = \int_{\Omega} f\ v\,dx.$$

Now, setting

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

12 and

$$L(v) = \int_{\Omega} f v\,dx,$$

problem (2.1) can be formulated thus:
Find $u \varepsilon V$ such that

$$a(u, v) = L(v) \quad \text{for all } v \varepsilon V, \tag{2.5}$$

where V has to be chosen suitably.

Since $a(\cdot\,,\cdot)$ is symmetric, (2.5) is Euler's condition for the minimization problem

$$J(u) = \inf_{v\varepsilon V} J(v), \tag{2.6}$$

where

$$J(v) = 1/2a(v, v) - L(v).$$

If u is the solution of the minimization problem then it can be shown that

$$(J'(u), v) = 0 \quad \text{for all } v\varepsilon V, \tag{2.7}$$

where $(J'(u), v)$ is the Gateaux derivative of J in the direction v.

When $a(\cdot\,,\cdot)$ is symmetric one can show that problems (2.5) and (2.6) are equivalent.

To have $J(v)$ finite, we want our space V to be such that $\nabla v \,\varepsilon L^2(\Omega)$, $f \varepsilon L^2(\Omega)$ for all $v \varepsilon V$. The largest space satisfying the above conditions and (2.2) is $H^1_\circ(\Omega)$ and hence we choose V to be $H^1_\circ(\Omega)$.

If u is a solution of (2.5) then

$$\int\limits_{\Omega} \nabla u . \nabla v \, dx = \int\limits_{\Omega} f v \, dx \quad \text{for all } v \varepsilon \mathscr{D}(\Omega) \subset H^1_\circ(\Omega).$$

This implies that 13

$$\langle -\Delta u, v \rangle = \langle f, v \rangle \quad \text{for all } v \varepsilon \mathscr{D}(\Omega);$$

so

$$-\Delta u = f \quad \text{in} \quad \mathscr{D}'. \tag{2.8}$$

Conversely, if $u \varepsilon H^1_\circ(\Omega)$ satisfies (2.8), retracing the above steps we obtain

$$a(u, v) = f(v) \quad \text{for all } v \varepsilon \mathscr{D}(\Omega). \tag{2.9}$$

Since $\mathscr{D}(\Omega)$ is dense in $H^1_\circ(\Omega)$, (2.9) holds for all $v \varepsilon V = H^1_\circ(\Omega)$. Thus u is the solution of (2.5).

2.2 Abstract Variational Problem.

We now prove the existence and uniqueness theorem for the abstract variational problem.

THEOREM 1. *Let V be a Hilbert space and $a(\cdot\,,\cdot) : V \times V \to \mathbb{R}$ be continuous and bilinear. Further, assume that $a(\cdot\,,\cdot)$ is* **coercive:** *there exists $\alpha > 0$ such that $a(v, v) \geq \alpha \parallel v \parallel^2_V$ for all $v \varepsilon V$. Let L be a continuous linear functional on V. Then the problem:*

To find $u \varepsilon V$ such that

$$a(u, v) = L(v), \quad \textit{for all } v \varepsilon V \tag{2.10}$$

has a unique solution.

Proof. (i) **Uniqueness.** Let $u_1, u_2 \varepsilon V$ be two solutions of (2.10). Therefore

$$a(u_1, v) = L(v),$$
$$a(u_2, v) = L(v), \quad \text{for all } v\varepsilon V.$$

14 Subtracting one from the other, taking $v = u_2 - u_1$ and using V-coercivity of $a(\cdot\,,\cdot)$, we obtain

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

Thus $u_1 = u_2$.

(ii) **Existence when** $a(\cdot\,,\cdot)$ **is symmetric**. Since $a(\cdot\,,\cdot)$ is symmetric, the bilinear form $a(u, v)$ is a scalar product on V and the associated norm $a(v, v)^{1/2}$ is equivalent to the norm in V. Hence, by the Riesz representation theorem there exists $\sigma L \varepsilon V$ such that

$$a(\sigma L, v) = L(v) \quad \text{for all } v\varepsilon V.$$

Hence the theorem is true in the symmetric case.

(iii) **Existence in the general case.** Let $w\varepsilon V$. The function $L_w : V \to \mathbb{R}$ defined by

$$L_w(v) = (w, v) - \rho(a(w, v) - L(v))$$

is linear and continuous. Hence by the Riesz representation theorem there exists a $u\varepsilon V$ such that

$$L_w(v) = (u, v).$$

Let $T : V \to V$ be defined by

$$Tw = u$$

where u is the solution of the equation

$$L_w(v) = (u, v) \quad \text{for all } v\varepsilon V.$$

15 we will show that T is a contraction mapping. Hence T has a unique fixed point which will be the solution of (2.10).

Let

$$u_1 = Tw_1, u_2 = Tw_2.$$

Thus

$$(u_1 - u_2, v) = (w_1 - w_2, v) - \rho a(w_1 - w_2, v) \ \forall v \varepsilon V \tag{2.11}$$

Let $A : V \to V$, where Au is the unique solution of

$$(Au, v) = a(u, v) \quad \text{for all } v \varepsilon V,$$

which exists by the Riesz representation theorem.

$$\| Au \| = \sup_{v \varepsilon V} \frac{|(Au, v)|}{\| v \|} = \sup_{v \varepsilon V} \frac{|a(u, v)|}{\| v \|} \leq M \| u \|,$$

where $|a(u, v)| \leq M \| u \| \| v \|$. So A is continuous. Equation (2.11) can be written as

$$(u_1 - u_2, v) = (w_1 - w_2.v) - \rho(A(w_1 - w_2)v) \quad \text{for all } v \varepsilon V,$$

which implies that

$$u_1 - u_2 = (w_1 - w_2) - \rho A(w_1 - w_2).$$

So

$$\begin{aligned}
\| u_1 - u_2 \|^2 &= \| w_1 - w_2 \|^2 - 2\rho(A(w_1 - w_2), w_1 - w_2) \\
&\qquad + \rho^2 \| A(w_1 - w_2) \|^2 \\
&\leq \| w_1 - w_2 \|^2 - 2\rho a(w_1 - w_2, w_1 - w_2) \\
&\qquad + \rho^2 M^2 \| w_1 - w_2 \|^2, \quad \text{(using the continuity of } A) \\
&\leq \| w_1 - w_2 \|^2 - 2\rho\alpha \| w_1 - w_2 \|^2 + \rho^2 M^2 \| w_1 - w_2 \|^2,
\end{aligned}$$

since 16

$$a(w_1 - w_2, w_1 - w_2) \geq \alpha \| w_1 - w_2 \|^2 .$$

So
$$\| u_1 - u_2 \|^2 \leq (1 - 2\rho\alpha + \rho^2 M^2) \| w_1 - w_2 \|^2 .$$
That is,
$$\| Tw_1 - Tw_2 \| \leq \sqrt{(1 - 2\rho\alpha + \rho^2 M^2)} \| w_1 - w_2 \| .$$
Choosing ρ in $]0, \alpha/2M[$, we obtain that T is a contraction. ◻

This proves the theorem.

REMARK 1. *This theorem also gives an algorithm to find the solution of equation* (2.10)*. Let $u^\circ \varepsilon V$ be given. Let $u^{n+1} = Tu^n$. Then $u^n \to w_\circ$, which is the fixed point of T, and also the solution of* (2.10).

2.3 Neumann's problem.

Neumann's problem is to find an u such that
$$-\Delta u + cu = f \quad \text{in} \quad \Omega, \tag{2.12}$$
$$\frac{\partial u}{\partial n} = g \quad \text{on} \quad \Gamma. \tag{2.13}$$
We now do the calculations formally to find out the bilinear form a $(\cdot, \cdot)$, the linear functional $L(\cdot)$ and the space V.

For smooth v, (2.12) implies
$$\int\limits_\Omega (-\Delta u + cu) v \, dx = \int\limits_\Omega f v \, dx. \tag{2.14}$$
17 From Green's formula,
$$\int\limits_\Omega \nabla u . \nabla v \, dx = \int\limits_\Gamma \frac{\partial u}{\partial n} v \, d\Gamma - \int\limits_\Omega v \Delta u \, dx,$$
and by (2.14) we obtain
$$\int\limits_\Omega (\nabla u . \nabla v + cuv) \, dx = \int\limits_\Omega f v \, dx + \int\limits_\Gamma \frac{\partial u}{\partial n} v \, d\Gamma = \int\limits_\Omega f v \, dx + \int\limits_\Gamma g v \, d\Gamma,$$

since $\dfrac{\partial u}{\partial n} = g$ on Γ, by (2.13). This suggests the definitions:

$$a(u,v) = \int\limits_{\Omega} (\nabla u . \nabla v + cuv)\, dx \tag{2.15}$$

$$L(v) = \int\limits_{\Omega} fv\, dx + \int\limits_{\Gamma} gv\, d\Gamma, \tag{2.16}$$

$$V = H^1(\Omega), \tag{2.17}$$

where $f \varepsilon L^2(\Omega)$ and $g \varepsilon L^2(\Gamma)$.

Clearly $a(u,v)$ is bilinear, continuous and symmetric.

$$\begin{aligned} a(v,v) &= \int\limits_{\Omega} \left((\nabla v)^2 + cv^2\right)\, dx \\ &\geq \min\{1, c\} \parallel v \parallel_1^2, \end{aligned}$$

which shows $a(\cdot,\cdot)$ is $H^1(\Omega)$-coercive.

$L(\cdot)$ is a continuous linear functional on $H^1(\Omega)$. Hence by the theorem there exists a unique $u \varepsilon V = H^1(\Omega)$ such that

$$\int\limits_{\Omega} (\nabla u . \nabla v + cuv)\, dx = \int\limits_{\Omega} fv\, dx + \int\limits_{\Gamma} \quad \text{for all } v \varepsilon H^1(\Omega) \tag{2.18}$$

From (2.18) we obtain that for all $v \varepsilon \mathscr{D}(\Omega)$, **18**

$$\langle -\Delta u + cu, v \rangle = \langle f, v \rangle.$$

Hence

$$-\Delta u + cu = f \quad \text{in} \quad \mathscr{D}'(\Omega) \tag{2.19}$$

To find the boundary condition we use Green's formula:

$$\int\limits_{\Omega} \nabla u . \nabla v\, dx = \int\limits_{\Omega} -\Delta u . v\, dx + \int\limits_{\Gamma} \frac{\partial u}{\partial n} v\, d\Gamma,$$

which holds for all $u\varepsilon H^2(\Omega)$ and for all $v\varepsilon H^1(\Omega)$.

Assuming that our solution $u\varepsilon H^2(\Omega)$, from (2.19) we have

$$\int_\Omega (-\Delta u + cu)v = \int_\Omega fv \quad \text{for all} \quad v\varepsilon H^1(\Omega).$$

Using Green's formula we obtain

$$\int_\Omega (\nabla u \nabla v + cuv)\, dx = \int_\Gamma \frac{\partial u}{\partial n} v\, dx + \int_\Omega fv\, dx.$$

This, together with (2.18), implies

$$\int_\Gamma \left(g - \frac{\partial u}{\partial n}\right) v\, d\Gamma = 0 \quad \text{for all } v\varepsilon H^1(\Omega).$$

Hence we get the desired boundary condition

$$\frac{\partial u}{\partial n} = g \quad \text{on} \quad \Gamma.$$

If $u\varepsilon H^2(\Omega)$, these are still valid in "some sense" which is given in LIONS–MAGENES [29].

REMARK 2. *Even when $g = 0$ we cannot take the space*

$$V_1 = \left\{ v\varepsilon H^1(\Omega) : \frac{\partial v}{\partial n} = 0 \quad on \quad \Gamma \right\}$$

19 *to be the basic space V, since V_1 is not closed. In the Neumann problem 2.3, we obtain the boundary condition from Green's formula. In the case of Dirichlet problem 2.1, we impose the boundary condition in the space itself.*

REGULARITY THEOREM (FOR DIRICHLET PROBLEM) 2. *If Γ is C^2 or Ω is a convex polygon and $f\varepsilon L^2(\Omega)$, then the solution u of the Dirichlet problem* (2.1), (2.2) *belongs to $H^2(\Omega)$.*

REGULARITY THEOREM (FOR THE NEUMANN PROBLEM) 3. *If Γ is C^2 or Ω is a convex polygon, $f\varepsilon L^2(\Omega)$ and g belongs to a space finer than $L^2(\Omega)$ (for example $g\varepsilon H^1(\Gamma)$), then the solution u of the Neumann problem* (2.12), (2.13) *belongs to $H^2(\Omega)$.*

For a proof of these theorems the reader is referred to NECAS [33].

2.4 Mixed Problem.

In Sections 2.1 and 2.3 we found the variational formulation from the partial differential equation. In the general case it is difficult to formulate the variational problem from the p.d.e. In fact a general p.d.e. need not give rise to a variational problem. So in this section, we will take a general variational problem and find out the p.d.e. satisfied by its solution.

Let Ω be a bounded open set with boundary Γ. Let $\Gamma = \Gamma_\circ \cup \Gamma_1$ **20**
where $\Gamma_\circ$ and Γ_1 are disjoint. Let

$$V = \left\{ v \varepsilon H^1(\Omega) : v = 0 \quad \text{on} \quad \Gamma_\circ \right\}. \tag{2.20}$$

It is easy to see that V is closed and hence a Hilbert space with $\|\cdot\|_1$ norm

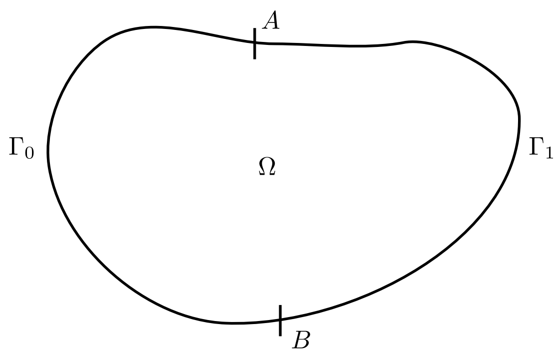

Figure 2.1:

We will use summation convention here afterwards. Let

$$a(u,v) = \int\limits_\Omega \left(a_{ij}(x) \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} + a_\circ \, uv \right) dx, \tag{2.21}$$

$$L(v) = \int\limits_\Omega f v \, dx + \int\limits_\Gamma g v \, d\Gamma, \tag{2.22}$$

where $a_\circ > 0, a_{ij}$ are smooth and there exists two constants $\alpha_\circ$ and α_1 such that

$$\alpha_1 \xi_i \xi_i \geq a_{ij}(x)\xi_i\xi_j \geq \alpha_\circ \xi_i \xi_i \quad \text{for all} x\varepsilon\Omega, \xi\varepsilon\mathbb{R}^n \tag{2.23}$$

i.e. the quadratic form $a_{ij}(x)\xi_i\xi_j$ is uniformly continuous and uniformly positive definite.

Inequality (2.23) implies that the bilinear form $a(\cdot\,,\cdot)$ is continuous and V-coercive. Formally we have

$$a(u,v) = \int_\Omega \left[-\frac{\partial}{\partial x_j}\left(a_{ij}\frac{\partial u}{\partial x_i}\right)v + a_\circ uv\right] dx + \int_\Gamma a_{ij}\frac{\partial u}{\partial x_i} n_j v \, d\Gamma \tag{2.24}$$

21 Let

$$\frac{\partial u}{\partial v_A} = a_{ij}\frac{\partial u}{\partial x_i} n_j,$$

and

$$Au = -\frac{\partial}{\partial x_j}\left(a_{ij}\frac{\partial u}{\partial x_i}\right).$$

If $v\varepsilon\mathscr{D}(\Omega)$ then the equation

$$a(u,v) = L(v) \tag{2.25}$$

becomes

$$\langle Au, v\rangle = \langle f, v\rangle.$$

Therefore

$$Au = f \quad \text{in} \quad \mathscr{D}'(\Omega). \tag{2.26}$$

Now for all $v\varepsilon V$, we have

$$\begin{aligned} a(u,v) &= \int_\Omega Au.v + \int_\Gamma \frac{\partial u}{\partial v_A} v \, d\Gamma \\ &= \int_\Omega Au.v + \int_{\Gamma_1} \frac{\partial u}{\partial v_A} v \, d\Gamma \\ L(v) &= \int_\Omega f v \, dx + \int_{\Gamma_1} g v \, d\Gamma. \end{aligned}$$

Equations (2.25) and (2.26) imply, for all $v \varepsilon V$,

$$\int_{\Gamma_1} \frac{\partial u}{\partial \nu_A} v \, d\Gamma = \int_{\Gamma_1} g \, v \, d\Gamma.$$

From this we obtain formally

$$\frac{\partial u}{\partial \nu_A} = g \quad \text{on} \quad \Gamma_1. \tag{2.27}$$

Thus the boundary value problem corresponding to the variational problem **22**

$$a(u, v) = L(v) \quad \text{for all } v \varepsilon V,$$

with $a(\cdot\,,\cdot\,), L(\cdot\,)$ and V given by the equations (2.20) - (2.22) is

$$\begin{aligned} Au &= f \quad \text{in} \quad \Omega, \\ \frac{\partial u}{\partial \nu_A} &= g \quad \text{on} \quad \Gamma_1, \\ u &= 0 \quad \text{on} \quad \Gamma_\circ. \end{aligned} \tag{2.28}$$

REMARK 3. *Even when f and g are smooth the solution u of the problem* (2.28) *may not be in $H^2(\Omega)$. In general, we will have a singularity at the transition points A, B on Γ. But if $\Gamma_\circ$ and Γ_1 make a corner then the solution u may be in $H^2(\Omega)$ provided that the boundary functions f, g satisfy some compatibility conditions. For regularity theorems the reader is referred to an article by PIERRE GIRSVARD [22].*

EXERCISE 1. Transmission Problem. Let $\Omega, \Omega_1, \Omega_2$ be open sets such that $\Omega = \Omega_1 \cup \Omega_2 \cup S$ where Ω_1 and Ω_2 are disjoint subsets of Ω and S is the interface between them. Let

$$a(u, v) = \sum_{i=1}^{2} \int_{\Omega_i} a_i \nabla u . \nabla v \, dx,$$

$$L(v) = \int_{\Omega} f v \, dx,$$

where $a_i > 0, i = 1, 2$, and $f \varepsilon L^2(\Omega)$. If u is the solution of the problem **23**

$$a(u, v) = L(v) \quad \text{for all } v \varepsilon H^1_\circ(\Omega),$$

and

$$u_i = u|_{\Omega_i}, f_i = f|_{\Omega_i}$$

then show that

$$\begin{aligned} -a_i \Delta u_i &= f_i \quad \text{on} \quad \Omega_i, i = 1, 2; \\ u_1 &= u_2 \quad \text{on} \quad S, \\ a_1 \frac{\partial u_1}{\partial n} &= a_2 \frac{\partial u_2}{\partial n} \quad \text{on} \quad S. \end{aligned}$$

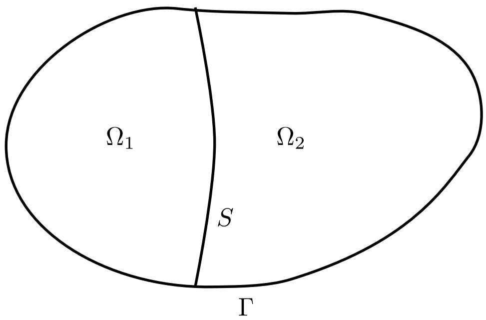

Figure 2.2:

EXERCISE 2. Fourier Condition. Let

$$\begin{aligned} V &= H^1(\Omega), \\ a(u, v) &= \int_\Omega \nabla u . \nabla v \, dx + \int_\Gamma uv \, d\Gamma, \\ L(v) &= \int_\Omega fv \, dx + \int_\Gamma gv \, d\Gamma, \end{aligned}$$

What is the boundary value problem associated with this ? Interpret the problem.

2.5 Elasticity Problem.

(a) **3-DIMENSIONAL CASE.** Let $\Omega \subset \mathbb{R}^3$ be a bounded, connected open set. Let Γ be the boundary of Ω and let Γ be split into two parts $\Gamma_\circ$ and Γ_1. Let Ω be occupied by an elastic medium, which we assume to be continuous. Let the elastic material be fixed along $\Gamma_\circ$. Let (f_i) be the body force acting in Ω and (g_i) be the pressure load acting along Γ_1. Let $(u_i(x))$ denote the displacement at x. **24**

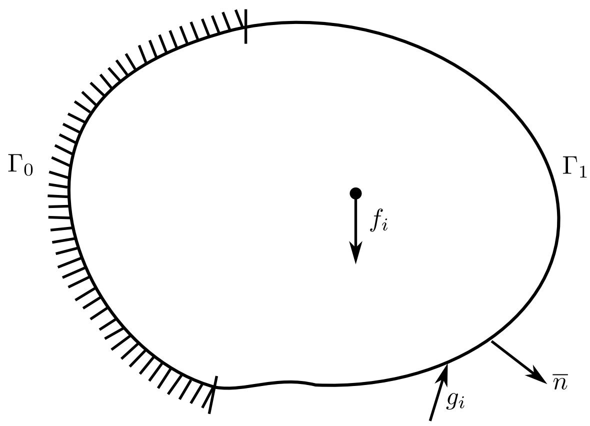

Figure 2.3:

In linear elasticity the stress-strain relation is

$$\sigma_{ij.}(u) = \lambda(div\ u)\delta_{ij} + 2\mu\varepsilon_{ij}(u), \varepsilon_{ij}(u) = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \quad (2.29)$$

where σ_{ij} and ε_{ij} denote the components of the stress and strain tensors respectively.

The problem is to find σ_{ij} and u_i, given (f_i) in Ω, (g_i) on Γ_1 and $(u_i) = 0$ on $\Gamma_\circ$.

The equations of equilibrium are

$$\frac{\partial}{\partial x_j}\sigma_{ij} + f_i = 0 \quad \text{in} \quad \Omega, \quad (2.30)$$

$$\sigma_{ij}n_j = g_i \quad \text{on} \quad \Gamma_1,\ i = 1, 2, 3 \tag{2.30b}$$

$$u_i = 0 \quad \text{on} \quad \Gamma_\circ. \tag{2.30c}$$

25 We have used the summation convention in the above equations. We choose

$$v = \left\{v\varepsilon(H^1(\Omega))^3 : v = 0 \quad \text{on} \quad \Gamma_\circ\right\}, \tag{2.31}$$

$$a(u, v) = \int\limits_\Omega \sigma_{ij}(u)\varepsilon_{ij}(v)\,dx, \tag{2.32}$$

$$L(v) = \int\limits_{\Gamma_1} g_i v_i\,d\Gamma + \int\limits_\Omega f_i v_i\,dx. \tag{2.33}$$

Using (2.29), $a(u, v)$ can be written as

$$a(u, v) = \int\limits_\Omega (\lambda \operatorname{div} u. \operatorname{div} v + 2\mu\varepsilon_{ij}(u)\varepsilon_{ij}(v))\,dx,$$

from which it is clear that $a(\cdot\,,\cdot\,)$ is symmetric. That $a(\cdot\,,\cdot\,)$ is V-elliptic is a nontrivial statement and the reader can refer to CIARLET [9]. Formal application of Green's formula will show that the boundary value problem corresponding to the variational problem (2.31) - (2.33) is (2.30).

$a(\cdot\,,\cdot\,)$ can be interpreted as the internal work and $L(\cdot\,)$ as the work of the external loads. Thus, the equation

$$a(u, v) = L(v) \quad \text{for all } v\varepsilon V$$

is a reformulation of the theorem of virtual work.

26 (b) **PLATE PROBLEM.** Let 2η be the thickness of the plate. By allowing $\eta \to 0$ in (a) we obtain the equations for the plate problem. It will be a two dimensional problem.

We have to find the bending moments M_{ij} and displacement (u_i). These two satisfy the equations:

$$M_{ij} = \alpha\Delta u\delta_{ij} + \beta\frac{\partial^2 u}{\partial x_i\,\partial x_j}, \tag{2.34}$$

$$\frac{\partial^2 M_{ij}}{\partial x_i\,\partial x_j} = f \quad \text{in} \quad \Omega, \tag{2.35}$$

$$u = 0 \quad \text{on} \quad \Gamma, \tag{2.36}$$

and

$$\frac{\partial u}{\partial n} = 0 \quad \text{if the plate is clamped,} \tag{2.37}$$

$$M_{ij} n_i n_j = 0 \quad \text{if the plate is simply supported} \tag{2.37a}$$

We take

$$V = \begin{cases} H^2_\circ(\Omega) = \left\{ v\varepsilon H^2 : V = \frac{\partial v}{\partial n} = 0 \quad \text{on} \quad \Gamma \right\}, \\ \qquad\qquad \text{if the plate is clamped;} \\ H^2(\Omega) \cap H^1_\circ(\Omega), \quad \text{if the plate is simply supported} \end{cases} \tag{2.38}$$

Formally, using Green's formula we obtain

$$\begin{aligned} \int\limits_\Omega \frac{\partial^2 M_{ij}}{\partial x_i \partial x_j} v\,dx &= -\int\limits_\Omega \frac{\partial M_{ij}}{\partial x_j}\frac{\partial v}{\partial x_i}\,dx + \int\limits_\Gamma \frac{\partial M_{ij}}{\partial x_j} v n_i\,d\Gamma \\ &= \int\limits_\Omega M_{ij}\frac{\partial^2 v}{\partial x_i \partial x_j}\,dx - \int\limits_\Gamma M_{ij} n_j \frac{\partial v}{\partial x_i}\,d\Gamma \\ &\quad + \int\limits_\Gamma \frac{\partial M_{ij}}{\partial x_j} v n_i\,d\Gamma \quad \text{for all } v\varepsilon V. \end{aligned} \tag{2.39}$$

But **27**

$$\int\limits_\Gamma \frac{\partial M_{ij}}{\partial x_j} v n_i\,d\Gamma = 0 \quad \text{for all } v\varepsilon V, \quad \text{since} \quad v = 0 \quad \text{on} \quad \Gamma,$$

and

$$\int\limits_\Gamma M_{ij} n_j \frac{\partial v}{\partial x_i}\,d\Gamma = \int\limits_\Gamma M_{ij} n_j \left(n_i \frac{\partial v}{\partial n} + s_i \frac{\partial v}{\partial s} \right) d\Gamma,$$

where $\partial v/\partial n$ denotes the normal derivative of v and $\partial v/\partial s$ denotes the tangential derivative. By (2.37) and (2.37a) we have

$$\int_\Gamma M_{ij} n_j \frac{\partial v}{\partial x_i} \, d\Gamma = 0.$$

Hence

$$\int_\Omega f v \, dx = \int_\Omega \frac{\partial^2 M_{ij}}{\partial x_i \partial x_j} v \, dx = \int_\Omega M_{ij} \frac{\partial^2 v}{\partial x_i \partial x_j} \, dx \quad \text{for all } v \varepsilon V.$$

We therefore choose

$$a(u, v) = \int_\Omega M_{ij} \frac{\partial^2 v}{\partial x_i \partial x_j} \, dx = \int_\Omega \left(\alpha \Delta u . \Delta v + \beta \frac{\partial^2 u}{\partial x_i \partial x_j} \frac{\partial^2 v}{\partial x_i \partial x_j} \right) dx \tag{2.39}$$

and

$$L(v) = \int_\Omega f v \, dx. \tag{2.40}$$

$a(\cdot, \cdot)$ can be proved to be V-coercive if $\beta \geq 0$ and $\alpha \geq 0$.

REGULARITY THEOREM 4. *When Ω is smooth and $f \varepsilon L_2(\Omega)$, then the solution u of the problem*

$$\begin{aligned} -\Delta u &= f \quad \textit{in} \quad \Omega, \\ u &= 0 \quad \textit{on} \quad \Gamma, \end{aligned}$$

28 *belongs to $H^2(\Omega)$. Moreover, we have*

$$\| u \|_2 \leq C \| f \|_\circ = C \| \Delta u \|_\circ,$$

where C is a constant.

This proves the coerciveness of $a(u, v)$ above for $\beta = 0$ and $\alpha > 0$.

2.6 Stokes Problem.

The motion of an incompressible, viscous fluid in a region Ω is governed by the equations

$$-\Delta u + \nabla p = f \quad \text{in} \quad \Omega, \tag{2.41}$$
$$\operatorname{div} u = 0 \quad \text{in} \quad \Omega, \tag{2.42}$$
$$u = 0 \quad \text{in} \quad \Gamma; \tag{2.43}$$

where $u = (u_i)_{i=1,\ldots,n}$ denotes the velocity of the fluid and p denotes the pressure. We have to solve for u and p, given f.

We impose the condition (2.42) in the space V itself. That is, we define

$$V = \{v\varepsilon(H^1_\circ(\Omega))^n : \operatorname{div} v = 0\} \tag{2.44}$$

Taking the scalar product on both sides of equation (2.41) with $v\varepsilon V$ and integrating, we obtain

$$\int\limits_\Omega f.v\,dx = \int\limits_\Omega \frac{\partial u_i}{\partial x_j}\frac{\partial v_i}{\partial x_j},$$

since

$$\begin{aligned}-\int\limits_\Omega v.\Delta u &= -\int\limits_\Omega v_j \frac{\partial^2 u_j}{\partial x_i \partial x_i}\\ &= \int\limits_\Omega \frac{\partial v_j}{\partial x_i}.\frac{\partial u_j}{\partial x_i} - \int\limits_\Gamma v_j \frac{\partial u_j}{\partial x_i} n_i,\end{aligned}$$

and 29

$$\int\limits_\Omega \nabla p.v = \int\limits_\Omega \frac{\partial p}{\partial x_i} v_i = -\int\limits_\Omega p\frac{\partial v_i}{\partial x_i} + \int\limits_\Gamma p v_i n_i = 0$$

as $v\varepsilon V$. Therefore we define

$$a(u,v) = \int\limits_\Omega \frac{\partial u_i}{\partial x_j}\frac{\partial v_i}{\partial x_j}\,dx \tag{2.45}$$

$$L(v) = \int_{\Omega} f.v\,dx. \tag{2.46}$$

We now have the technical lemma.

LEMMA 5. *The space*

$$\vartheta = \{v\varepsilon(\mathscr{D}(\Omega))^n : \operatorname{div} v = 0\}$$

is dense in V.

The proof of this Lemma can be found in LADYZHENSKAYA [27].

The equation $a(u, v) = L(v)$ for all $v\varepsilon V$ with $a(,), L(\ ), v$ defined by (2.44) - (2.46) is then equivalent to

$$\langle \Delta u + f, \phi \rangle = 0 \quad \text{for all } \phi\varepsilon\vartheta, \tag{2.47}$$

where $\langle , \rangle$ denotes the duality bracket between $(\mathscr{D}'(\Omega))^n$ and $(\mathscr{D}(\Omega))^n$. Notice that (2.46) is not valid for all $\phi\varepsilon(\mathscr{D}(\Omega))^n$ since $(\mathscr{D}(\Omega))^n$ is not contained in ϑ. To prove conversely that the solution of (2.46) satisfies (2.41), we need

THEOREM 6. *The annihilator $\vartheta^\perp$ of ϑ in $(\mathscr{D}'(\Omega))^n$ is given by $\vartheta^\perp = \{v :$ there exists a $p\varepsilon\mathscr{D}'(\Omega)$ such that $v = \nabla p\}$.*

30 Theorem 2.6 and Equation (2.46) imply that there exists a $p\varepsilon\mathscr{D}'(\Omega)$ such that

$$\Delta u + f = \Delta p.$$

Since

$$u\varepsilon(H^1(\Omega))^n \quad \text{and} \quad f\varepsilon(L^2(\Omega))^n, \Delta u + f\varepsilon(H^{-1}(\Omega))^n.$$

Therefore

$$\nabla p\varepsilon(H^{-1}(\Omega))^n.$$

We now state

THEOREM 7. *If $p\varepsilon\mathscr{D}'(\Omega)$ and $\nabla p\varepsilon(H^{-1}(\Omega))^n$, then $p\varepsilon L^2(\Omega)$ and*

$$\| p \|_{L^2(\Omega)/\mathbb{R}} \leq C \| \nabla p \|_{(H^{-1}(\Omega))^n}$$

where C is a constant.

From this Theorem we obtain that $p \varepsilon L^2(\Omega)$. Thus, if $f \varepsilon L^2(\Omega)$ and Ω is smooth, we have proved that the problem (2.44) - (2.46) has a solution $u \varepsilon V$ and $p \varepsilon L^2(\Omega)$.

Chapter 3

Conforming Finite Element Methods

IN CHAPTER 2 WE dealt with the abstract variational problems and 31
some examples. In all our examples the function space V is infinite dimensional. Our aim is to approximate V by means of finite dimensional subspaces V_h and study the problem in V_h. Solving the variational problem in V_h will correspond to solving some system of linear equations. In this Chapter we will study an error estimate, the construction of V_h and examples of finite elements.

3.1 Approximate Problem.

The abstract variational problem is:

$$\text{find } u\varepsilon V \text{ such that } a(u,v) = L(v). \quad \text{for all } v\varepsilon V, \tag{3.1}$$

where $a(\cdot\,,\cdot), L(\cdot), V$ are as in Chapter 2.

Let V_h be a finite dimensional subspace of V. Then the approximate problem corresponding to (3.1) is:

$$\text{find } u_h\varepsilon V_h \text{ such that } a(u_h,v) = L(v) \quad \text{for all } v\varepsilon V_h. \tag{3.2}$$

By the Lax-Milgram Lemma (Chapter 2, Theorem 2.1), (3.2) has a unique solution.

Let dimension $(V_h) = N(h)$ and let $(w_i)_{i=1,\ldots,N(h)}$ be a basis of V_h. Let

$$u_h = \sum_{i=1}^{N(h)} u_i w_i, v_h = \sum_{j=1}^{N(h)} v_j w_j,$$

32 where $u_i, v_j \varepsilon \mathbb{R}$, $1 \leq i, j \leq N(h)$. Substituting these in (3.2), we obtain

$$\sum_{i,j=1}^{N(h)} u_i v_j \ a(w_i, w_j) = \sum_{j=1}^{N(h)} L(w_j) v_j \tag{3.3}$$

Let

$$A^T = (a(w_i, w_j))_{i,j}, U = (u_i)_i, \ V = (v_i)_i, b = (L(w_i))_i$$

Then (3.3) can be written as

$$V^T A U = V^T b.$$

This is true for all $V \varepsilon \mathbb{R}^{N(h)}$. Hence

$$AU = b. \tag{3.4}$$

If the linear system (3.4) is solved, then we know the solution u_h of (3.2). This approximation method is called the Rayleigh-Galerkin method.

A is positive definite since

$$\begin{aligned} V^T A V = \sum_{i,j=1}^{N(h)} a(w_i, w_j) v_i \ v_j &= a\left(\sum_{i=1}^{N(h)} w_i v_i, \sum_{j=1}^{N(h)} w_j v_j\right) \\ &\geq \alpha \parallel \sum v_i w_i \parallel^2 \quad \text{for all } V \varepsilon \mathbb{R}^{N(h)}. \end{aligned}$$

A is symmetric if the bilinear form $a(\cdot\,,\cdot)$ is symmetric.

From the computational point of view it is desirable to have A as a
33 sparse matrix, i.e. A has many zero elements. Usually $a(\cdot\,,\cdot)$ will be given by an integral and the matrix A will be sparse if the support of the basis functions is "small". For example, if

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

then $a(w_i, w_j) = 0$ if Supp $w_i \cap$ Supp $w_j = \phi$.

Now we will prove a theorem regarding the error committed when the approximate solution u_h is taken instead of the exact solution u.

THEOREM 1. *If u and u_h denote the solutions of* (3.1) *and* (3.2) *respectively, then we have*

$$\| u - u_h \|_V \leq C \inf_{v_h \varepsilon V_h} \| u - v_h \|_V$$

Proof. We have

$$\begin{aligned} a(u, v) &= L(v) \quad \text{for all } v \varepsilon V, \\ a(u_h, v) &= L(v) \quad \text{for all } v \varepsilon V_h; \end{aligned}$$

so

$$a(u - u_h, v) = 0 \quad \text{for all } v \varepsilon V_h \tag{3.5}$$

By the V-coerciveness of $a(\cdot\,,\cdot)$ we obtain

$$\begin{aligned} \| u - u_h \|^2 &\leq 1/\alpha \; a(u - u_h, u - u_h) \\ &= 1/\alpha \; a(u - u_h, u - v + v - u_h), \quad \text{for all } v \varepsilon V_h \\ &= 1/\alpha \; a(u - u_h, u - v), \quad \text{by (3.5)} \\ &\leq M/\alpha \; \| u - u_h \| \, \| u - v \| \end{aligned}$$

□

This proves the theorem with $C = M/\alpha$. **34**

3.2 Internal Approximation of $H^1(\Omega)$.

Let $\Omega \subset \mathbb{R}^2$ be a polygonal domain. Let T_h be a triangulation of Ω: that is T_h a finite collection of triangles such that

$$\overline{\Omega} = \bigcup_{K \varepsilon T_h} \overline{K} \quad \text{and} \quad K \cap K' = \phi \quad \text{for} \quad K, K' \varepsilon T_h, K \neq K'.$$

Let $P(K)$ be a function space defined on K such that $P(K) \subset H^1(K)$. Usually we take $P(K)$ to be the space of polynomials of some degree. We have

THEOREM 2. *If*

$$V_h = \{v_h \varepsilon C^\circ(\Omega) : v_h|_K \varepsilon P(K), K \varepsilon T_h\}$$

where $P(K) \subset H^1(K)$*, then* $V_h \subset H^1(\Omega)$*.*

Proof. Let $u \varepsilon V_h$ and v_i be a function defined on Ω such that $v_i|_K = \dfrac{\partial}{\partial x_i}(u|_K)$. This makes sense since $u|_K \varepsilon H^1(K)$. Moreover $v_i \varepsilon L^2(\Omega)$, since $v_i|_K = \dfrac{\partial}{\partial x_i}(u|_K) \varepsilon L^2(K)$. We will show that $v_i = \dfrac{\partial u}{\partial x_i}$ in $\mathscr{D}'(\Omega)$.

For any $\phi \varepsilon \mathscr{D}(\Omega)$, we have

$$\begin{aligned}\langle v_i, \phi\rangle &= \int\limits_\Omega v_i \phi\, dx = \sum_{K \varepsilon T_h} \int\limits_K v_i \phi\, dx = \sum_{K \varepsilon T_h} \int\limits_K \frac{\partial}{\partial x_i}(u|_K)\phi\, dx \\ &= \sum_{K \varepsilon T_h} - \int\limits_K (u|_K) \frac{\partial \phi}{\partial x_i}\, dx + \int\limits_K (u|_K)\phi\, n_i^K\, d\Gamma,\end{aligned}$$

35 where n_i^K is the i^{th} component of the outward drawn normal to ∂K. So

$$\langle v_i, \phi\rangle = -\int\limits_\Omega u \frac{\partial \phi}{\partial x_i}\, dx + \sum_{K \varepsilon T_h} \int\limits_{\partial K} (u|_K)\phi n_i^K\, d\Gamma \tag{3.6}$$

The second term on the right hand side of (3.6) is zero since u is continuous in Ω and if K_1 and K_2 are two adjacent triangles then $n_i^{K_1} = -n_i^{K_2}$. Therefore

$$\langle v_i, \phi\rangle = -\int\limits_\Omega u . \frac{\partial \phi}{\partial x_i}\, dx = \left\langle \frac{\partial u}{\partial x_i}, \phi \right\rangle$$

which implies

$$v_i = \frac{\partial u}{\partial x_i} \quad \text{in} \quad \mathscr{D}'(\Omega).$$

Hence $u \varepsilon H^1(\Omega)$. Thus $V_h \subset H^1(\Omega)$.

We assume that the triangulation T_h is such that if $K_1, K_2 \varepsilon T_h$ are distinct, then either $\overline{K}_1 \cap \overline{K}_2$ is empty or equal to the common edge of the triangles K_1 and K_2. By this assumption we eliminate the possibility of a triangulation as shown in figure.

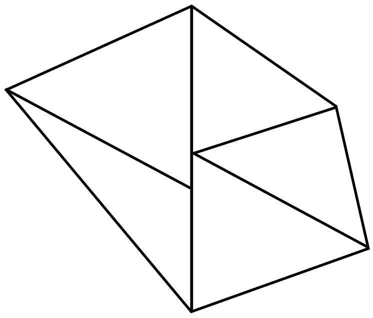

Figure 3.1:

Construction of V_h. 36

Let Ω be a polygonal domain and T_h be a triangulation of Ω, where

$$h = \max_{K \varepsilon T_h} \quad \text{(diameter of } K\text{)}.$$

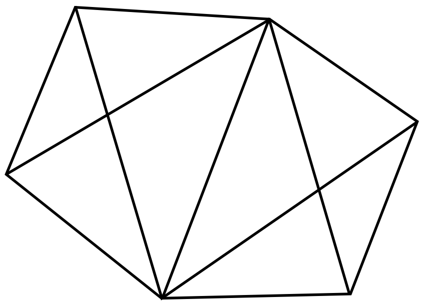

Figure 3.2:

Let

$$N(h) = \# \quad \text{nodes of the triangulation}, \tag{3.8}$$

$$P(K) = P_1(K) = \text{ polynomial of degree less than or equal to } 1 \text{ in } x \text{ and } y \tag{3.9}$$

Let

$$V_h = \{v_h : v_h|_K \varepsilon P_1(K), K \varepsilon T_h\} \tag{3.10}$$

We know that a polynomial of degree 1 in x and y is uniquely determined if its values on three non-collinear points are given. Using this we construct a basis for V_h. A function in V_h is uniquely determined if its value at all the nodes of the triangulation is given. Let the nodes of the triangulation be numbered $\{1, 2, \ldots, N(h)\}$. Let $W_i \varepsilon V_h$ be

$$w_i = \begin{cases} 1 & \text{at the } i^{th} \text{ node,} \\ 0 & \text{at other nodes.} \end{cases} \tag{3.11}$$

37 It is easy to see that w_i are linearly independent. If $v \varepsilon V_h$, then

$$v = \sum_{i=1}^{N(h)} v_i w_i \tag{3.12}$$

where v_i the value of v at the i^{th} node. This proves that $\{w_i\}_{1,\ldots,N(h)}$ is a basis of V_h and dimension of $V_h = N(h)$.

Moreover, $\operatorname{Supp} w_i \subset \cup \overline{K}$, where the union is taken over all the triangles whose one of the vertices is the i^{th} node.

Hence if i^{th} node and j^{th} node are not the vertices of a triangle K, for any $K \varepsilon T_h$, then

$$\operatorname{Supp} w_i \cap \operatorname{Supp} w_j = \phi.$$

We will show that V_h given by (3.10) is contained in $C^\circ(\overline{\Omega})$. Let $v \varepsilon V_h$ and let $K_1, K_2 \varepsilon T_h$ be adjacent triangles. Let 'ℓ' be the side common to both K_1 and K_2. $v|_{K_1}$ and $V|_{K_2}$ are polynomials of degree less than or equal to one in x and y. Let $\tilde{v}_1$ and $\tilde{v}_2$ be the extensions of $v|_{K_1}$ and $v|_{K_2}$ to $\overline{K}_1$ and $\overline{K}_2$ respectively. $\tilde{v}_1|_\ell$ and $\tilde{v}_2|_\ell$ can be thought of as a polynomial of degree less than or equal to one in a **single variable** and hence can be determined uniquely if their values at two distinct points
38 are known. But, by the definition of V_h in (3.10), $\tilde{v}_1|_\ell$ and $\tilde{v}_2|_\ell$ agree at the common vertices of K_1 and K_2. Hence $\tilde{v}_1|_\ell = \tilde{v}_2|_\ell$. This proves that v is continuous across K_1 and K_2. Thus $v \varepsilon C^\circ(\overline{\Omega})$. Hence $V_h \subset C^\circ(\overline{\Omega})$.

Using the theorem 3.2 we conclude that $V_h \subset H^1(\Omega)$. When we impose certain restrictions on T_h, it is possible to prove that $d(u, V_h) \to 0$ as $h \to 0$ where $d(u, V_h)$ is the distance between the solution u of (3.1) and the finite dimensional space V_h. The reader can refer to CIARLET [9]. Thus V_h "approximates" $H^1(\Omega)$.

The finite element method and the finite difference scheme are the "same" when the triangulation is uniform. For elliptic problems the finite element method gives better results than the finite difference scheme.

□

3.3 Finite Elements of Higher Degree.

DEFINITION. *Let K be a triangle with vertices $(a_i, i = 1, 2, 3)$. Let the coordinates of a_i be a_{ij}, $j = 1, 2$. For any $x \varepsilon \mathbb{R}$, the* barycentric coordinates $\lambda_i(x), i = 1, 2, 3$*, of x are defined to be the unique solution of the linear system*

$$\begin{aligned} \sum_{i=1}^{3} \lambda_i\, a_{ij} &= x_j,\ j = 1, 2; \\ \sum_{i=1}^{3} \lambda_i &= 1 \end{aligned} \tag{3.13}$$

Notice that the determinant of the coefficient matrix of the system (3.13) is twice the area of the triangle K. It is easy to see that the barycentric coordinates of a_1, a_2, a_3 are $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ respectively. The barycentric coordinate of the centroid G of K is $(1/3, 1/3, 1/3)$. 39

Using Cramer rule we find from (3.13) that

$$\lambda_1 = \frac{\begin{vmatrix} x_1 & a_{21} & a_{31} \\ x_2 & a_{22} & a_{32} \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ 1 & 1 & 1 \end{vmatrix}}$$

$$\text{i.e.} \qquad \lambda_1 = \frac{\text{area of the triangle} \quad xa_2a_3}{\text{area of the triangle} \quad a_1a_2a_3}$$

$$\text{Similarly, } \lambda_2 = \frac{\text{area of the triangle} \quad a_1xa_3}{\text{area of the triangle} \quad a_1a_2a_3}$$

$$\lambda_3 = \frac{\text{area of the triangle} \quad a_1a_2x}{\text{area of the triangle} \quad a_1a_2a_3}$$

This geometric interpretation of the barycentric coordinates will be helpful in specifying the barycentric coordinates of a point. For example, the equation of the side a_2a_3 in barycentric coordinates is $\lambda_1 = 0$.

40 **DEFINITION.** *A* **finite element** *is a triple* $(K, P_K, \sum_K)$*, where* K *is a polyhedron,* P_K *is polynomial space whose dimension is* m *and* $\sum_K$ *is a set of distributions, whose cardinality is* m*. Further* $\sum_K = \{L_i \varepsilon \mathscr{D}'; i = 1, 2, \ldots, m\}$ *is such that for given* $d_i \varepsilon \mathbb{R}$, $1 \leq i \leq m$*, the equations*

$$L_i(p) = d_i,\ 1 \leq i \leq m$$

have a unique solution $p \varepsilon P_K$*.*
The elements L_i *are called* degree of freedom of P.

EXAMPLE 1. (Finite Element of Degree 1). Let $K = a$ triangle,

$$P_K = P_1(K) = \text{Polynomials of degree} \leq 1.$$
$$= \text{Span} \quad \{1, x, y\}$$

$\dim P_K = 3$,
$\sum_K = \{\delta_{a_i} : a_i \text{ vertices},\ i = 1, 2, 3\}$,
where δ_{a_i} denotes the dirac mass at the point a_i. Then $(K, P_K, \sum_K)$ is a finite element.

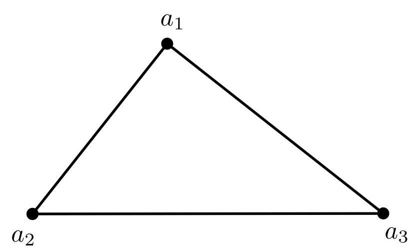

Figure 3.3:

This follows from the fact that $p\varepsilon P_1(K)$ is uniquely determined if its values at three non collinear points are given.

$$\delta_{a_i}(\lambda_j) = \lambda_j(a_i) = \delta_{ij}.$$

Hence λ_j, $j = 1, 2, 3$, form a basis for $P_1(K)$ and if $p\varepsilon P_1(K)$ then **41**

$$p = \sum_{i=1}^{3} p(a_i)\lambda_i$$

REMARK 1. *In the definition, dimension of P_K is m and we require that the equations*

$L_i(p) = d_i$, $1 \leq i \leq m$, *for given $d_i\varepsilon\mathbb{R}$ have a solution. So, in examples, to prove existence we have to prove only uniqueness. To prove the uniqueness it is enough to show that*

$$L_i(p) = 0, 1 \leq i \leq m, \quad \textit{implies} \quad p \equiv 0.$$

REMARK 2. *If $p_j\varepsilon P_K, 1 \leq j \leq m$, are such that*

$$L_i(p_j) = \delta_{ij}, 1 \leq i \leq m,\ 1 \leq j \leq m,$$

then $\{p_j\}$ form a basis for P_K and any $p\varepsilon P_K$ can be written as

$$p = \sum_{i=1}^{m} L_i(p)P_i.$$

EXAMPLE 2. (Finite Element of Degree 2). Let $K = a$ triangle,

$$P_K = P_2(K) = \quad \text{Span} \quad \{1, x, y, x^2, xy, y^2\},$$
$$\sum_K = \{\delta_{a_i}, 1 \leq i \leq 3, \delta_{a_{ij}} : 1 \leq i < j \leq 3\},$$

where a_i denote the vertices of K and a_{ij} denote the mid point of the side $a_i a_j$.

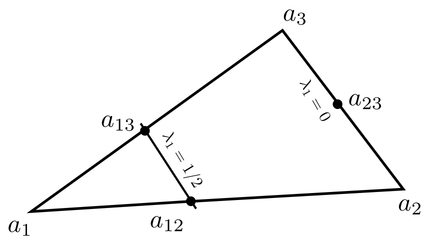

Figure 3.4:

42 The equations of the lines a_3a_2 and $a_{13}a_{12}$ are $\lambda_1 = 0$ and $\lambda_1 = 1/2$ respectively. Hence the function $\lambda_1(\lambda_1 - 1/2)$ vanishes at the points $a_2, a_3, a_{12}, a_{23}, a_{13}$. The value of $\lambda_1(\lambda_1 - 1/2)$ at a_1 is $1/2$. Hence $\lambda_1(2\lambda_1 - 1)$ takes the value 1 at a_1 and 0 at other nodes.

The equations of the lines a_1a_3 and a_2a_3 are $\lambda_2 = 0$ and $\lambda_1 = 0$ respectively. Therefore the function $\lambda_1\lambda_2$ vanishes at $a_1, a_2, a_{13}, a_{23}, a_3$ and takes the value $1/4$ at a_{12}. Thus $4\lambda_1\lambda_2$ is 1 at a_{12} and zero at the other nodes.

Thus any $p \varepsilon P_2(K)$ can be written in the form

$$p = \sum_{i=1}^{3} p(a_i)\lambda_i(2\lambda_i - 1) + \sum_{\substack{i<j \\ i,j=1}}^{3} 4p(a_{ij})\lambda_i\lambda_j.$$

EXAMPLE 3. (Finite Element of Degree 3). Let $K = a$ triangle,

$$P_K = P_3(K) = \quad \text{Span} \quad \{1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3\}.$$

Thus $\dim P_3 = 10$.

$$\sum_K = \{\delta_{a_i}, 1 \leq i \leq 3;\ \delta_{a_{iij}} 1 \leq i \leq j \leq 3, \delta_{a_{123}}\}.$$

where a_i denote the vertices of K and $a_{iij} = \dfrac{2}{3}a_i + \dfrac{1}{3}a_j$.

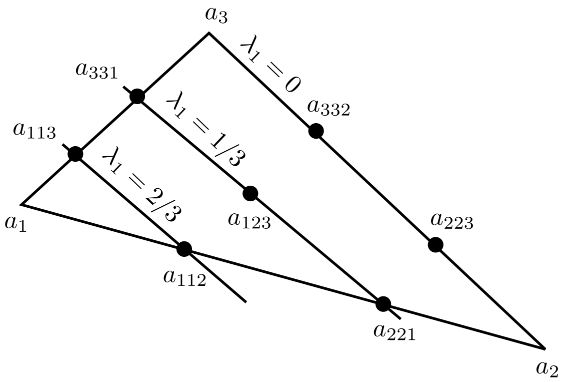

Figure 3.5:

It is easy to see that **43**

$$\begin{aligned} p_i &= 1/2\ \lambda_i(3\lambda_i - 1)\ (3\lambda_i - 2), \\ p_{iij} &= 9/2\ \lambda_i\lambda_j\ (3\lambda_i - 1), \\ p_{123} &= 27\ \lambda_1\lambda_2\lambda_3, \end{aligned}$$

$1 \leq i, j \leq 3$, is a basis of $P_3(K)$.
Moreover, p_i is 1 at the node a_i and zero at the other nodes; p_{iij} is 1 at the node a_{iij} and vanishes at the other nodes; p_{123} is zero at all nodes except a_{123} where its value is 1.

REMARK 3. *In the above three examples $\sum_K$ contains only Dirac masses and not derivatives of Dirac masses. All the above three finite elements are called* Lagrange finite elements.

Let Ω be a polygonal domain and let T_h be a triangulation of Ω, i.e. $\overline{\Omega} = \bigcup\limits_{K\varepsilon T_h} \overline{K}$. Let $P_K = P_\ell(K)$ consists of polynomials of degree $\leq \ell$. Let $(K, P_K, \sum_K)$ be a finite element for each $K\varepsilon T_h$. Let $\sum_h = \bigcup\limits_{K\varepsilon T_h} \sum_K$ and

$$V_h = \{v_h : v_h|_K \varepsilon P_K, K\varepsilon T_h\}$$

From the definition of finite element it follows that a function in V_h is uniquely determined by the distributions in $\sum_h$. **44**

REMARK 4. *In the example*

$\sum_h = \{\delta_{a_i} : a_i$ *is a vertex of a triangle in the triangulation*$\}$. *We proved in Sec. 3.2 that* $V_h \subset H^1(\Omega)$. *In this case we say that the finite element is* conforming.

REMARK 5. *The* V_h *so constructed above need not be contained in* $H^1(\Omega)$. *If* $V_h \not\subset H^1(\Omega)$ *we say that the finite element method is* non-conforming.

Let K *be a triangle,* $P_K = P_1(K)$ *and*

$$\sum_K = \{\delta_{a_{ij}} : 1 \leq i < j \leq 3\}.$$

Then $(K, P_K, \sum_K)$ *will be a finite element, but the space* $V_h \not\subset H^1(\Omega)$.

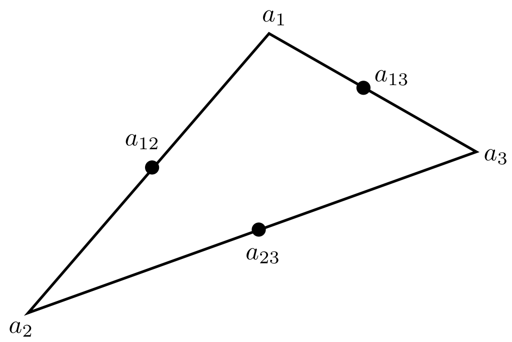

Figure 3.6:

45 When $(K, P_K, \sum_K)$ is as in Example 2, we will prove that $V_h \subset C^\circ(\overline{\Omega})$. This together with theorem 3.2 implies $V_h \subset H^1(\Omega)$. Thus the finite element in Example 2 is conforming.

To prove that $V_h \subset C^\circ(\overline{\Omega})$, let K_1 and K_2 be two adjacent triangles in the triangulation.

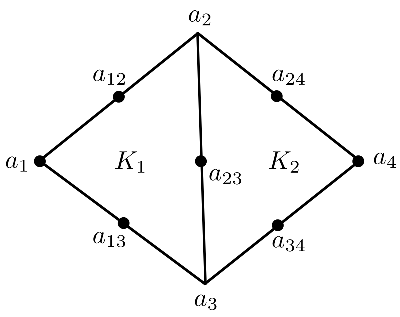

Figure 3.7:

A polynomial of degree 2 in x and y when restricted to a line in the plane is a polynomial of degree 2 in a single variable and hence can be determined on the line if the value of the polynomial at three distinct points on the line are known. Let $v_h \varepsilon V_h$. Let $\tilde{v}_1$ and $\tilde{v}_2$ be the continuous extension of $v_h|_{K_1}$ and $v_h|_{K_2}$ to $\overline{K}_1$ and $\overline{K}_2$ respectively; $\tilde{v}_1$ and $\tilde{v}_2$ are polynomials of degree 2 in one variable along the common side; $\tilde{v}_1$ and $\tilde{v}_2$ agree at the two common vertices and at the midpoint of the common side. Hence $\tilde{v}_1 = \tilde{v}_2$ on the common side. This shows v_h is continuous. Hence $V_h \subset C^\circ(\overline{\Omega})$.

Exercise 1. Taking $(K, P_K, \sum_K)$ as in Example 3, show that $V_h \subset H^1(\Omega)$.

3.4 Internal Approximation of $H^2(\Omega)$.

In this section we give an example of a finite element which is such that the associated space V_h is contained in $H^2(\Omega)$. This finite element can be used to solve some fourth order problems. We need **46**

THEOREM 3. *If $V_h = \{v_h : v_h|_K \varepsilon P_K \subset H^2(K)$, for all $K \varepsilon T_h\}$ is contained in $C^1(\overline{\Omega})$ then V_h is contained in $H^2(\Omega)$.*

The proof of this theorem is similar to that of theorem 2 of this section.

EXAMPLE 4. Let K be a triangle

$$P_K = P_3(K) = \quad \text{Span} \quad \{1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3\},$$
$$\sum_K = \left\{\delta_{a_i}, \frac{\partial}{\partial x}\delta_{a_i}, \frac{\partial}{\partial y}\delta_{a_i}, \delta_{a_{123}}, 1 \le i \le 3\right\}$$

where a_i are vertices of K, a_{123} is the centroid of K and

$$\dim P_K = \text{Card} \sum_K = 10.$$

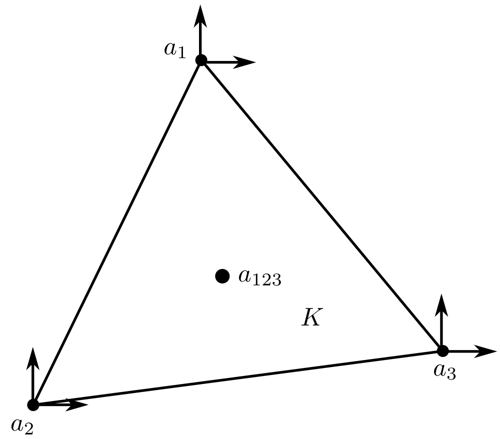

Figure 3.8:

The arrows in the figure denote that the values of the derivatives at the vertices are given. Using the formula

$$\begin{aligned} p = & \sum_{i=1}^{3} \left(-2\lambda_i^3 + 3\lambda_i^2 - 7\lambda_1\lambda_2\lambda_3\right) p(a_i) + 27\lambda_1\lambda_2\lambda_3 p(a_{123}) \\ & + \sum_{1 \le i < j \le 3} \lambda_i\lambda_j(2\lambda_i + \lambda_j - 1)Dp(a_i)(a_j - a_i) \quad \text{for all} \quad p\varepsilon P_3 \end{aligned}$$

47

where

$$Dp(a_i) = \left(\frac{\partial p}{\partial x}(a_i), \frac{\partial p}{\partial y}(a_i)\right).$$

We obtain that $p \equiv 0$ if

$$p(a_i) = p(a_{123}) = \frac{\partial p}{\partial x}(a_i) = \frac{\partial p}{\partial y}(a_i) = 0, 1 \leq i \leq 3.$$

Hence $(K, P_K, \sum_K)$ is a finite element.

The corresponding V_h is in $C^\circ(\overline{\Omega})$ but not in $C^1(\overline{\Omega})$. This shows that $V_h \not\subset H^2(\Omega)$. Hence this is not a conforming finite element for fourth order problems.

We now give an example of a finite element with $V_h \subset H^2(\Omega)$.

EXAMPLE 5. The Argyris Triangle. The Argyris triangle has 21 degrees of freedom. Here the values of the polynomial, its first and second derivatives are specified at the vertices; the normal derivative is given at the mid points.

In the figure we denote the derivatives by circles and normal derivative by a straight lines.

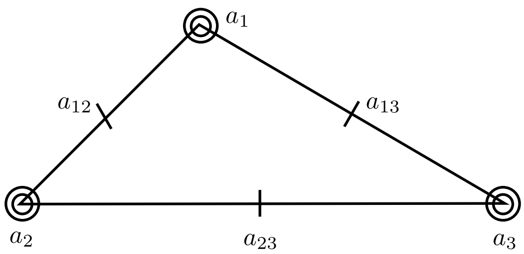

Figure 3.9:

48

We take $P_K = P_5$ = Space of polynomials of degree less than or

equal to5.

$$\dim P_K = 21;$$

$$\sum_K = \left\{\delta_{a_i}, \frac{\partial}{\partial x}\delta_{a_i}, \frac{\partial}{\partial y}\delta_{a_i}, \frac{\partial^2}{\partial x^2}\delta_{a_i},\right.$$

$$\left.\frac{\partial^2}{\partial x\,\partial y}\delta_{a_i}, \frac{\partial^2}{\partial y^2}\delta_{a_i}, 1 \leq i \leq 3, \frac{\partial}{\partial n}\delta_{a_{ij}} 1 \leq i < j \leq 3\right\},$$

where a_i denote the vertices of K, a_{ij} the midpoint of the line joining a_i and a_j, and $\partial/\partial n$, the normal derivative.

Let $p\varepsilon P_K$ be such that $L(p) = 0, L\varepsilon \sum_K$ We will prove that $p = 0$ in K. p is a polynomial of degree 5 in one variable along the side a_1a_2. By assumption p, and its first and second derivatives vanish at a_1 and a_2. Hence $p = 0$ along a_1a_2. Consider $\partial p/\partial n$ along a_1a_2. By assumption $\partial p/\partial n$ vanishes at a_1a_2 and a_{12}. Since the second derivatives of p vanish at a_1, a_2 we have the first derivatives of $\partial p/\partial n$ vanish at a_1 and a_2; $\partial p/\partial n$ is polynomial of degree 4 in one variable along a_1a_2. Hence $\partial p/\partial n = 0$ along a_1a_2. Since $p = 0$ along a_1a_2, $\partial p/\partial\tau = 0$ along a_1a_2, where $\partial/\partial\tau$ denote the tangential derivative, i.e. derivative along a_1a_2. Therefore we
49 have p and its first derivatives zero along a_1a_2.

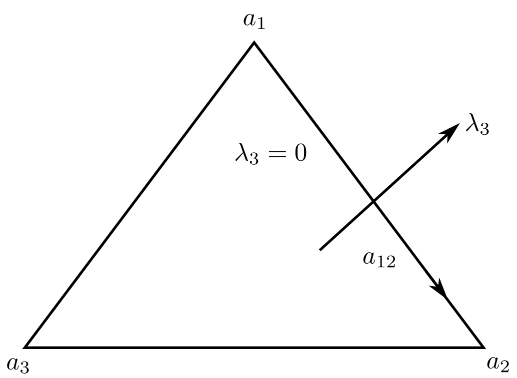

Figure 3.10:

The equation of a_1a_2 is $\lambda_3 = 0$. Hence we can choose a line perpendicular to a_1a_2 as the λ_3 axis. Let τ denote the variable along a_1a_2. Changing the coordinates from (x, y) to (y_3, τ) we can write the polynomial p as

$$p = \sum_{i=0}^{5} \lambda_3^i q_i(\tau)$$

where $q_i(\tau)$ is a polynomial in τ of degree $\leq 5 - i$. Now

$$\frac{\partial p}{\partial \lambda_3} = \sum_{i=1}^{5} i\lambda_3^{i-1} q_i(\tau).$$

Since p and its first derivatives vanish along $a_1 a_2$ we have

$$0 = p(0,\tau) = q_\circ(\tau),$$
$$0 = \frac{\partial p}{\partial \lambda_3}(0,\tau) = q_1(\tau)$$

Hence

$$p = \lambda_3^2 \sum_{i=2}^{5} \lambda_3^{i-2} q_i(\tau).$$

Thus λ_3^2 is a factor of p. By taking the other sides we can prove that λ_1^2 and λ_2^2 are also factors of p. Thus

$$p = c\lambda_1^2 \lambda_2^2 \lambda_3^2.$$

But $\lambda_1^2 \lambda_2^2 \lambda_3^2$ is a polynomial of degree 6 which does not vanish identi- **50**
cally in K and p is a polynomial of degree 5. Hence $c = 0$. Therefore $p \equiv 0$ in K. Thus $(K, P_K, \sum_K)$ is a finite element with 21 degrees of freedom.

THEOREM 4. *If V_h is the space associated with the Argyris finite element, then $V_h \subset C^1(\overline{\Omega})$.*

Proof. Let K_1, K_2 be two adjacent finite elements in the triangulation. Let $v \varepsilon V_h$ and let $p_1 = v|_{K_1}$, $p_2 = v|_{K_2}$. We denote the continuous extensions of p_1 and p_2 to $\overline{K}_1$ and $\overline{K}_2$ also by p_1 and p_2. We have to show that

$$p_1 = p_2, Dp_1 = Dp_2$$

along the common side Q of K_1 and K_2.

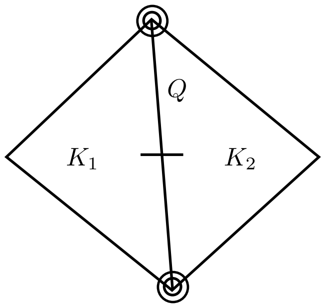

Figure 3.11:

Since $v \varepsilon V_h$, p_1, p_2 their first and second derivatives agree at the two
51 common vertices of K_1 and K_2. p_1 and p_2 are polynomials of degree 5 in one variable along the common side Q. Hence, along Q,

$$p_1 = p_2. \tag{3.14}$$

Therefore

$$\frac{\partial p_1}{\partial \tau} = \frac{\partial p_2}{\partial \tau} \quad \text{along} \quad Q. \tag{3.15}$$

The normal derivatives $\dfrac{\partial p_1}{\partial n}$ and $\dfrac{\partial p_2}{\partial n}$ are polynomials of degree 4 in one variable along Q. Since $v \varepsilon V_h$, $\dfrac{\partial p_1}{\partial n} = \dfrac{\partial p_2}{\partial n}$ at the two common vertices and at the midpoint of the common side Q. Moreover, the first derivatives of $\dfrac{\partial p_1}{\partial n}$ and $\dfrac{\partial p_2}{\partial n}$ coincide at the common vertices. Hence

$$\frac{\partial p_1}{\partial n} = \frac{\partial p_2}{\partial n} \quad \text{along} \quad Q. \tag{3.16}$$

Equations (3.14) - (3.16) show that

$$p_1 = p_2, Dp_1 = Dp_2 \quad \text{along} \quad Q.$$

$\square$

This proves that v and its first derivative are continuous across Q. Therefore $v \varepsilon C^1(\overline{\Omega})$. Hence $V_h \subset C^1(\overline{\Omega})$.

Theorems 3.3 and 3.4 imply that $V_h \subset H^2(\Omega)$. For other examples of finite element, the reader can refer to CIARLET [9].

Chapter 4

Computation of the Solution of the Approximate Problem

4.1 Introduction

THE SOLUTION OF THE APPROXIMATE PROBLEM. Find $u_h \varepsilon V_h$ 52
such that

$$a(u_h, v_h) = L(v_h)\ \forall v_h \varepsilon V_h, \tag{4.1}$$

can be found using either iterative methods or direct methods. We describe these methods in this chapter.

Let $\{w_i\}_{1 \leq i \leq N(h)}$ be a basis of V_h. Let $A = (a(w_i, w_j))$ and $b = (L(w_i))$. If $a(\cdot\,,\cdot)$ is symmetric, then (4.1) is equivalent to the minimization problem

$$J(u_h) = \min_{v_h \varepsilon V_h} J(v_h), \tag{4.2}$$

where $J(v) = \frac{1}{2} v^T A v - v^T b,\ \ v \varepsilon \mathbb{R}^{N(h)}$. Here we identify V_h and $\mathbb{R}^{N(h)}$ through the basis $\{w_i\}$ and the natural basis $\{e_i\}$ of $\mathbb{R}^{N(h)}$. u_h is a solution of (4.2) iff $Au_h = b$. Iterative methods are applicable only when $a(\cdot\,,\cdot)$ is symmetric.

4.2 Steepest Descent Method

Let $J : \mathbb{R}^N \to \mathbb{R}$ be differentiable.

In the steepest descent method, at each iteration we move along the direction of the negative gradient to a point such that the functional value of J is reduced. That is, let $x^{\bullet} \varepsilon \mathbb{R}^N$ be given. Knowing $x^n \varepsilon \mathbb{R}^N$, we define $x^{n+1} \varepsilon \mathbb{R}^N$ by

$$x^{n+1} = x^n - \lambda^n \ J'(x^n), \tag{4.3}$$

where λ^n minimizes the functional

$$\phi(\lambda) = J\left(x^n - \lambda J'(x^n)\right) \tag{4.4}$$

53

In the case

$$J(x) = 1/2 \ x^T A x - x^T b,$$

λ^n can be computed explicitly. It is easy to see that

$$J'(x) = Ax - b.$$

Since λ^n minimizes $\phi(\lambda)$, we have

$$\phi'(\lambda^n) = (J'(x^n - \lambda^n J'(x^n)), -J'(x^n)) = 0 \tag{4.5}$$

Let

$$r^n = J'(x^n) = Ax^n - b.$$

Then (4.5) implies

$$\lambda^n = \frac{(r^n, \ r^n)}{(Ar^n, r^n)}. \tag{4.6}$$

For proving an optimal error estimate for this scheme we need Kantorovich's inequality which is left as an exercise.

Exercise 1. (See LUENBERGER [30]).

Prove the Kantorovich's inequality

$$\frac{(Ax, x)\ (A^{-1}x, x)}{\| \ x \ \|^4} \leq \frac{(M + m)^2}{4mM} \tag{4.7}$$

where A is symmetric, positive definite matrix with

$$m = \operatorname*{Inf}_{x \neq 0} \frac{(Ax, x)}{\| x \|^2} > 0, M = \operatorname*{Sup}_{x \neq 0} \frac{(Ax, x)}{\| x \|^2}$$

THEOREM 1. *For any $x_\circ \varepsilon X$ the sequence $\{x_n\}$ defined by*

$$x_{n+1} = x_n + \frac{(r_n, r_n)}{(r_n, Ar_n)} r_n,$$

where **54**

$$r_n = b - Ax_n,$$

converges to the unique solution $\overline{x}$ of $Ax = b$. Furthermore, defining

$$E(x) = ((x - \overline{x}), A(x - \overline{x}))$$

we have the estimate

$$\| x_n - \overline{x} \|^2 \leq \frac{1}{m} E(x_n) \leq \frac{1}{m} \left(\frac{M - m}{M + m} \right)^{2n} E(x_\circ).$$

Proof. We have

$$\begin{aligned} E(x) &= (x - \overline{x}, A(x - \overline{x})) \\ &= 2J(x) + (\overline{x},\ A\overline{x}), \end{aligned}$$

where

$$J(x) = 1/2(x, Ax) - (b, x).$$

It is easy to see that

$$\frac{E(x_n) - E(x_{n+1})}{E(x_n)} = \frac{(r_n, r_n)^2}{(r_n, Ar_n)(r_n, A^{-1} r_n)} \geq \frac{4Mm}{(M + m)^2}$$

by Kantorovich inequality. Therefore

$$\frac{E(x_{n+1})}{E(x_n)} \leq \left(\frac{M - m}{M + m} \right)^2.$$

This implies

$$E(x_{n+1}) \leq \left(\frac{M-m}{M+m}\right)^{2(n+1)} E(x_\circ).$$

From the definition of m we obtain 55

$$\| x_n - \overline{x} \|^2 \leq \frac{1}{m} E(x_n) \leq \frac{1}{m} \left(\frac{M-m}{M+m}\right)^{2.n} E(x_\circ).$$

The condition number of A is defined by $\mathrm{Cond}(A) = \frac{M}{m}$. We have

$$\left(\frac{M-m}{M+m}\right)^2 \sim 1 - \frac{2m}{M} = 1 - \frac{2}{\mathrm{Cond}(A)}$$

If the condition number of A is smaller, then the convergence is faster. The steepest descent method is not a very good method for finite elements, since $\mathrm{Cond}(A) \sim C/h^2$ when $V_h \subset H^1(\Omega)$. □

4.3 Conjugate Gradient method

DEFINITION. *The directions $w_1, w_2 \varepsilon \mathbb{R}^N$ are said to be* conjugate *with respect to the matrix A if $w_1^T\ A\ w_2 = 0$.*

In the conjugate gradient method, we construct conjugate directions using the gradient of the functional. Then the functional is minimized by proceeding along the conjugate direction. We have

THEOREM 2. *Let $w^1, w^2, \ldots, w^N$ be N mutually conjugate directions. Let*

$$x^{k+1} = x^k - \lambda^k w^k$$

where λ^k minimizes

$$\phi(\lambda) = J(x^k - \lambda w^k),\ \lambda \varepsilon \mathbb{R}.$$

56 *When $x^1 \varepsilon \mathbb{R}^N$ is given, we have*

$$x^{N+1} = x^*$$

where

$$Ax^* = b.$$

Proof. Let
$$r^n = -J'(x^n) = b - Ax^n.$$
Since λ^k minimizes $\phi(\lambda)$ we have
$$\phi(\lambda^k) = (J'(x^k - \lambda^k w^k), -w^k) = 0.$$
This gives
$$\lambda^k = \frac{(r^k)^T w^k}{(w^k)^T A w^k} \tag{4.8}$$

Since $w^1, w^2, \ldots, w^N$ are mutually conjugate directions, they are linearly independent. Therefore there exist $\alpha_i, 1 \leq i \leq N$, such that
$$x^1 - x^* = \sum_{k=1}^{N} \alpha_k \, w^k.$$

From this, using the fact that w^j are mutually conjugate, we obtain
$$(x^1 - x^*)^T A w^j = \alpha_j (w^j)^T A w^j.$$
This gives
$$\alpha_j = \frac{(x^1 - x^*)^T \; A w^j}{(w^j)^T \; A w^j}. \tag{4.9}$$
Using induction we show that 57
$$\alpha_k = \lambda^k.$$
Since $Ax^* = b$, we have
$$r^1 = Ax^1 - b = A(x^1 - x^*).$$
This shows that
$$\alpha_1 = \lambda^1.$$

Let $\alpha_i = \lambda^i$ for $1 \leq i \leq k - 1$.
From the definition of x^k we obtain
$$x^k = x^1 - \sum_{i=1}^{k-1} \lambda^i w^i = x^1 - \sum_{i=1}^{k-1} \alpha_i \, w^i,$$

(by induction hypothesis). Since

$$(w^i)^T\ Aw^k = 0 \quad \text{for} \quad 1 \leq i \leq k-1,$$

we get

$$(x^k - x^1)^T\ Aw^k = 0$$

This together with (4.8) and (4.9) shows that

$$\alpha_k = \lambda^k$$

Thus $\alpha_k = \lambda^k$ for $1 \leq k \leq N$.

The definition of x^k implies

$$x^{N+1} = x^1 - \sum_{i=1}^{N} \lambda^i\ w^i = x^1 - \sum_{i=1}^{N} \alpha_i w^i = x^*$$

$\square$

Algorithm for Conjugate Gradient Method

58 **THEOREM 3.** *Let $x_\circ \varepsilon \mathbb{R}^N$. Define $w^1 = b - Ax^1$. Knowing x^n and w^{n-1} we define x^{n+1} and w^n by*

$$x^{n+1} = x^n + \alpha_n w^n$$
$$w^n = r^n + \beta_n w^{n-1},$$

where

$$r^n = b - Ax^n, \alpha_n = \frac{(r^n, w^n)}{(w^n, Aw^n)}, \beta_n = \frac{(r^n, r^n)}{(r^{n-1}, r^{n-1})}$$

Then w^n are mutually conjugate directions and x^{N+1} is the unique solution of $Ax = b$.

A proof of this theorem can be found in LUENBERGER [31]. It can be shown that

$$x^n - x^{N+1} \sim \left(\frac{1 - \sqrt{c}}{1 + \sqrt{c}}\right)^n,$$

where $c = m/M$. Thus the convergence rate in the conjugate gradient method is faster than in the steepest descent method, at least for quadratic functionals.

4.4 Computer Representation of a Triangulation

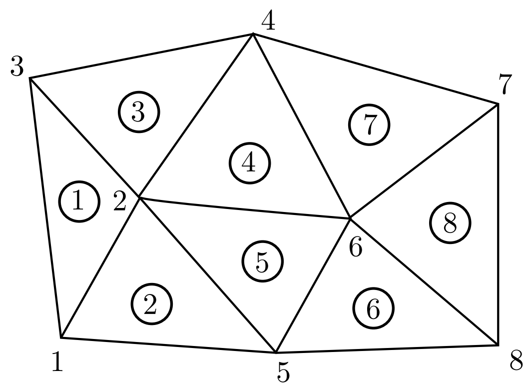

Figure 4.1:

Let T_h be a triangulation of the domain Ω. We number the nodes of the triangulation and the triangles in T_h. Let 59

$$
\begin{aligned}
NS &= \# \quad \text{of nodes of} \quad T_h, \\
NT &= \# \quad \text{of triangles in} \quad T_h.
\end{aligned}
$$

The triangulation is uniquely determined by the two matrices

$$Q(2,\ NS) = (q_{ij}) \quad \text{and} \quad ME(3,\ NS) = (m_{jk}),$$

where q_{ij} denotes the i^{th} coordinate of the j^{th} node and m_{jk} denotes the j^{th} vertex of the k^{th} triangle.

The matrix ME, corresponding to the triangulation in the above figure is

$$ME = \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 6 & 4 & 6 \\ 2 & 5 & 4 & 6 & 5 & 5 & 6 & 8 \\ 3 & 2 & 3 & 4 & 6 & 8 & 7 & 7 \end{pmatrix}$$

In some problems it is better to know the boundary nodes. The array $NG(NS)$ defined by

$$NG(i) = \begin{cases} 1 & \text{if} \quad i\varepsilon\Gamma \\ 0 & \text{otherwise} \end{cases}$$

is used for picking boundary nodes.

Exercise 2. Draw the triangulation

$$Q = \begin{pmatrix} 0 & 1 & 1 & 0 & 0.5 & 0.5 & 0.5 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0.5 & 1 & 0 & 0 & 0.5 \end{pmatrix}$$

$$ME = \begin{pmatrix} 2 & 2 & 6 & 3 & 4 & 4 & 1 & 5 \\ 7 & 5 & 5 & 6 & 6 & 9 & 7 & 7 \\ 5 & 8 & 8 & 8 & 9 & 5 & 9 & 9 \end{pmatrix}$$

60

4.5 Computation of the Gradient.

In the Neumann problem we have

$$(J'(u),\ v) = \int_\Omega (\nabla u.\nabla v + a_\circ uv) - \int_\Omega fv.$$

We now give a practical way of computing the gradient $J'(u)$.

Let w_i be the basis function in V_h which takes the value 1 at the i^{th} node and zero at the other nodes. Let V_h be defined by $\mathbb{P}_1$ Lagrange finite element. Let

$$u = (u_i)_{1 \le i \le NS}$$

be given. We want to compute $J'(u)$. Let $(J'(u))_i = (J'(u); w_i)$ and

$$B_i = \int_\Omega \nabla u.w_i\, dx = \sum_{K \varepsilon T_h} \int_K \nabla u.\nabla w_i\, dx.$$

We know that

$$w_i = \begin{cases} \lambda_j^k & \text{if} \quad i = m_{jk}, \quad \text{for some} \quad j \quad \text{and} \quad k \\ 0 & \text{otherwise,} \end{cases}$$

where λ_j^k is the j^{th} rycentric coordinate of the k^{th} triangle. Therefore

$$B_i = \sum_{1 \le j \le 3, 1 \le k \le NT, i = m_{jk}} a_j^k,$$

$$a_j^k = \int\limits_{K_k} \nabla u . \nabla \lambda_j^k \, dx, \quad \text{where} \quad K_k$$

is the triangle corresponding to the k^{th} element. **61**

Algorithm to Compute B.

Set $\quad B = 0$
for $\quad k = 1, 2, \ldots, NT;$
for $\quad j = 1, 2, 3,$
do $\quad B_{m_{jk}} = B_{m_{jk}} + a_j^k.$

Computation of the a_j^k's.

Since $u = (u_i)_{1 \le i \le NS}$ and we take $\mathbb{P}_1$ Lagrange finite elements, we have

$$\begin{aligned} u &= \sum_{j=1}^{3} \lambda_j^k u_{m_{jk}} \quad \text{in} \quad K_k, \\ &= ax + by + c, \quad \text{say.} \end{aligned}$$

Let $(\xi_i, \eta_i), 1 \le i \le 3$, be the coordinates of the vertices of K_k. Let $w_j = u_{m_{jk}}$. Then a, b are found from the equations

$$\begin{aligned} a\xi_1 + b\eta_1 + c &= w_1, \\ a\xi_2 + b\eta_2 + c &= w_2, \\ a\xi_3 + b\eta_3 + c &= w_3, \end{aligned} \tag{4.10}$$

as

$$a = \frac{(w_1 - w_3)(\eta_2 - \eta_3) - (w_2 - w_3)(\eta_1 - \eta_3)}{C_2} \tag{4.11}$$

$$b = -\frac{(w_1 - w_3)(\xi_2 - \xi_3) - (w_2 - w_3)(\xi_1 - \xi_3)}{C_2} \tag{4.12}$$

where

$$C_2 = (\xi_1 - \xi_3)(\eta_2 - \eta_3) - (\xi_1 - \xi_2)(\eta_1 - \eta_3).$$

Hence $\nabla u = \binom{a}{b}$ is determined. 62

If $\lambda_j^k = a^j x + b^j y + c$, then a^j and b^j are got from (4.11) and (4.12) by taking $w_i = \delta_{ij}$. Note that $C_2 = 1/2$ area K_k. Then $a_j^k = C_2/2\ (aa_j + bb_j)$.

4.6 Solution by Direct Methods

In Section 4.6.2 and 4.6.3 we gave algorithms to solve the equation $Ax = b$ when A is symmetric and positive definite. When A is not symmetric or A is sparse, direct methods can be used to solve the equation $Ax = b$.

4.6.1 Review of the Properties of Gaussian Elimination

The principle of Gaussian elimination is to decompose A into a product LU where L is lower triangular and U is upper triangular so that the linear system

$$Ax = b$$

is reduced to solving 2 linear systems with triangular matrices

$$\begin{aligned} Ly &= b, \\ Ux &= y. \end{aligned}$$

Each of these is very easy to solve. For $Ly = b, y = (y_i), y_1$ is given by the first equation, hence y_2 is given by the second since y_1 is already known, etc.

To decompose A into LU, one proceeds iteratively. Let

$$A^{(k)} = \left(a_{ij}^{(k)}\right) 1 \leq i, j \leq N \quad ,$$

63 be such that

$$a_{ij}^{(k)} = 0 \quad \text{for} \quad 1 \leq j \leq k-1 \quad \text{and} \quad i > j.$$

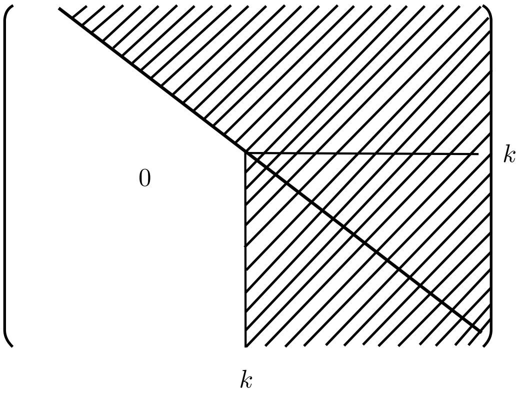

Figure 4.2:

Then U is $A^{(N)}$.

To get $A^{(k+1)}$ from $A^{(k)}$ one adds the k^{th} equation multiplied by a scaling factor of the i^{th} equation in order to have $a_{ik}^{(k+1)} = 0$. The scaling factor has to be

$$-\frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, \quad \text{and hence}$$
$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - \frac{a_{ik}^{(k)}\ a_{kj}^{(k)}}{a_{kk}^{(k)}} i, \quad j = k+1, \ldots, n. \tag{4.13}$$

For a full matrix the order of operations for this process is $N^3/3$. For a band matrix, i.e. a matrix such that

$$a_{ij} = 0 \quad \text{if} \quad |i-j| \geq w,$$

We see that if $|i-j| \geq w$, then either $|k-i|$ or $|k-j|$ is greater than $w+1$, provided that $i, j \geq k+1$. Hence in the formula (4.13) an element which is outside the band is never modified since the corrective term in (4.13) is a always zero.

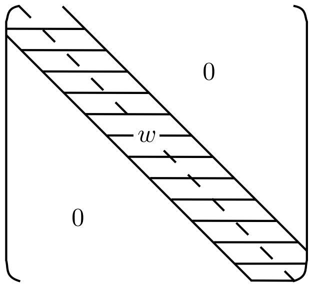

Figure 4.3:

64 More precisely we have

PROPOSITION 4. For a band matrix A with bandwidth w, at the k^{th} step of the Gaussian elimination, only w^2 "corrective elements" have to be computed and added to the submatrix.

$$\begin{pmatrix} a_{k+1,k+1} & \cdots & a_{k+1,k+w} \\ \vdots & & \\ a_{k+w,k+1} & \cdots & a_{k+w,k+w} \end{pmatrix}$$

Note that we get an evaluation of the number of operations for the process which is about Nw^2 (instead of $N^3/3$).

4.6.2 Stiffness Matrix and Stiffness Submatrix

For simplicity we consider the Neumann problem. Find $u_h \varepsilon V_h$ such that

$$a(u_h, v) = L(v) \forall\ v \varepsilon V_h,$$

where V_h is a finite dimensional subspace of $H^1(\Omega)$ constituted with functions which are continuous and piecewise linear on the elements of the triangulation T_h and

$$a(u, v) = \int\limits_{\Omega} (\nabla u . \nabla v + uv)\, dx.$$

65 For $(w_i)_{1 \leq i \leq N}$, a basis of V_h (where N denotes the number of vertices of

T_h), we have

$$a(u, v) = \sum_{i,j=1}^{N} u_i a_{ij} v_j,$$

where v_i (respectively u_i) denote the value of v (respectively u) at the i^{th} vertex of T_h and

$$a_{ij} = \int_{\Omega} (\nabla w_i \ \nabla w_j + w_i w_j)\, dx.$$

But this is not a practical way to compute the elements a_{ij} of the matrix A of the linear system to be solved, since the support of the w_i involves several elements of T_h. Instead one writes

$$a(u, v) = \sum_{K \varepsilon T_h} \int_{K} (\nabla u . \nabla v + uv)\, dx.$$

Hence, as

$$u(x) = \sum_{\alpha=1}^{3} u_{m_{\alpha K}} \lambda_{\alpha}^{K}(x),$$
$$v(x) = \sum_{\beta=1}^{3} v_{m_{\beta K}} \lambda_{\beta}^{K}(x),$$

in the element K, where $(m_{\alpha K})_{\alpha=1,2,3}$ denotes the 3 vertices of the element K and $\lambda_{\alpha}^{K}(x)$ the associated barycentric coordinates. One has

$$\sum_{i,j=1}^{N} u_i a_{ij} v_j = \sum_{K \varepsilon T_h} \sum_{\alpha,\beta=1}^{3} u_{m_{\alpha K}} v_{m_{\beta K}} a_{\alpha\beta}^{K},$$

where **66**

$$a_{\alpha\beta}^{K} = \int_{K} (\nabla \lambda_{\alpha}^{K} . \nabla \lambda_{\beta}^{K} + \lambda_{\alpha}^{K} \lambda_{\beta}^{K})\, dx.$$

The matrix $A^K = (a_{\alpha\beta}^{K})_{1 \le \alpha, \beta \le 3}$ is called the *element stiffness matrix* of K.

A convenient algorithm to compute A is then the following *Assembling algorithm*.

$$\begin{cases} 1. \quad \text{Set} \quad A = 0 \\ 2. \quad \text{For} \quad K\varepsilon T_h, \quad \text{compute } A^K \text{ and for } \alpha, \beta = 1, 2, 3 \text{ make} \\ \quad a_{m_{\alpha K}, m_{\beta K}} = a_{m_{\alpha K}, m_{\beta K}} + a^K_{\alpha\beta}. \end{cases}$$

Exercise 3. Write a Fortran subroutine performing the assembling algorithm (without the computation of the element stiffness matrices A^K which will be assumed to be computed in another subroutine).

4.6.3 Computation of Element Stiffness Matrices

We shall consider more sophisticated elements, e.g. the triangular, quadratic element with 6 nodes.

The midside points have to be included in the numbering of the vertices to describe properly the triangulation. For each element K, one has to give the 6 numbers of its 6 nodes in the global numbering,

$$m_{\alpha K}, \alpha = 1, \ldots, 6.$$

67 The assembling algorithm of last section is still valid except that α and β range now from 1 to 6 and that λ^K_α has to be replaced by $p^K_\alpha, \alpha = 1, 2, \ldots, 6$, the local basis functions of the interpolation (see chapter 3).

To compute the element stiffness matrix

$$A^K = \left(a^K_{\alpha\beta}\right)_{\alpha,\beta=1,\ldots,6},$$

one introduces the mapping,

$$F : \hat{K} \to K$$

where $\hat{K}$ is the triangle $(0, 0), (1, 0), (0, 1)$. Since F is affine, we have

$$F(\xi) = B\xi + b,$$

where B is 2×2 matrix and $b\varepsilon\mathbb{R}^2$.

Let $\hat{u}(\xi) = u(F(\xi))$ and $\hat{v}(\xi) = v(F(\xi))$. One has

$$\int_K uv\,dx = \int_{\hat{K}} \hat{u}\hat{v}\ \det(B)\,d\xi \tag{4.14}$$

In the same way one has

$$\nabla\hat{u}|_{\xi} = B^T\ \nabla u|_{F(\xi)},$$

since B^T is the Jacobian matrix of F. Therefore,

$$\int_K \nabla u.\nabla v\,dx = \int_{\hat{K}} (B^{-T}\nabla\hat{u}).(B^{-T}\ \nabla\hat{v})\det(B)\,d\xi \tag{4.15}$$

Finally to compute the coefficients $a^K_{\alpha\beta}$ of the element stiffness matrix A^K, one notices that

$$\hat{u}(\xi) = \sum_{\alpha=1}^{6} u_{m_{\alpha K}} p_\alpha(\xi),$$

where $(p_\alpha(\xi),\ \alpha = 1,\ldots,6)$ are the basis functions of K which are **68**
easily computed once for all. Note that

$$\lambda_1 = 1 - \xi_1 - \xi_2,\ \lambda_2 = \xi_{1,}\lambda_3 = \xi_2.$$

As p are polynomials (even for higher degree elements) the integrals in (4.14) and (4.15) can be computed by noticing that

$$\int_{\hat{K}} \xi_1^i \xi_2^j\,d\xi = \frac{i!\ j!}{(i+j+2)!}$$

However, for the simplicity of the programming they are usually computed by numerical integration: every integral of the type $\int_K f(\xi)\,d\xi$ is replaced by

$$\sum_{\ell=1}^{L} w_\ell\ f(b_\ell)$$

where $(b_\ell)_{\ell=1,\dots,L}$ are called the nodes of the numerical integration formula and $(w_\ell)_{\ell=1,\dots,L}$ the coefficients.

The programming is easier since one may compute (in view of (4.14) and (4.15) only the values of p_α and $\partial p_\alpha/\partial\xi_i$ at the points b_ℓ. For more details and model programs we refer to Mercier – Pironneau [32].

Chapter 5

Review of the Error Estimates for the Finite Element Method

THE PURPOSE OF this chapter is to state the theorems on error estimates which are useful for our future analysis. The proof of the theorems can be found in CIARLET [9]. **69**

DEFINITION. *Let* $\Omega \subset \mathbb{R}^n$ *be an open subset,* $m \geq 0$ *be an integer and* $1 \leq p \leq +\infty$*. Then the Sobolev Space* $W^{m,p}(\Omega)$ *is defined by*

$$W^{m,p}(\Omega) = \{v \varepsilon L^p(\Omega) : \partial^\alpha v \varepsilon L^p(\Omega), \quad \text{for all } |\alpha| \leq m\}.$$

On the space $W^{m,p}(\Omega)$ *we define a norm* $\| \,.\, \|_{m,p,\Omega}$ *by*

$$\| v \|_{m,p,\Omega} = \left(\int\limits_\Omega \sum_{|\alpha| \leq m} |D^\alpha v|^p \, dx \right)^{1/p},$$

and a semi norm $| \,.\, |_{m,p,\Omega}$ *by*

$$|v|_{m,p,\Omega} = \left(\int\limits_\Omega \sum_{|\alpha| = m} |D^\alpha v|^p \, dx \right)^{1/p}.$$

If k is an integer, then we consider the quotient space

$$W^{-k+1,p}(\Omega) = W^{k+1,p}(\Omega)/p_k(\Omega)$$

with the quotient norm

$$\| \tilde{v} \|_{k+1,p,\Omega} = \operatorname*{Inf}_{\ell \varepsilon \mathbb{P}_k} \| v + \ell \|_{k+1,p,\Omega},$$

70 *where $\tilde{v}$ is the equivalence class containing v.*

We introduce a semi norm in $\tilde{W}^{k+1,p}(\Omega)$ by

$$|\tilde{v}|_{k+1,p,\Omega} = |v|_{k+1,p,\Omega}.$$

Then we have

THEOREM 1. (CIARLET - RAVIART). *In $\tilde{W}^{k+1,p}(\Omega)$ the semi norm $|\tilde{v}|_{k+1,p,\Omega}$ is a norm equivalent to the quotient norm $\| v \|_{k+1,p,\Omega}$.*

Using this theorem it is easy to prove

THEOREM 2. *Let $W^{k+1,p}(\Omega)$ and $W^{m,q}(\Omega)$ be such that $W^{k+1,p}(\Omega) \hookrightarrow W^{m,q}(\Omega)$ (continuous injection). Let*

$$\pi \in \mathscr{L}(W^{k+1,p}(\Omega),\ W^{m,q}(\Omega))$$

be such that for each $p \varepsilon \mathbb{P}_k$, $\pi p = p$. Then there exists a $c = c(\Omega, \pi)$ such that for each $v \in W^{k+1,p}(\Omega)$

$$|v - \pi v|_{m,q,\Omega} \leq c|v|_{k+1,p,\Omega}.$$

DEFINITION. *Two open subsets $\hat{\Omega}, \Omega$ of $\mathbb{R}^n$ are said to be affine equivalent if there exists an affine map F from $\hat{\Omega}$ onto Ω such that $F(\hat{x}) = B\hat{x} + b$, where B is a $n \times n$ non singular matrix and $b \varepsilon \mathbb{R}^n$.*

We have

THEOREM 3. *Let $\hat{\Omega}, \Omega$ be affine equivalent with F as their affine map.*
71 *Then there exist constants $\hat{c}, c$ such that for all $v \varepsilon W^{m,p}(\Omega)$,*

$$|\hat{v}|_{m,p,\hat{\Omega}} \leq c \| B \|^m \ |\det B|^{-1/p} |v|_{m,p,\Omega},$$

and for all $\hat{v}\varepsilon W^{m,p}(\hat{\Omega})$,

$$|v|_{m,p,\Omega} \leq \hat{c} \parallel B^{-1} \parallel^{m} \ |\det B|^{1/p} \ |\hat{v}|_{m,p,\hat{\Omega}},$$

where

$$\hat{v} = v \bullet F.$$

If h $(resp.\ \hat{h})$ is the diameter of Ω $(resp.\ \hat{\Omega})$ and p $(resp.\ \hat{p})$ is the supremum of the diameters of all balls that can be inscribed in Ω $(resp.\ \hat{\Omega})$, then we have

THEOREM 4. $\parallel B \parallel \leq h/\hat{\rho}$ *and* $\parallel B^{-1} \parallel \leq \hat{h}/\rho$.

DEFINITION. *Two finite elements* $(\hat{K}, \hat{\Sigma}, \hat{P})$ *and* (K, Σ, P) *are said to be affine equivalent if there exists an affine map* $F\hat{x} = B\hat{x} + b$ *on* $\mathbb{R}^n$, *where* B *is an* $n \times n$ *non singular matrix, and* $b\varepsilon\mathbb{R}^n$ *such that*

(i) $F(\hat{K}) = K$

(ii) $\hat{p} = \{\hat{p} = p \circ F : p\varepsilon P\}$,

(iii) $\hat{\Sigma} = \{\hat{\phi} = F^{-1} \circ \phi : \phi\varepsilon\Sigma\}$

where

$$F^{-1} \circ \phi(\hat{p}) = \phi(\hat{p} \circ F^{-1}).$$

DEFINITION. *Let* (K, Σ, P) *be a finite element and* $v : K \to \mathbb{R}$ *be a* **72**
smooth function on K. *Then by virtue of the* P*-unisolvency of* Σ *there exists a unique element, say,* $\pi_K v \in P$, *such that* $\phi(\pi_K v) = \phi(v)$ *for all* $\phi \in \Sigma$. *The function* $\pi_K v$ *is called the* P*-interpolate function of* v *and the operator* $\pi_K : C^\infty(K) \to P$ *is called the* P*-interpolation operator.*

Now we state an important theorem which is often used.

THEOREM 5. *Let* $(\hat{K}, \hat{\Sigma}, \hat{P})$ *be a finite element. Let* $s(= 0, 1, 2)$ *be the maximal order of derivatives occurring in* Σ. *Assume that*

(i) $W^{k+1,p}(\hat{K}) \hookrightarrow C^s(\hat{K})$,

(ii) $W^{k+1,p}(\hat{K}) \hookrightarrow W^{m,q}(\hat{K})$,

(iii) $P_k \subset \hat{P} \subset W^{m,q}(\hat{K})$,

Then there exists a constant $C = C(\hat{K}, \hat{\Sigma}, \hat{P})$ *such that for all affine equivalent finite elements* (K, Σ, P) *we have*

$$|v - \pi_K v|_{m,q,K} \leq C\ (\text{meas}\ K)^{1/q-1/p} \frac{h_K^{k+1}}{\rho_K^m} |v|_{k+1,p,K}$$

for all $v \in W^{k+1,p}(K)$, *where* π_K *is a* P*-interpolate operator,* h_K *is the diameter of* K *and* ρ_K *is the supremum of diameter of all balls inscribed in* K.

DEFINITION. *A family* (T_h) *of triangulations of* Ω *is regular if*

73 *(i)* *for all* h *and for each* $K \varepsilon T_h$ *the finite elements* (K, Σ, P) *are all affine equivalent to a single finite element* $(\hat{K}, \hat{\Sigma}, \hat{P})$*;*

(ii) *there exists a constant* σ *such that for all* T_h *and for each* $K \varepsilon T_h$ *we have*

$$\frac{h_K}{\rho_K} \leq \sigma;$$

(iii) *for a given triangulation* T_h*, if*

$$h = \max_{K \varepsilon T_h} h_K,$$

then $h \to 0$.

Exercise 1. Prove that there exists a constant C independent of h such that

$$|p|_{1,k} \leq C/h\ |p|_{\circ,K} \quad \text{for all } p \varepsilon \mathbb{P}_k.$$

A theorem which gives a global error bound is the following.

THEOREM 6. *Let us assume that*

(i) $\pi_h : H^{k+1}(\Omega) \to V_h$*, the restriction of* π_h *to* V_h *being the identity,*

(ii) $V_h \subset \underset{K \varepsilon T_h}{\pi} \mathbb{P}_k(K)$,

74 *(iii)* $V_h \subset H^m(\Omega)$,

(iv) $u \in H^m(\Omega)$ *(regularity assumption),*

Then we have

$$\| u - \pi_h u \|_{m,\Omega} \leq C\, h^{k+1-m} |u|_{k+1,\Omega},$$

where C is a constant independent of h and (T_h) *is a regular family of triangulations.*

For stating a theorem on L_2 -error estimates we need the definition of a regular adjoint problem.

DEFINITION. *Let* $V = H^1(\Omega)$ *or* $H^1_\circ(\Omega)$, $H = L^2(\Omega)$. *The adjoint problem:*

$$\begin{cases} \text{Find} \quad \phi \varepsilon V \quad \text{such that} \\ a(v, \phi) = (g, v) \quad \text{for all } v \varepsilon V, \end{cases}$$

is said to be regular if

(i) for all $g \varepsilon L^2(\Omega)$*, the solution* ϕ *of the adjoint problem for g belongs to* $H^2(\Omega) \cap V$*;*

(ii) there exists a constant C such that

$$\| \phi \|_{2,\Omega} \leq C |g|_{\circ,\Omega}.$$

We now have

THEOREM 7. *Let* (T_h) *be a regular family of triangulations on* Ω *with reference finite element* $(\hat{K}, \hat{\Sigma}, \hat{P})$. *Let* $s = 0$ *and* $n \leq 3$. *Suppose there exists an integer* $k \geq 1$ *such that* $u \varepsilon H^{k+1}(\Omega)$ $P_k \subset \hat{p} \subset H^1(\hat{K})$. *Assume further that the adjoint problem is regular. Then there exists a constant C independent of h such that* 75

$$|u - u_h|_{\circ,\Omega} \leq C\, h^{k+1}\, |u|_{k+1,\Omega}.$$

Chapter 6

Problems with an Incompressibility Constraint

6.1 Introduction

We recall the variational formulation of the Stokes problem (see Chapter 76
2).

Find $u \varepsilon V$ such that

$$a(u, v) = L(v) \ \forall \ v \varepsilon V,$$

where

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

$$L(v) = \int_\Omega f . v \ dx, \ \ f \varepsilon \ (L^2(\Omega))^n,$$

and

$$V = \{v \varepsilon (H^1_\circ(\Omega))^n : \operatorname{div} v = 0\}.$$

It is difficult to construct internal approximations of V because of the constraint $\operatorname{div} v = 0$. In two dimensional problem we know that

$v \varepsilon V \Leftrightarrow$ there exists $\psi \varepsilon H^2_\circ(\Omega)$ such that

$$v = \operatorname{rot} \psi,$$

where

$$\operatorname{rot} \psi = \left(\frac{\partial \psi}{\partial x_2}, -\frac{\partial \psi}{\partial x_1} \right)$$

Therefore, it seems logical that the difficulties we encountered in approximating $H^2(\Omega)$ in a conforming way are transferred to the conforming approximation of V.

6.2 Approximation Via Finite Elements of Degree 1

77 Let $W_h = \{v_h \ \varepsilon \ (C^\circ(\overline{\Omega}))^2 : v_h|_K \ \varepsilon (\mathbb{P}_1(K))^2$, for $K \varepsilon T_h, v_h = 0$ on $\partial\Omega\}$ It is natural to try for V_h the space

$$\{v_h \ \varepsilon \ W_h : \operatorname{div} \ v_h = 0\}.$$

But for most triangulations, $V_h = \{0\}$. This is due to the fact that the number of equations due to the constraint $\operatorname{div} v_h = 0$ is greater than the number of degrees of freedom of W_h. In fact,

$$\text{Dimension of} \quad W_h = 2 \quad (\text{\# internal vertices}).$$

Number of equations due to the constraint $\operatorname{div} v_h = 0$ is equal to number of triangles.

Hence V_h cannot be a good approximation to V. However, if the triangulation T_h is obtained by first taking quadrilaterals and then dividing each quadrilateral into four triangles by joining the diagonals (see figure 6.1), we obtain a 'good space' V_h. In this case only 3 of the four equations $\operatorname{div} v_h = 0$ are independent.

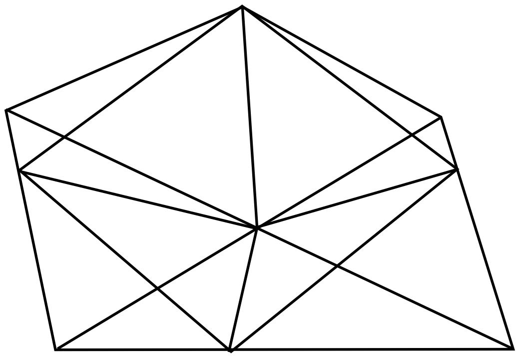

Figure 6.1:

Exercise 1. Let K be a quadrilateral. Let it be divided into four triangles 78
$T_i, 1 \leq i \leq 4$, by

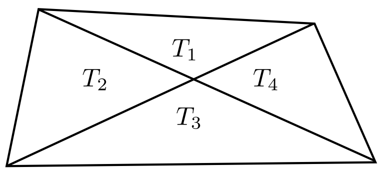

Figure 6.2:

joining the diagonals of K. Let $v \varepsilon C^\circ(\overline{K})$ such that

$$v|_{T_i} \; \varepsilon \mathbb{P}_1(T_i), 1 \leq i \leq 4. \quad \text{Let} \quad d_i = \operatorname{div} v|_{T_i} \; 1 \leq i \leq 4.$$

Then show that $d_1 + d_3 = d_2 + d_4$.

When the mesh is *uniform*, it is possible to prove convergence and to construct an interpolation operator,

$$\pi_h : V \to V_h$$

such that

$$\| v - \pi_h v \|_{1,\Omega} \leq \text{ch} \| v \|_{2,\Omega} .$$

Therefore the solution v_h of the approximate problem

$$a(u_h, v_h) = L(v_h) \;\; \forall \; v_h \; \varepsilon \; V_h,$$

satisfies

$$\| u - u_h \|_{1,\Omega} \leq \text{ch}; \| u - u_h \|_{0,\Omega} \leq \text{ch}^2 .$$

When the domain is a square and the mesh is uniform we define π_h as follows.

We choose for simplicity the diagonals of the square to be the coordinate axes.

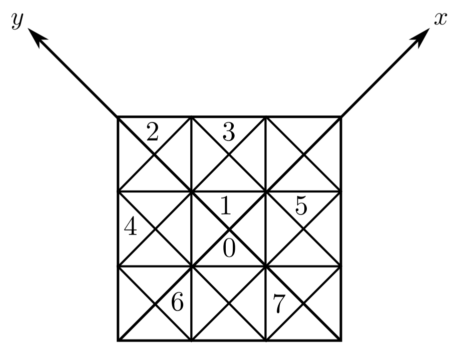

Figure 6.3:

79

$\pi_h u$ at each of the main nodes (like 1, 5, 6, 7) is chosen as an average of u. For example, at the node 1 the two components of $\pi_h u$ are given by

$$\begin{aligned}(\pi_h u)_1 &= \text{average of } u_1 \text{ on } 02,\\ (\pi_h u)_2 &= \text{average of } u_2 \text{ on } 34.\end{aligned}$$

At each of the secondary nodes (like 0, 2, 3, 4) $\pi_h u$ is chosen as an average of the values of $\pi_h u$ at the main nodes. For example, at the node 0

$$\begin{aligned}(\pi_h u)_1\ (0) &= \frac{(\pi_h u)_1\ (1) + (\pi_h u)_1\ (7)}{2},\\ (\pi_h u)_2\ (0) &= \frac{(\pi_h u)_2\ (5) + (\pi_h u)_2\ (6)}{2},\end{aligned}$$

Numerically the method works even for irregular, not too distorted, meshes.

6.3 The Fraeijs De Veubeke - Sander Element

We shall first describe a C^1 element of a non standard type: the Fraeijs de Veubeke - Sander element.

Let K be a quadrilateral. We divide K into four triangles $T_i, 1 \leq i \leq$ **80**
4, by joining the diagonals of K.

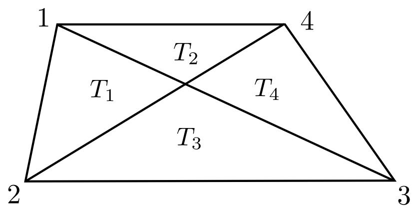

Figure 6.4:

Let

$$Q(K) = \{p \varepsilon C^1(K) : p \varepsilon \mathbb{P}_3(T_i), 1 \leq i \leq 4\}.$$

We have

LEMMA 1. $\dim Q(K) = 16$.

Proof. Indeed we choose $p = p_1$ on T_1 where $p_1 \varepsilon \mathbb{P}_3$. Then p_1 depends on 10 parameters. Let $p_2 \varepsilon \mathbb{P}_3$ be such that p_2 and $\partial p_2/\partial n$ vanish on the diagonal 13. Hence p_2 depends on 10-4-3=3 parameters. We choose $p = p_1 + p_2$ on T_2 and it is easy to see that p is C^1 across 13.

In the same way we choose $p = p_1 + p_3$ on T_3 with $p_3 = \partial p_3/\partial n = 0$ on 24. p_3 depends again on 3 parameters and p will be C^1 across 24.

Taking $p = p_1 + p_2 + p_3$ on T_4, it is easy to see that p is C^1 across 13 and 24. Finally, p depends on $10 + 3 + 3 = 16$ parameters. Hence the result. $\square$

We choose, **81**

$$\Sigma_K = \left\{\delta_i, \frac{\partial}{\partial x}\delta_i, \frac{\partial}{\partial y}\delta_i, 1 \leq i \leq 4; \frac{\partial}{\partial n}\delta_{b_i}, 1 \leq i \leq 4\right\}$$

as the set of degrees of freedom. Here b_i denotes the midpoint of a side of the quadrilateral K.

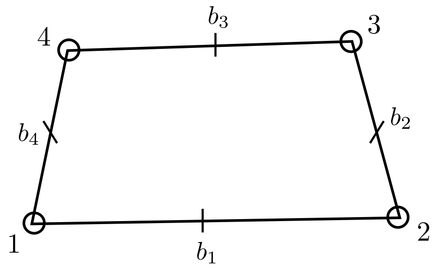

Figure 6.5:

We refer the reader to CIAVALDINI-NEDELEC [11] to prove that this is an admissible choice of degrees of freedom.

Let Q_h denote a regular family of *quadrangulations* of the polygonal domain Ω. Let

$$\chi_h = \{\phi_h \varepsilon C^1(\overline{\Omega}) : \ \phi_h|_K \varepsilon Q(K), \ \phi_h = \frac{\partial \phi_h}{\partial n} = 0 \quad \text{on} \quad \partial\Omega\}.$$

Then the following result has been proved by CIAVALDINI-NEDELEC [11].

THEOREM 2. *The operator $\pi_h : H^3(\Omega) \to \chi_h$ defined by the above choice of degrees of freedom satisfies*

$$\| \phi - \pi_h \phi \|_{2,\Omega} \leq \mathrm{ch}^s \| \phi \|_{s+2,\Omega},$$

where $s = 1, 2$.

82 As a consequence of this result the space χ_h can be used to approximate any variational problem on $H^2_\circ(\Omega)$ and in particular the biharmonic problem associated to the Stokes problem. However, as is the case with any Hermite finite element, the programming of the Fde V-S element is difficult and it is easier to use the corresponding element in the velocity formulation, which is quadratic.

6.4 Approximation of the Stokes Problem Via Quadratic Elements

Let T_h be the triangulation associated to Q_h by dividing each quadrilateral of Q_h into four triangles in the usual way. Let

$$W_h = \left\{v_h \ \varepsilon C^{\circ}(\overline{\Omega})^2 : v_h|_T \ \varepsilon(\mathbb{P}_2(T))^2, T\varepsilon T_h, v_h = 0 \quad \text{on} \quad \partial\Omega\right\}$$

and

$$V_h = \{v_h \varepsilon W_h : \text{div } \ v_h = 0\}.$$

Obviously $v_h \varepsilon V_h$ implies that there exists $\psi_h \varepsilon \chi_h$ such that $v_h = \text{rot}\,\psi_h$. Therefore we can state

THEOREM 3. *There exists a constant c such that, for $u \varepsilon V \cap H^{s+1}(\Omega)$ ($s = 1$ or 2),*

$$\operatorname*{Inf}_{v_h \varepsilon V_h} \| u - v_h \|_{1,\Omega} \leq \text{ch}^s \| u \|_{s+1,\Omega} \ .$$

Proof. As $u\varepsilon V$, there exists $\psi \varepsilon H^{s+2}(\Omega) \cap H^2_{\circ}(\Omega)$ such that $u = \text{rot}\,\psi$. Then choosing $v_h = \text{rot}\,\pi_h \psi$ we obtain the result. ◻

Numerically, one may solve the approximate problem: **83**

$$\begin{cases} \text{find} \quad u_h \ \varepsilon V_h \quad \text{such that} \\ a(u_h, v_h) = L(v_h) \ \ \forall \ v_h \ \varepsilon V_h, \end{cases}$$

via a penalty method.

Let $\varepsilon > 0$ and

$$a_\varepsilon \ (u, v) = a(u, v) + 1/\varepsilon \int_\Omega \text{div } \ v \, \text{div } \ u \ dx.$$

The penalised problem is:

$$\begin{cases} \text{Find} \quad u_h \ \varepsilon W_h \quad \text{such that} \\ a_\varepsilon(u_{\varepsilon h}, v_h) = L(v_h) \ \ \forall \ v_h \ \varepsilon W_h. \end{cases}$$

This problem is much easier to solve. We shall see in Chapter 7 that the order of the error due to penalization is only

$$0(\varepsilon) : \| u_{\varepsilon h} - u_h \|_1 \leq c.\varepsilon,$$

where c may depend on h.

6.5 Penalty Methods.

We now come back to the case of finite elements of degree 1 where the space W_h is

$$W_h = \left\{v_h \ \varepsilon \ (C^\circ(\overline{\Omega}))^2 : \ v_h|_K \ \varepsilon \ (\mathbb{P}_1(K))^2, K \ \varepsilon \ T_h, v_h = 0 \text{ on } \partial\Omega\right\}.$$

The approximate problem can be taken as

$$\begin{cases} \text{Find} \quad u_{\varepsilon h} \ \varepsilon \ W_h \quad \text{such that} \\ a_\varepsilon(u_{\varepsilon h}, v_h) = L(v_h) \quad \text{for all} \quad v_h \ \varepsilon \ W_h. \end{cases}$$

We have

84 **THEOREM 4.** *Assume that $u \varepsilon H^2(\Omega)$. Then there exists a constant c such that*

$$\| u_{\varepsilon h} - u \|_1^2 \leq c[h^2(1 + 1/\varepsilon) + h + \sqrt{\varepsilon}].$$

Hence by choosing $\varepsilon = h^{4/3}$ we obtain

$$\| u_{\varepsilon h} - u \|_1 \leq \mathrm{ch}^{1/3}.$$

Exercise 2. Prove the above theorem.

Note that the above convergence rate is very poor, which is confirmed by the poor numerical results obtained with this method.

6.6 The Navier-Stokes Equations.

The stationary flow of a viscous, Newtonian fluid subjected to gravity loads in a bounded region Ω of $\mathbb{R}^3$ is governed by the following dimensionless equations.

$$\left.\begin{aligned} -\gamma\Delta u + \sum_{i=1}^{3} u_i \frac{\partial u}{\partial x_i} + \nabla p = f \quad \text{in} \quad \Omega, \\ \operatorname{div} u = 0 \quad \text{in} \quad \Omega, \\ u = 0 \quad \text{in} \quad \partial\Omega, \end{aligned}\right\} \tag{6.1}$$

where u represents the velocity, p the pressure and f is the body force. All these quantities are in dimensionless form and $\gamma = \frac{\mu}{DV_\rho} = \frac{1}{Re}$ where Re is called the Reynolds number. Here is the viscosity of the fluid D a length characterizing the domain Ω, V a characteristic velocity of the 85
flow and ρ the density of the fluid (For more details the reader is referred to BIRD-STEWART-LIGHTFOOR 'Transport Phenomena, Wiley Ed. p. 108).

The Reynolds number is the only parameter in the equation and it measures how far the Navier-Stokes model is from the Stokes model. The limiting case $\gamma = 0$ corresponds to Euler's equations for inviscid fluids. However, at high Reynolds number, the flow develops a boundary layer near the boundary. Moreover, instability and bifurcation phenomena can be observed which correspond physically to turbulence. We are going to study only the flows at *low* Reynolds number.

6.7 Existence and Uniqueness of Solutions of Navier-Stokes

EQUATIONS AT LOW REYNOLDS NUMBERS. The variational formulation of Navier-Stokes equations is:

$$\begin{cases} \text{Find} \quad u \varepsilon V \quad \text{such that} \\ a(u,v) + b(u,u,v) = (f,v) \quad \forall v \varepsilon V, \end{cases} \tag{6.2}$$

where

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

$$b(u,v,w) = \int_\Omega \sum_{i,j=1}^{3} u_i \frac{\partial v_j}{\partial x_i} w_j \; dx$$

and

$$V = \{v \varepsilon \; (H^1_\circ(\Omega))^3 : \operatorname{div} v = 0\}.$$

Exercise 3. Show that if u is a solution of (6.2), then there exists a $p \varepsilon L^2(\Omega)$ such that $\{u, p\}$ is a solution of (6.1) in the sense of distributions. 86

Exercise 4. Show that for all $u, v, w \varepsilon V$ one has

$$b(u, v, w) = -b(u, w, v),$$
$$b(u, v, v) = 0.$$

As $H^1(\Omega) \hookrightarrow L^4(\Omega)$ (see LADYZHENSKAYA [27] for a proof of this), using Schwarz inequality twice we obtain

$$b(u, v, w) \leq \| u \|_{0,4} \| v \|_1 \| w \|_{0,4} \leq c \| u \|_1 \| v \|_1 \| w \|_1 .$$

Hence

$$\beta = \operatorname*{Sup}_{u,v,w\varepsilon V} \frac{b(u, v, w)}{\| u \| \| v \| \| w \|} < \infty.$$

where $\| u \| = |u|_1$. We have

THEOREM 5. *Assume that* $\beta/\gamma^2 \; \| f \|_* < 1$. *Then the problem* (6.2) *has a unique solution.*

Proof. Let $u^i \varepsilon V, i = 1, 2$. Let $v^i, i = 1, 2$ be the solution

$$a(v^i, w) + b(u^i, v^i, w) = (f, w) \quad \forall w\varepsilon \; V \; i = 1, 2 \tag{6.3}$$

Note that (6.3) is a linear problem and has a unique solution by virtue of the Lax-Milgram Lemma.

87 Choosing $w = v^i$ in (6.3) it is easy to see that

$$\gamma \| v^i \|^2 \leq (f, v^i) \leq \| f \|_* \| v^i \| .,$$

Thus

$$\| v^i \| \leq \frac{\| f \|_*}{\gamma}$$

Taking $w = v^2 - v^1$, we obtain

$$\gamma \| w \|^2 \leq a(v^2 - v^1, w) = b(u^1, v^1, w) - b(u^2, v^2, w) = b(u^1, -w, w) +$$
$$+ b(u^1 - u^2, v^2, w) \leq 0 + \beta \| v^2 \| . \| w \| \| u^1 - u^2 \| .$$

Hence we obtain

$$\| v^1 - v^2 \| \leq \frac{\beta}{\gamma^2} \| f \|_* \| u^1 - u^2 \| .$$

Since $\frac{\beta}{\gamma^2} \| f \|_* < 1$, the mapping $T : u^i \to v^i$ is a strict contraction and has a fixed point which obviously is the unique solution of (6.2) □

REMARK 1. *The proof is constructive in the sense that the algorithm $u^{n+1} = Tu^n$ gives a sequence which converges to the solution. At each step of this algorithm one has to solve the linearised problem* (6.3).

REMARK 2. *If $\gamma^2 \geq \beta \| f \|_*$ then there exists atleast one solution to* (6.2). *The solution of* (6.2) *in this case may not be unique (see LIONS [28]). Note that problem* (6.2) *is equivalent to solving a non-linear equation $F(u) = 0$ where $F : V \to V'$ is given by* 88

$$F(u)(v) = a(u, v) + b(u, u, v) - (f, v).$$

Let G_u be the linear operator which is tangent to F, i.e.

$$(G_u w, v) = \lim_{\theta \to 0} \frac{1}{\theta}(F(u + \theta w) - F(u), v) = a(w, v) + b(u, w, v) + b(w, u, v).$$

If G_u is not singular, then u is an isolated solution, otherwise there may be a bifurcation. The eigenvalue problem associated to the linearised problem is:

$$\begin{cases} \text{Find} \quad w \varepsilon V, \lambda \varepsilon \not\mathbb{C} \quad \text{such that} \\ a(w, v) + b(u, w, v) + b(w, u, v) = \lambda(w, v) \ \ v \varepsilon V \end{cases}$$

and a study of this problem is of fundamental interest.

We refer the reader to BREZZI-RAPPAZ-RAVIART [6] for a study of the convergence in the case where u is an isolated solution. In the next section we restrict ourselves to the case where u is unique.

6.8 Error Estimates for Conforming Method

Let $V_h \subset V$. We consider the approximate problem:

$$\begin{cases} \text{Find} \quad u_h \; \varepsilon V_h \quad \text{such that} \\ a(u_h, v_h) + b(u_h, u_h, v_h) = (f, v_h) \; \forall \; v_h \; \varepsilon \; V_h \end{cases} \tag{6.4}$$

Let

$$\beta_h = \operatorname*{Sup}_{u,v,w \varepsilon V_h} \frac{b(u, v, w)}{\| u \| \| v \| \| w \|}$$

89 and

$$\| f \|_{h*} = \operatorname*{Sup}_{u \varepsilon V_h} \frac{(f, v)}{\| v \|}.$$

Then it is easy to see that (6.4) has a unique solution when

$$\frac{\beta_h}{\gamma^2} \| f \|_{h*} < 1.$$

The iterative method mentioned in Remark 1 converges for all γ satisfying $\gamma^2 > \beta_h \| f \|_{h*}$ Note that $\beta_h \leq \beta$ and $\| f \|_{h*} \leq \| f \|_*$; however, JAMET-RAVIART [23] proved that $\beta_h \to \beta$ and $\| f \|_{h*} \to \| f \|_*$ as $h \to 0$.

THEOREM 6. *Assume that*

$$\frac{\beta}{\gamma^2} \| f \|_* < 1 - \delta$$

with $0 < \delta < 1$*; then one has*

$$\| u - u_h \| \leq 3/\delta \| u - v_h \| \quad \forall \; v_h \; \varepsilon V_h.$$

Proof. Let $w_h = v_h - u_h$. Then

$$\gamma \| w_h \|^2 \leq a(v_h - u, w_h) + a(u - u_h, w_h),$$
$$a(u, w_h) = (f, w_h) - b(u, u, w_h),$$
$$a(u_h, w_h) = (f, w_h) - b(u_h, u_h, w_h),$$
$$a(u - u_h, w_h) = b(u_h, u_h - u, w_h)$$

$$+b(u_h - u, u, w_h) \leq \beta \parallel u_h \parallel\parallel v_h - u \parallel\parallel w_h \parallel + \beta \parallel u_h - u \parallel . \parallel u \parallel$$
$$\parallel w_h \parallel \quad \text{since} \quad b(u_h, u_h - v_h, w_h) = 0.$$

But we know that

$$\parallel u \parallel \leq 1/\gamma \parallel f \parallel_*, \parallel u_h \parallel \leq 1/\gamma \parallel f \parallel_* .$$

Therefore we obtain 90

$$\parallel w_h \parallel^2 \leq (\parallel v_h - u \parallel + (1-\delta) \parallel v_h - u \parallel + (1-\delta) \parallel u_h - u \parallel) . \parallel w_h \parallel .$$

As

$$\parallel u - u_h \parallel \leq \parallel u - v_h \parallel + \parallel w_h \parallel$$

we get

$$\delta \parallel w_h \parallel^2 \leq (3 - 2\delta) \parallel v_h - u \parallel\parallel w_h \parallel .$$

Hence

$$\parallel u - u_h \parallel \leq \parallel v_h - u \parallel + \frac{3 - 2\delta}{\delta} \parallel u - v_h \parallel .$$

This gives the desired result. ◻

An immediate consequence of Theorem 6 is that, when the solution u of (6.2) is sufficiently regular, we obtain the same error estimate for Navier-Stokes equations at low Reynolds number as for Stokes.

The method described in Section 6.4 is probably one of the best methods for Navier-Stokes equations at low Reynolds number.

However, when the Reynolds number is large, a major disadvantage is that the velocity field is required to be continuous (since our method is conforming) and this is not good to take into account boundary layer phenomena.

Indeed the velocity profile near the boundary has the behaviour as 91
shown in figure.

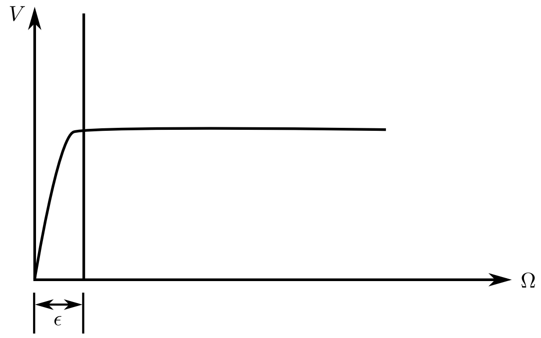

Figure 6.6:

$\in$ is called the thickness of the boundary layer. In fact, $\in = \gamma^{1/2}$, so that for high Reynolds number, this requires a very high refinement of the mesh near the boundary and therefore very expensive computer time. This can be partly avoided with mixed finite element since we shall work with discontinuous velocity fields.

Chapter 7

Mixed Finite Element Methods

7.1 The Abstract Continuous Problem

Let V, M, H be Hilbert spaces with $V \hookrightarrow H$. The continuous problem is: 92

Find $\{u, \lambda\}$ $\varepsilon V \times M$ such that

$$a(u, v) + b(v, \lambda) = (f, v) \quad \forall\, v\, \varepsilon V, \tag{7.1}$$

$$b(u, \mu) = (\phi, \mu) \quad \forall\, \mu\, \varepsilon M, \tag{7.1b}$$

where $a(\cdot\,,\cdot) : H \times H \to \mathbb{R}$ and $b(\cdot\,,\cdot) : V \times M \to \mathbb{R}$ are continuous bilinear forms and $f \varepsilon V', \phi \varepsilon M'$.

Let $a(\cdot\,,\cdot)$ and $b(\cdot\,,\cdot)$ satisfy

$$a(v, v) \geq \alpha \parallel v \parallel_H^2 \ \forall\, v\, \varepsilon H \quad \text{ellipticity,} \tag{7.2}$$

$$\operatorname*{Sup}_{v \varepsilon V} \frac{b(v, \mu)}{\parallel v \parallel_V} \geq \beta \parallel \mu \parallel_M \quad \text{Brezzi's condition} \tag{7.3}$$

$|a(u, v)| \leq |a| \parallel u \parallel \parallel v \parallel$, $|b(v, \mu)| \leq |b| \parallel v \parallel \parallel \mu \parallel$. We have

THEOREM 1. *If $H = V$ then, under the above assumptions problem* (7.1) *has a unique solution.*

Proof. Let us consider the regularised problem:

$$a(u_\varepsilon, v) + b(v, \lambda_\varepsilon) = (f, v) \ \ \forall\, v \ \varepsilon\ V, \tag{7.4}$$

$$-\, b(u_\varepsilon, \mu) + \varepsilon(\lambda_\varepsilon, \mu) = -(\phi, \mu) \ \ \forall\, \mu\ \varepsilon M, \tag{7.4b}$$

Let $\Phi = \{u, \lambda\}, \Psi = \{v, \mu\} \varepsilon V \times M$. We define

$$A_\varepsilon(\Phi, \Psi) = a(u, v) + b(v, \lambda) - b(u, \mu) + \varepsilon(\lambda, \mu), L(\Psi) = (f, v) - (\phi, \mu)$$

93 Then $A_\varepsilon(\Phi, \Psi)$ is $V \times M$ coercive and $L(\Psi)$ is a continuous, linear form on $V \times M$.

It is easy to see that problem (7.4) is equivalent to:

$$\left.\begin{array}{r} \text{Find} \quad \Phi \varepsilon V \times M \quad \text{such that} \\ A_\varepsilon(\Phi, \Psi) = L(\Psi)\ \forall\ \Psi \varepsilon V \times M \end{array}\right\} \tag{7.5}$$

By Lax-Milgram Lemma, problem (7.5) has a unique solution which implies that the regularized problem (7.4) has a unique solution.

Taking $v = u_\varepsilon$ in (7.4), $\mu = \lambda_\varepsilon$ in (7.4b) and adding and using the continuity of bilinear forms and H-ellipticity of $a(\cdot\,,\cdot)$, we get

$$\alpha \parallel u_\varepsilon \parallel^2 + \varepsilon \parallel \lambda_\varepsilon \parallel^2 \leq C\,(\parallel u_\varepsilon \parallel + \parallel \lambda_\varepsilon \parallel)\,, \tag{7.6}$$

where C is a constant.

Since

$$b(v, \lambda_\varepsilon) = (f, v) - a(u_\varepsilon, v) \leq (\parallel f \parallel_* + |a| \parallel u_\varepsilon \parallel) \parallel v \parallel$$

we obtain, using Brezzi's condition,

$$\beta \parallel \lambda_\varepsilon \parallel \leq \operatorname*{Sup}_{v \varepsilon V} \frac{b(v, \lambda_\varepsilon)}{\parallel v \parallel} \leq \parallel f \parallel_* + |a| \parallel u_\varepsilon \parallel\,.$$

This implies

$$\parallel \lambda_\varepsilon \parallel \leq C\,(1 + \parallel u_\varepsilon \parallel)\,. \tag{7.7}$$

From (7.6) and (7.7) we obtain

$$\parallel u_\varepsilon \parallel \leq C \quad \text{and} \quad \parallel \lambda_\varepsilon \parallel \leq C,$$

94 where C is a constant. Hence there exists a subsequence $\{\in'\}$, $u \in V$, $\lambda \in M$, such that

$$u_{\in'} \rightharpoonup u \quad \text{and} \quad \lambda_{\in'} \rightharpoonup \lambda.$$

Obviously $\{u, \lambda\}$ is a solution of (7.1).

If $\{u_1, \lambda_1\}$ and $\{u_2, \lambda_2\}$ are solutions of (7.1), then

$$\begin{aligned} a(u_1 - u_2, v) + b(v, \lambda_1 - \lambda_2) = 0 \ \forall v \in V, \\ b(u_1 - u_2, \mu) = 0 \ \forall \mu \in M. \end{aligned} \tag{7.8}$$

Taking $v = u_1 - u_2$ and $\mu = \lambda_1 - \lambda_2$, we obtain

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

Therefore

$$u_1 = u_2.$$

Since $u_1 = u_2$, using Brezzi's condition, we obtain from (7.8) that

$$\lambda_1 = \lambda_2.$$

Hence the solution of (7.1) is unique. □

REMARK 1. *We now give an error estimate for the solution of* (7.1) *and the regularized problem* (7.4). *We have*

$$a(u - u_\varepsilon, v) + b(v, \lambda - \lambda_\varepsilon) = 0 \ \forall v \in V.$$

Hence

$$\beta \parallel \lambda - \lambda_\in \parallel \leq |a|. \parallel u - u_\in \parallel . \tag{7.9}$$

From (7.1b) *and* (7.4b) *we obtain*

$$b(u - u_\in, \mu) + \in (\lambda_\in, \mu) = 0 \forall \mu \in M.$$

Choosing $v = u - u_\in, \mu = \lambda - \lambda_\in$, we get 95

$$\alpha \parallel u - u_\in \parallel^2 \leq a(u - u_\in, u - u_\in) = \in (\lambda_\in, \lambda - \lambda_\in) \leq \in \parallel \lambda_\in \parallel \parallel \lambda - \lambda_\in \parallel$$

Thus, using (7.7)

$$\parallel u - u_\in \parallel \leq C \in .$$

Thus from (7.9) *we get*

$$\| \lambda - \lambda_\in \| \leq C \in .$$

So we obtain

$$\| u - u_\in \| = 0(\in) \quad \textit{and} \quad \| \lambda - \lambda_\in \| = 0(\in).$$

REMARK 2. *Let $B : V \to M$ be such that*

$$(Bv, \mu)_M = b(v, \mu) \ \forall \ \mu \ \in \ M.$$

One has, from (7.4b) *and* (7.4)*,*

$$\lambda_\in = 1/ \in \ (Bu_\in - \phi),$$
$$a(u_\in, v) + 1/ \in \ (Bv, Bu_\in - \phi) = (f, v) \ \forall \ v \ \in \ V,$$

which correspond to penalization of (7.1b)*.*

REMARK 3. *We proved the existence and uniqueness of* (7.1) *only under the assumption $V = H$. If $V \neq H$ then we do not have a general existence theorem. But existence theorems for particular examples when $V \neq H$ are proved.*

We now give examples of (7.1)*.*

96 **EXAMPLE 1. The Stokes Problem.** We recall that the Stokes problem is:

Find $\{u, p\}$ such that

$$-\gamma\Delta u + \nabla p = f \quad \text{in} \quad \Omega,$$
$$\operatorname{div} u = 0 \quad \text{in} \quad \Omega,$$
$$u = 0 \quad \text{on} \quad \Gamma.$$

Without loss of generality we can take $\gamma = 1$. Using the standard technique of integration by parts we find that this corresponds to the problem:

Find

$$\{u, p\} \ \varepsilon \ (H^1_\circ(\Omega))^n \ \times \ (L^2(\Omega)/\mathbb{R})$$

such that

$$a(u,v) + b(v,p) = (f,v) \ \forall\ v\ \varepsilon\ (H^1_\circ(\Omega))^n,$$
$$b(u,\mu) = 0\ \forall\ \mu\ \varepsilon\ L^2(\Omega)/\mathbb{R};$$

where

$$a(u,v) = \int_\Omega \nabla u.\nabla v\, dx,$$
$$b(v,\mu) = -\int_\Omega \mu\ \text{div}\ v\, dx.$$

We take

$$V = H = (H^1_\circ(\Omega))^n,$$
$$M = (L^2(\Omega)/\mathbb{R}).$$

Clearly $a(\cdot,\cdot)$ is H-elliptic, continuous and bilinear. Let us prove Brezzi's condition,

$$\operatorname*{Sup}_{v\varepsilon V}\frac{b(v,\lambda)}{\|v\|_V} = \operatorname*{Sup}_{v\varepsilon(H^1_\circ(\Omega))^n}\frac{-\int\limits_\Omega \lambda\,\text{div}\ v}{\|v\|_V} = \operatorname*{Sup}_{v\varepsilon(H^1_\circ(\Omega))^n}\frac{<\nabla\lambda, v>}{\|v\|_V}$$
$$= \|\nabla\lambda\|_{(H^{-1}(\Omega))^n} \geq \frac{1}{C}\ \|\lambda\|_{L^2(\Omega)/\mathbb{R}}$$

where $\langle,\rangle$ denotes the duality between $(H^1_\circ(\Omega))^n$ and $(H^{-1}(\Omega))^n$. **97**

Thus Stokes problem has a unique solution.

EXAMPLE 2. The Biharmonic Problem. Let

$$V = H^1(\Omega),\ M = H^1_\circ(\Omega),\ H = L^2(\Omega).$$

Consider the problem:

Find

$$\{u,\lambda\}\ \varepsilon\ H^1(\Omega)\ \times\ H^1_\circ(\Omega)$$

such that

$$a(u, v) + b(v, \lambda) = 0 \quad \forall\ v\ \varepsilon\ H^1(\Omega), \tag{7.10}$$

$$b(u, \mu) = \int_\Omega \phi\mu\, dx \quad \forall\ \mu\ \varepsilon\ H^1_\circ(\Omega). \tag{7.11}$$

where

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

and $\phi\ \varepsilon\ L^2(\Omega)$.

Using integration by parts, we obtain from (7.10) that

$$\int_\Omega uv\, dx - \int_\Omega \Delta\lambda.v\, dx + \int_\Gamma \frac{\partial\lambda}{\partial n} v\, d\Gamma = 0,$$

which implies

$$\left.\begin{aligned} u - \Delta\lambda = 0 \quad \text{in} \quad \Omega,\\ \frac{\partial\lambda}{\partial n} = 0 \quad \text{on} \quad \Gamma.\end{aligned}\right\} \tag{7.12}$$

98 From (7.11) we get

$$-\Delta u = \phi. \tag{7.13}$$

Thus we have, from (7.12) and (7.13), the biharmonic problem

$$\begin{cases} \Delta^2\lambda = 0 \quad \text{in} \quad \Omega,\\ \frac{\partial\lambda}{\partial n} = 0 \quad \text{on} \quad \Gamma,\\ \lambda = 0 \quad \text{on} \quad \Gamma. \end{cases} \tag{7.14}$$

In the variational form of the biharmonic equation, we notice that $V = H^1(\Omega) \neq L^2(\Omega) = H$. It is easy to see when (7.10), (7.11) has one solution. If a solution λ of (7.14) is in $H^3(\Omega) \cap H^2_\circ(\Omega)$ then u defined by (7.13) is in $H^1(\Omega)$. Moreover, one can check that this $\{u, \lambda\}$ is a solution of (7.10), (7.11).

7.2 The Approximate Problem

Let $V_h \subset V$ and $M_h \subset M$ be two families of finite-dimensional spaces approximating V and M. We shall study the approximate problem:

Find

$$\{u_h, \lambda_h\} \in V_h \times M_h$$

such that

$$a(u_h, v_h) + b(v_h, \lambda_h) = (f, v_h) \ \ \forall\, v_h \in V_h, \tag{7.15}$$

$$b(u_h, \mu_h) = (\phi, \mu_h) \ \ \forall\, \mu_h \in M_h. \tag{7.15b}$$

99

Exercise 1. Show that the problem (7.15) leads to solving a linear system with matrix

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$$

where A is a $m \times m$ positive definite matrix and B is $n \times m$ matrix where $m = \dim V_h$ and $n = \dim M_h$.

Since V_h is finite-dimensional, the norms $\|\cdot\|_v$ and $\|\cdot\|_H$ on V_h are equivalent, that is there exists a function $S(h)$ such that

$$\| v_h \|_v \leq S(h) \| v_h \|_H \quad \forall\, v_h \in V_h. \tag{7.16}$$

We introduce the affine spaces

$$Z_h(\phi) = \{v_h \in V_h : \ b(v_h, \mu_h) = (\phi, \mu_h) \ \ \forall\, \mu_h \in M_h\}$$

$$Z(\phi) = \{v \in V : \ b(v, \mu) = (\phi, \mu) \ \ \forall\, \mu \in M\}$$

Exercise 2. Let $\phi = 0$. Show that (7.1) is equivalent to the problem:

Find

$$\left.\begin{aligned} &u \in Z = Z(0) \quad \text{such that} \\ &a(u, v) = (f, v) \ \forall\, v \in Z \end{aligned}\right\} \tag{7.17}$$

In the same way show that (7.15) is equivalent to:

Find

$$\begin{cases} u_h \in Z_h = Z_h(0) \quad \text{such that} \\ a(u_h, v_h) = (f, v_h) \ \ \forall\, v_h \in Z_h \end{cases} \tag{7.18}$$

The present framework allows us to deal with $Z_h \not\subset Z$ and hence we can consider non-conforming approximations of (7.17). 100

We now give an error bound in H-norm.

THEOREM 2. *Assuming that the continuous problem has at least one solution $\{u, \lambda\}$ one has the error bound*

$$\| u - u_h \|_H \leq \left(1 + \frac{|a|}{\alpha}\right) \inf_{v_h \varepsilon Z_h(\phi)} \| u - v_h \|_H + \frac{|b|}{\alpha} S(h) \inf_{\mu_h \varepsilon M_h} \| \lambda - \mu_h \|_M \tag{7.19}$$

Proof. Let $w_h = v_h - u_h$ and we have

$$a(w_h, w_h) = a(v_h - u, w_h) + a(u - u_h, w_h)$$

From (7.1) and (7.15), we obtain

$$a(u - u_h, w_h) = b(w_h, \lambda_h - \lambda) \ \ \forall\ w_h\ \varepsilon\ V_h.$$

We notice that for $v_h \varepsilon Z_h(\phi)$,

$$b(v_h - u_h, \mu_h) = 0 \ \ \forall\ \mu_h\ \varepsilon\ M_h$$

and hence

$$a(u - u_h, v_h - u_h) = b(v_h - u_h, \mu_h - \lambda) \ \ \forall\ v_h\ \varepsilon\ Z_h(\phi), \mu_h\ \varepsilon\ M_h.$$

Using the H-coerciveness of $a(\cdot\,,\cdot)$ and the continuity of $a(\cdot\,,\cdot)$ and $b(\cdot\,,\cdot)$ we obtain

$$\begin{aligned} \alpha \| v_h - u_h \|_H^2 &\leq |a| \| v_h - u_h \|_H \| v_h - u \|_H + |b| \| \lambda - \mu_h \|_M \| v_h - u_h \|_V \\ &\leq |a| \| v_h - u_h \|_H \| v_h - u \|_H + |b| \| \lambda - \mu_h \|_M\ S(h). \| v_h - u_h \|_H \end{aligned}$$

101 Hence

$$\| v_h - u_h \|_H \leq \frac{|a|}{\alpha} \| v_h - u \|_H + \frac{|b|}{\alpha} S(h) \| \lambda - \mu_h \|_M\ .$$

We get the desired result by noticing that

$$\| u - u_h \|_H \leq \| u - v_h \|_H + \| v_h - u_h \|_H$$

□

REMARK 4. *If $Z_h(0) \subset Z(0)$ then the error estimate* (7.19) *reduces to*

$$\| u - u_h \|_H \leq \left(1 + \frac{|a|}{\alpha}\right) \inf_{v_h \varepsilon Z_h(\phi)} \| u - v_h \|_H, \tag{7.20}$$

using the fact $b(v_h - u_h,\ \lambda_h - \lambda) = 0$ as $v_h - u_h\ \varepsilon\ Z_h(0) \subset Z(0)$.

When $\phi = 0$, the error estimate (7.20) *is obvious, inview of exercise 2. Then* (7.20) *is the error bound obtained in the conforming case.*

7.3 Application to the Stokes Problem

In what follows T_h will denote a regular family of triangulations of the polygonal domain Ω, ϑ_h will denote the set of vertices of T_h, m_h the set of mid side points and $\mathscr{E}_h$ the set of edges.

We consider the Stokes problem (example 1) where u is the velocity **102**
and λ is the pressure.

We shall choose for V_h a conforming $\mathbb{P}_2$ space and for M_h a piece-wise constant space, namely

$$V_h = \{v_h\ \varepsilon\ (C^\circ(\overline{\Omega}))^n : v_h|_K\ \varepsilon (\mathbb{P}_2(K))^n, K\ \varepsilon\ T_h, v_h = 0 \quad \text{on} \quad \partial\Omega\}$$

and

$$M_h = \{\mu_h\ \varepsilon\ L^2(\Omega) : \mu_h|_K\ \varepsilon\ \mathbb{P}_0(K), K\ \varepsilon\ T_h\}.$$

We notice that

$\dim V_h = n$ (# internal vertices + # internal edges) and choose Σ_h for the set of degrees of freedom in each component where

$$\Sigma_h = \{\delta_N,\ N\ \varepsilon \vartheta_h;\ M_\gamma,\ \gamma\ \varepsilon\ \mathscr{E}_h\},$$

$M_\gamma(p) = \frac{1}{|\gamma|} \int\limits_\gamma p\, ds$ denoting the *average* on the edge.

With this choice of Σ_h, the interpolation operator

$$\pi_h : (H^2(\Omega))^n \to V_h$$

defined by

$$\delta_N(\pi_h u) = \delta_N(u),\ N\ \varepsilon \vartheta_h;$$

$$M_\gamma(\pi_h u) = M_\gamma(u),\ \gamma\ \varepsilon\ \mathscr{E}_h,$$

will have some nice properties. Note that π_h is defined only on a subset of V since $u(N)$ is not defined for all u in $(H^1_\circ(\Omega))^n$; π_h is defined on
103 $V \cap (H^2(\Omega))^n$ since the functions in $(H^2(\Omega))^n$ are continuous.

Exercise 3. Let K be a triangle.

Let $P_K = \mathbb{P}_2(K)$ and

$$\Sigma_K = \{\delta_{a_i},\ M_{\gamma_i},\ 1 \leq i \leq 3\},$$

where a_i are the vertices of K and γ_i are edges of K. M_{γ_i} is defined by

$$M_{\gamma_i}(p) = \frac{1}{|\gamma_i|} \int\limits_{\gamma_i} p\ d\ s.$$

Show that Σ_K is P_K unisolvent.

We have

LEMMA 3. *One has*

$$\int\limits_K \operatorname{div}\,(\pi_h v)\,dx = \int\limits_K \operatorname{div}\ v\,dx\ \ \forall\ v\ \varepsilon\ (H^2(\Omega))^n$$

Proof. Indeed, by Green's formula,

$$\begin{aligned}\int\limits_K \operatorname{div}\,(\pi_h v)\,dx &= \int\limits_{\partial K} (\pi_h v.n)\,ds\\ &= \int\limits_{\partial K} v.n\,ds\end{aligned}$$

since n is constant on each side of K.

Applying again Green's formula we get the desired result. □

104 **Error Estimates for** $\|\ u - u_h\ \|_1$ If $u\varepsilon H^2(\Omega)$ (which is true when Ω is convex) then, since $u\varepsilon Z(0)$ (i.e. $\operatorname{div} u = 0$) and M_h contains functions

which are piecewise constant, we have, by Lemma 3, $\pi_h u \varepsilon Z_h(0)$. Hence we obtain

$$\inf_{v_h \varepsilon Z_h(0)} \| u - v_h \|_1 \leq \| u - \pi_h u \|_1 \leq \text{ch} \| u \|_2 .$$

If the solution $u \varepsilon H^3(\Omega)$ (which is unlikely since Ω is a polygon) the error bound becomes $\text{ch}^2 \| u \|_3$. Indeed, the interpolation operator π_h leaves invariant the polynomial space $\mathbb{P}_2(K)$ on each element K and the above error bound follows from Chapter 5.

However, we have

$$\inf_{u_h \varepsilon M_h} \| \lambda - \mu_h \|_0 \leq \text{ch} \| \lambda \|_1,$$

provided that the pressure $\lambda \varepsilon H^1(\Omega)$.

Finally, Theorem 2 gives

$$\| u - u_h \|_1 \leq \text{ch}(\| u \|_2 + \| \lambda \|_1),$$

which is only $0(h)$ due to the low degree of approximation for the multiplier.

Other Choices for V_h and M_h.

Let

$$Q(K) = \mathbb{P}_2(K) + \{\lambda_1 \ \lambda_2 \ \lambda_3\},$$

where $\lambda_1\lambda_2\lambda_3$ is called the *bubble* function. Note that the dimension of **105**
$Q(K)$ is 7.

The choice

$$V_h = \left\{ v_h \ \varepsilon \ (C^\circ(\overline{\Omega}))^2, v_h|_K \ \varepsilon \ (Q(K))^2, K \ \varepsilon \ T_h, v_h = 0 \quad \text{on} \quad \partial\Omega \right\}$$
$$M_h = \prod_K \mathbb{P}_1(K)$$

leads to the error estimates

$$\| u - u_h \|_1 \leq \text{ch}^2, \ \| u - u_h \|_0 \leq \text{ch}^3 .$$

(See CROUZEIX - RAVIART [14].

The choice

$$V_h = \left\{ v_h \ \varepsilon \ (C^\circ(\overline{\Omega}))^2 : v_h|_K \ \varepsilon \ (\mathbb{P}_2(K), K \ \varepsilon \ T_h, \ v_h = 0 \quad \text{on} \quad \partial\Omega \right\}$$
$$M_h = \left\{ \mu_h \ \varepsilon \ C^\circ(\overline{\Omega}) : \ \mu_h|_K \ \varepsilon \ \mathbb{P}_1(K), \ K \ \varepsilon \ T_h \right\},$$

in which M_h contains continuous piecewise linear functions, leads to the same error estimates. (See BERCOVIER-PIRONNEAU [3]).

This last method, due to TAYLOR-HOOD [42], is widely used by engineers.

7.4 Dual Error Estimates for $u - u_h$

Let

$$V_2 \hookrightarrow V \hookrightarrow V_0$$
$$M_1 \hookrightarrow M.$$

106 We denote by $\|\cdot\|_i$ the norms in V_i $(i = 0, 2)$ and $\|\cdot\|_1$ the norm in M_1. We assume that $V_0 \equiv V_0'$. (In practical applications V_0 will be a L^2 space) and that $H = V$ (i.e. $a(\cdot,\cdot)$ is V-coercive).

Let $g \varepsilon V_0$ and $\{w, \psi\} \varepsilon V \times M$ satisfy

$$a(v, w) + b(v, \psi) = (g, v) \ \forall \ v \ \varepsilon \ V, \tag{7.21}$$
$$b(w, \mu) = 0 \ \forall \ \mu \ \varepsilon \ M. \tag{7.21b}$$

We assume (regularity result) that

$$\| \ w \ \|_2 + \| \ \psi \ \|_1 \leq c \ \| \ g \ \|_0 \tag{7.22}$$

and

$$\inf_{v_h \varepsilon Z_h(\phi)} \| \ w - v_h \ \|_V \leq e(h) \ \| \ w \ \|_2, \tag{7.23}$$
$$\inf_{\mu_h \varepsilon M_h} \| \ \mu - \mu_h \ \|_M \leq e(h) \ \| \ \psi \ \|_1 \ . \tag{7.24}$$

A V_0-error estimate is given by

THEOREM 4. *Under the above assumptions, we have*

$$\| u - u_h \|_0 \leq ce(h) \left(\| u - v_h \|_V + \inf_{\mu_h \varepsilon M_h} \| \lambda - \mu_h \|_M \right).$$

Proof. We know that

$$\| u - u_h \|_0 = \operatorname*{Sup}_{g \varepsilon V_0} \frac{(g, u - u_h)}{\| g \|_0} \tag{7.25}$$

From (7.21) we have

$$(g, u - u_h) = a(u - u_h, w) + b(u - u_h, \psi) \tag{7.26}$$

Moreover, we have 107

$$a(u - u_h, v_h) + b(v_h, \lambda - \lambda_h) = 0 \quad \forall\ v_h\ \varepsilon\ V_h,$$
$$a(u - u_h, v_h) + b(v_h,\ \lambda - \mu_h) = 0 \quad \forall\ v_h\ \varepsilon\ Z_h(\phi), \quad \forall\ \mu_h\ \varepsilon\ M_h, \tag{7.27}$$

and

$$b(u - u_h, v_h) = 0 \quad \forall\ v_h\ \varepsilon\ M_h. \tag{7.28}$$

From (7.26), (7.21b), (7.27) and (7.28) we obtain

$$(g, u - u_h) = a(u - u_h, w - v_h) + b(w - v_h, \lambda - \mu_h) + b(u - u_h, \psi - v_h)$$
$$\forall\ v_h\ \varepsilon\ Z_h(\phi),$$
$$\forall\ \mu_h,\ v_h\ \varepsilon\ M_h.$$

Using the continuity of $a(\cdot\,,\cdot)$ and $b(\cdot\,,\cdot)$ we get

$$(g, u - u_h) \leq c\,(\| u - u_h \|_V \| w - v_h \|_V + \| w - v_h \|_V \| \lambda - \mu_h \|_M +$$
$$+ \| u - u_h \|_V \| \psi - v_h \|_M) \quad \forall\ v_h\ \varepsilon\ Z_h(\phi),\ \forall\ \mu_h, v_h\ \varepsilon\ M_h.$$

Taking the infimum over $v_h \varepsilon Z_h(\phi)$ and $\mu_h, v_h \varepsilon M_h$ and using (7.23), (7.24), we obtain

$$(g, u - u_h) \leq c[\| w \|_2 \left(\| u - u_h \|_V + \inf_{\mu_h \varepsilon M_h} \| \lambda - \mu_h \|_M \right) +$$
$$+ \| u - u_h \|_V \| \psi \|_1]\ e(h)$$

Finally, using the regularity result (7.22) and (7.25) we get the desired result. □

Application to Stokes Problem.

We choose

$$V_0 = (L^2(\Omega))^n, V_2 = (H^2(\Omega))^n,\ M_1 = H^1(\Omega).$$

108 From the error estimates of section 7.3, we have

$$\inf_{v_h \varepsilon Z_h(0)} \parallel w - v_h \parallel_1 \leq \text{ch} \parallel w \parallel_2$$

and

$$\inf_{\mu_h \varepsilon M_h} \parallel \psi - \mu_h \parallel_0 \leq \text{ch} \parallel \psi \parallel_1 .$$

The regularity result (7.22) is nothing but the regularity result for the Stokes problem.

Hence applying Theorem 4, we obtain L^2-error estimate.

$$\parallel u - u_h \parallel_0 \leq \text{ch}^2 (\parallel u \parallel_2 + \parallel \lambda \parallel_1) .$$

7.5 Nonconforming Finite Element Method for Dirichlet Problem

We recall that the variational formulation of the Dirichlet problem

$$\left.\begin{aligned} -\Delta u &= f \quad \text{in} \quad \Omega \\ u &= 0 \quad \text{on} \quad \partial\Omega \end{aligned}\right\} \tag{7.29}$$

is:

Find $u \varepsilon H^1_\circ(\Omega)$ such that

$$a(u, v) = (f, v) \ \ \forall\ v\ \varepsilon\ H^1_\circ(\Omega), \tag{7.30}$$

where

$$a(u, v) = \int_\Omega \nabla u . \nabla v = \sum_{K \varepsilon T_h} \int_K \nabla u . \nabla v,$$

$$(f, v) = \int_{\Omega} fv,$$

109 and T_h is a triangulation of Ω.

We like to consider the nonconforming finite element approximation of (7.30), namely

Find $u_h \varepsilon Z_h$ such that

$$a(u_h, v_h) = (f, v_h) \quad \forall\ v_h\ \varepsilon\ Z_h, \tag{7.31}$$

where

$Z_h = \{v_h : v_h|_K \varepsilon \mathbb{P}_1(K), K \varepsilon T_h, v_h$ is continuous

across the midside points of internal edges,

$v_h = 0$ on $\partial\Omega\}$.

Notice that $Z_h \not\subset H^1_\circ(\Omega)$.

We will construct a mixed finite element which is equivalent to (7.31). Multiplying the first equation in (7.29) by $v \varepsilon \prod_K H^1(K)$ and integrating we obtain, using integration by parts in each triangle K,

$$\sum_K \int_K \nabla u.\nabla v - \sum_K \int_{\partial K} (\nabla u.n)\ v = \int_{\Omega} fv.$$

This suggests

$$a(u, v) = \sum_K \int_K \nabla u.\nabla v, \tag{7.32}$$

$$b(v, \mu) = -\sum_K \int_K (\mu.n)\ v, \tag{7.33}$$

where μ belongs to some *suitable space*.

We have to construct finite-dimensional subspaces V_h and M_h such that the problem

$$\left.\begin{array}{l} \text{Find} \quad \{u_h, \lambda_h\}\ \varepsilon\ V_h \times M_h \quad \text{with} \\ a(u_h, v_h) + b(v_h, \lambda_h) = (f, v_h)\ \forall\ v_h\ \varepsilon\ V_h \\ b(u_h, \mu_h) = 0\ \forall\ \mu_h\ \varepsilon\ M_h \end{array}\right\} \tag{7.34}$$

is equivalent to (7.31). Here $a(\cdot\,,\cdot), b(\cdot\,,\cdot)$ are as in (7.32) and (7.33). 110

We take $V_h = \prod\limits_K \mathbb{P}_1(K)$.

It is easy to see that if $(\mu_h.n)$ is constant and continuous along internal edges, then $b(v_h, \mu_h) = 0 \;\; \forall v_h \varepsilon Z_h$.

Define

$$Q(K) \subset (\mathbb{P}_1(K))^2$$

by

$$Q(K) = \{q = (q_1, q_2) : q_1 = a + bx,\; q_2 = c + \text{ by}\}.$$

If $\alpha x + \beta y = \ell$ is the equation of an edge γ then $q.n$ is constant on γ for $q \varepsilon Q(K)$, where n is normal to γ. The set

$\sum_K = \{(q.n)(a_{ij}) : a_{ij}$ are mid points of the sides of $K\}$ is $Q(K)$-unisolvent. Hence

$$\begin{aligned} M_h &= \{q \;\varepsilon\; (L^2(\Omega))^2 : q|_K \;\varepsilon\; Q(K), K \;\varepsilon\; T_h,\ \text{div}\;\; q \;\varepsilon\; L^2(\Omega)\} \\ &= \{q \;\varepsilon\; (L^2)^2 : q|_K \;\varepsilon\; Q(K),\; K \;\varepsilon\; T_h,\; q.n \\ &\qquad \text{is continuous across the edges of} \quad T_h\} \end{aligned}$$

serves our purpose.

111 **Exercise 4.** With the above constructed V_h and M_h show that (7.31) and (7.34) are equivalent. Further show that $Z_h(0) = Z_h$.

$$(\text{Recall} \quad Z_h(0) = \{v_h \;\varepsilon\; V_h :\; b(v_h, \mu_h) = 0 \;\; \forall\; \mu_h \;\varepsilon\; M_h)$$

The continuous problem corresponding to (7.34) can be obtained as follows:

It is natural to take

$$V = \prod_K H^1(K).$$

When μ is smooth we can write

$$\begin{aligned} b(v, \mu) &= -\sum_K \int\limits_{\partial K} (\mu.n)v \\ &= -\sum_K \left(\int\limits_K \text{div}\;\; \mu.v + \int\limits_K \mu.\nabla v \right) \end{aligned}$$

Hence we take

$$M = \left\{\mu \,\varepsilon\, (L^2(\Omega))^2 : \text{ div } \mu \,\varepsilon\, L^2(\Omega)\right\}.$$

Thus the continuous problem is:

$$\left.\begin{array}{l} \text{Find} \quad \{u, \lambda\} \,\varepsilon\, V \times M \quad \text{such that} \\ a(u, v) + b(v, \lambda) = (f, v) \ \ \forall\, v \,\varepsilon\, V, \\ b(u, \mu) = 0 \ \ \forall\, \mu \,\varepsilon\, M, \end{array}\right\} \tag{7.35}$$

where

$$a(u, v) = \sum_K \int_K \nabla u . \nabla v.$$

We have the characterisation:

LEMMA 5.

$$Z = \{v \,\varepsilon\, V : \ b(v, \mu) = 0 \ \ \forall\, \mu \,\varepsilon\, M\} = H^1_\circ(\Omega).$$

112

Proof. Let $v \varepsilon \mathscr{D}(\Omega)$. Then

$$\begin{aligned} b(v, \mu) &= -\sum_K \left(\int_K \text{div } \mu . v + \int_K \mu . \nabla v \right) \\ &= -\left(\int_\Omega \text{div } \mu . v + \int_\Omega \mu . \nabla v \right) \\ &= -\langle \text{div } \mu, v \rangle - \langle \mu, \ \nabla v \rangle \\ &= -\langle \text{div } \mu, v \rangle + \langle \text{div } \mu, v \rangle \\ &= 0. \end{aligned}$$

Since $b(\cdot\,,\cdot\,)$ is continuous on $V \times M$ and $\mathscr{D}(\Omega)$ is dense in $H^1_\circ(\Omega)$ in the $\|\cdot\|_1$ norm topology, we obtain

$$H^1_\circ(\Omega) \subset Z$$

We have to prove the other inclusion. Let $v \varepsilon Z$. Define

$$v_i \quad \text{by} \quad v_i|_K = \frac{\partial}{\partial x_i}(v|_K),\ \forall\ K\ \varepsilon\ T_h.$$
$$\text{Then} \quad v_i\ \varepsilon\ L^2(\Omega).$$

Let $\phi \varepsilon \mathscr{D}(\Omega)$. Then

$$\begin{aligned}
\langle v_1, \phi\rangle &= \sum_K \int_K v_1\phi = \sum_K \int_K \frac{\partial v}{\partial x_1}\phi \\
&= \sum_K \left(-\int_K v\frac{\partial \phi}{\partial x_1} + \int_{\partial K} v\ \phi\ n_1 \right) \\
&= -\left\langle v, \frac{\partial \phi}{\partial x_1} \right\rangle + \sum_K \int_{\partial K} v\ \phi\ n_1.
\end{aligned}$$

113 Since $v\varepsilon Z$, $b(v,\mu) = 0\ \ \forall \mu \varepsilon M$. Taking $\mu = (\phi, 0)$, we obtain

$$\begin{aligned}
0 = b(v,\mu) &= -\left(\sum_K \int_K \text{div}\ \mu.v + \int_K \nabla v.\mu \right) \\
&= -\sum_K \int_{\partial K} (\mu.n)v \quad \text{since} \quad \mu \quad \text{is smooth} \\
&= -\sum_K \int_{\partial K} \phi v\ n_1.
\end{aligned}$$

Therefore,
$$\langle v_1, \phi\rangle = -\left\langle v, \frac{\partial \phi}{\partial x_1} \right\rangle.$$

Hence $v_1 = \frac{\partial v}{\partial x_1}$ in $\mathscr{D}'$. Similarly we have $v_2 = \frac{\partial v}{\partial x_2}$ in $\mathscr{D}'$. Therefore $v\varepsilon H^1(\Omega)$. Further

$$0 = b(v,\mu) = -\sum_K \int_{\partial K} (\mu.n)v\ \forall\ \mu\ \varepsilon\ (H^1(\Omega))^2 \subset M$$

This implies $v = 0$ on $\partial\Omega$. Thus $v \varepsilon H^1_\circ(\Omega)$. Hence $Z \subset H^1_\circ(\Omega)$.

If $\{u, \lambda\}$ is a solution of (7.35) then Lemma 5 and the second equation in (7.35) imply that $u \varepsilon H^1_\circ(\Omega)$. The first equation in (7.35) gives that $-\Delta u = f$ in $\mathscr{D}'$. Thus if $\{u, \lambda\}$ is a solution of (7.35), then u is the solution of the Standard Dirichlet problem:

$$\begin{cases} -\Delta u & = f \quad \text{in} \quad \Omega \\ u & = 0 \quad \text{on} \quad \partial\Omega, \end{cases} \tag{7.36}$$

and λ is the Lagrange multiplier for the continuity constraint for u from 114
one element to the other; (7.35) is called the *primal hybrid formulation* of the Dirichlet problem.

If u is the solution of the Dirichlet problem (7.36) then $\{u, -\nabla u\}$ is a solution of (7.35). Note that when λ is smooth then $b(\cdot\,,\cdot)$ contains only the trace of $\lambda.n$ on the edges γ, so that λ is certainly not unique and the Brezzi condition does not hold. □

Error Estimates for $u - u_h$.

We notice that the interpolation operator

$$\pi_h : H^2(\Omega) \to Z_h$$

such that $\pi_h u = u$ at the mid side points of T_h satisfies

$$\| u - \pi_h u \|_{1,K} \leq \text{ch} \| u \|_{2,K} .$$

Therefore

$$\| u - \pi_h u \|_V \leq \text{ch} \| u \|_2 . \tag{7.37}$$

On the other hand, we define

$$\pi'_h : (H^1(\Omega))^2 \to M_h$$

on γ by

$$(\pi'_h \mu).n = \frac{1}{|\gamma|} \int_\gamma \lambda.n$$

We now state a theorem whose proof can be found in RAVIART-THOMAS [38].

THEOREM 6. *There exists a constant c such that* 115

$$\| \mu - \pi'_h \mu \|_M \leq \text{ch} (\| \mu \|_1 + \| \text{div } \mu \|_1), \tag{7.38}$$

$$\| \pi'_h \mu \|_M \leq c \| \mu \|_1 .$$

Using Theorem 2, (7.37) *and* (7.38) *we obtain*

$$\| u - u_h \|_V \leq \text{ch} \| u \|_2 . \tag{7.39}$$

The dual error estimate of section 7.4 together with (7.39) *gives*

$$\| u - u_h \|_0 \leq \text{ch}^2 \| u \|_2 \tag{7.40}$$

REMARK 5. *In practice, one solves the approximate problem;*

$$\begin{aligned} &\textit{Find} \quad u_h \in Z_h \quad \textit{such that} \\ &a(u_h, v_h) = (f, v_h) \;\; \forall\, v_h \in Z_h, \end{aligned}$$

since Z_h is a simple finite element space. The fact that $Z_h \not\subset H^1_\circ(\Omega)$ has no importance for practical purposes. The same Z_h can be used to approximate the Stokes problem. One chooses $V_h = (Z_h)^2$ (for the velocities) and piecewise constant pressure; i.e. M_h as in section 7.3. For the error analysis we refer the reader to CROUZEIX-RAVIART [14].

REMARK 6. *Other primal hybrid finite elements can be obtained with other choices for $V_h \mathscr{E} M_h$, some of them giving other well known non-conforming finite elements. (See RAVIART-THOMAS [37]).*

7.6 Approximate Brezzi Condition

116 We assume that the approximate Brezzi condition

$$\operatorname*{Sup}_{v_h \varepsilon V_h} \frac{b(v_h, \mu_h)}{\| v_h \|_V} \geq \gamma \| \mu_h \|_M \;\; \forall\, \mu_h \,\varepsilon\, M_h, \tag{7.41}$$

where γ is independent of h, holds. The approximate Brezzi condition guarantees a unique solution for the approximate problem. Notice that the continuous Brezzi condition need not imply the approximate Brezzi condition.

We have the following result:

THEOREM 7. *Under the assumption* (7.41) *there exists a constant c_1 such that*

$$\underset{v_h \varepsilon V_h(\phi)}{\text{Inf}} \| u - v_h \|_V \leq c_1 \underset{w_h \varepsilon V_h}{\text{Inf}} \| u - w_h \|_V \tag{7.42}$$

Proof. We will show that for each $w_h \varepsilon V_h$ there exists a $v_h \varepsilon Z_h(\phi)$ such that

$$\| u - v_h \| \leq \ c \ \| u - w_h \|,$$

where c is a constant. This will imply (7.42).

Let $w_h \varepsilon V_h$. Let $\{y_h, v_h\} \varepsilon V_h \times M_h$ be the solution of

$$(y_h, v_h)_V + b(v_h, v_h) = 0 \ \ \forall \ v_h \ \varepsilon \ V_h,$$
$$b(y_h, \mu_h) = b(u - w_h, \mu_h) \ \ \forall \ \mu_h \ \varepsilon \ M_h.$$

Using the approximate Brezzi condition and the continuity of $b(,)$, **117**
it is easy to prove that

$$\| v_h \|_M \leq 1/\gamma \ \ \| y_h \|_V,$$
$$\| y_h \|_V \leq |b|/\gamma \ \ \| u - w_h \|_V \ .$$

Let $v_h = y_h + w_h$. Then

$$b(v_h, \mu_h) = b(u - w_h, \mu_h) + b(w_h, \mu_h)$$
$$= b(u, \mu_h).$$

Hence $v_h \varepsilon Z_h(\phi)$. Further

$$\| u - v_h \|_V \leq \| u - w_h \|_V + \| y_h \|_V$$
$$\leq (1 + |b|/\gamma) \ \ \| u - w_h \|_V \ .$$

Hence the theorem. □

We now give an error estimate for the multiplier.

THEOREM 8. *If* (7.41) *holds then one has the error estimate*

$$\| \lambda - \lambda_h \|_M \leq c \left(\| u - u_h \|_H + \underset{\mu_h \varepsilon M_h}{\text{Inf}} \| \lambda - \mu_h \|_M \right) \tag{7.43}$$

Proof. We have

$$\begin{aligned} b(v_h, \lambda_h - \mu_h) &= b(v_h, \lambda_h - \lambda) + b(v_h, \lambda - \mu_h) \\ &= a(u - u_h, v_h) + b(v_h, \lambda - \mu_h) \end{aligned}$$

From (7.41) we obtain

$$\begin{aligned} \gamma \parallel \lambda_h - \mu_h \parallel_M &\leq \operatorname*{Sup}_{v_h \varepsilon V_h} \frac{b(v_h, \lambda_h - \mu_h)}{\parallel v_h \parallel_V} \\ &\leq |a|c \parallel u - u_h \parallel_H + |b| \parallel \lambda - \mu_h \parallel_M . \end{aligned} \tag{7.44}$$

118 Further

$$\parallel \lambda - \lambda_h \parallel_M \leq \parallel \lambda - \mu_h \parallel_M + \parallel \lambda_h - \mu_h \parallel_M \tag{7.45}$$

From (7.44) and (7.45) we obtain (7.43). □

We now give a practical way of verifying the approximate Brezzi condition (7.41).

LEMMA 9. *If $b(\cdot\,,\cdot)$ satisfies continuous Brezzi condition and if*

$$\Lambda_h : V \to V_h$$

is such that

$$b(v - \Lambda_h v, \mu_h) = 0 \;\; \forall\, \mu_h \;\varepsilon\; M_h \tag{7.46}$$

(in fact Λ_h maps $Z(\phi)$ into $Z_h(\phi)$) and

$$\parallel \Lambda_h v \parallel_V \leq c \parallel v \parallel_V, \tag{7.47}$$

then (7.41) *holds with $\gamma = \beta/c$.*

Exercise 5. Prove Lemma 9.

For further details see FORTIN [18].

7.7 Dual Error Estimate for the Multiplier

119 This section is an analogue to Section 7.4. We assume that $V_2 \hookrightarrow V$ and $M_3 \hookrightarrow M \hookrightarrow M_0$. Further, we assume that $M_0' = M_0$ and that for $g \varepsilon M_0$, the problem:

Find $\{w, \psi\}\ \varepsilon V \times M$ such that

$$a(v, w) + b(v, \psi) = 0 \quad \forall\ v\ \varepsilon V \tag{7.48}$$

$$b(w, \mu) = (g, \mu) \quad \forall\ \mu\ \varepsilon\ M \tag{7.48b}$$

has one solution such that

$$\| w \|_2 + \| \psi \|_3 \leq c \| g \|_0 \tag{7.49}$$

We also assume that

$$\operatorname*{Inf}_{v_h \varepsilon V_h} \| w - v_h \|_V \leq e(h) \| w \|_2, \tag{7.50}$$

$$\operatorname*{Inf}_{\mu_h \varepsilon M_h} \| \psi - \mu_h \|_M \leq \varepsilon(h) \ \| \psi \|_3 . \tag{7.51}$$

Here $\| \ \|_0$, $\| \ \|_2$, $\| \ \|_3$ denote the norms in M_0, V_2, M_3 respectively. The dual error estimate is given by

THEOREM 10. *One has*

$$\| \lambda - \lambda_h \|_0 \leq ce(h) (\| \lambda - \lambda_h \|_M + \| u - u_h \|_H) +$$
$$+ c\ \varepsilon(h) \left(S(h) \| u - u_h \|_H + \operatorname*{Inf}_{y_h \varepsilon V_h} (\| u - y_h \|_V + S(h) \| u - y_h \|_H) \right) \tag{7.52}$$

Proof. We have

$$\| \lambda - \lambda_h \|_0 = \operatorname*{Sup}_{g \varepsilon M_0} \frac{(g, \lambda - \lambda_h)}{\| g \|_0}$$

and **120**

$$(g, \lambda - \lambda_h) = b(w, \lambda - \lambda_h)$$
$$= b(w - v_h, \lambda - \lambda_h) - a(u - u_h, v_h).$$

where the above is obtained from (7.1) and (7.15). Using $b(u-u_h, v_h) = 0 \ \forall\ v_h\ \varepsilon\ M_h$ and (7.48), we obtain

$$(g, \lambda - \lambda_h) = b(w - v_h, \lambda - \lambda_h) + a(u - u_h, w - v_h) + b(u - u_h, \psi - v_h)$$

This implies,

$$(g, \lambda-\lambda_h) \leq c(e(h)(\| \lambda-\lambda_h \|_M + \| u-u_h \|_H)+ \| u-u_h \|_V \varepsilon(h)) \| g \|_0 .$$

Here (7.49) - (7.51) are used.

To estimate $\| u - u_h \|_V$, we remark that

$$\begin{aligned}\| u - u_h \|_V &\leq \| u - y_h \|_V + \| y_h - u_h \|_V,\\ \| y_h - u_h \|_V &\leq S(h) \| y_h - u_h \|_H\\ &\leq S(h)(\| y_h - u \|_H + \| u - u_h \|_H).\end{aligned}$$

Finally,

$$\begin{gathered}\| \lambda - \lambda_h \|_0 \leq c[e(h)(\| \lambda - \lambda_h \|_M + \| u - u_h \|_H)+\\ \varepsilon(h)(S(h) \| u - u_h \|_H + \operatorname*{Inf}_{y_h \varepsilon V_h} (\| u - y_h \|_V + S(h) \| u - y_h \|_H))]\end{gathered}$$

□

7.8 Application to Biharmonic Problem

121 We shall now study a finite element approximation to the biharmonic problem. (Example 2, Section 7.1). We recall that in the variational formulation of the biharmonic problem

$$\begin{aligned}\Delta^2 \lambda &= -\phi \quad \text{in} \quad \Omega,\\ \lambda &= \frac{\partial \lambda}{\partial n} = 0 \quad \text{on} \quad \partial\Omega,\end{aligned}$$

we have

$$V = H^1(\Omega),\ M = H^1_\circ(\Omega), H = L^2(\Omega),$$

$$a(u,v) = \int_\Omega uv\,dx,\ b(v,\mu) = -\int_\Omega \nabla v.\nabla\mu$$

and $f = 0$.

In the terminology of hydrodynamics $\lambda = \psi$ is called the *stream function* and $w = -\Delta\psi = u$ is called the *Vortex* function.

For both V_h and M_h we shall use standard Lagrange finite element space of degree k:

$$V_h = \left\{v_h\ \varepsilon\ C^\circ(\overline{\Omega}) : v_h|_K\ \varepsilon\ \mathbb{P}_k(K),\ \forall\ K\ \varepsilon\ T_h\right\}$$

and

$$M_h = \{\mu_h\ \varepsilon\ V_h : \mu_h = 0 \quad \text{on} \quad \partial\Omega\}.$$

We immediately notice that the approximate Brezzi condition (7.41) holds. Indeed,

$$\operatorname*{Sup}_{v_h \varepsilon V_h} \frac{\int_\Omega \nabla v_h.\nabla\mu_h}{\| v_h \|_1} \geq \frac{\int_\Omega |\nabla\mu_h|^2}{\| \mu_h \|_1} \geq \gamma \| \mu_h \|_1 .$$

since $M_h \subset V_h$. Here γ is the constant occurring in Poincare's inequality. 122

Moreover, assuming that the triangulation satisfies the uniformity condition

$$h \leq c\rho_K,\ \forall\ K\ \varepsilon\ T_h,$$

where ρ_K denotes the diameter of the largest ball included in K and c is a constant independent of h, one always has

$$\| v \|_1 \leq \frac{c}{\min\limits_{K\varepsilon T_h} \rho_K} \| v_h \|_0 .$$

Therefore one has the inverse inequality $\| v_h \|_1 \leq \frac{c}{h} \| v_h \|_0$ which gives an evaluation of $S(h)$.

We can state a convergence result for $k > 2$.

THEOREM 11. *If $\lambda\varepsilon H^{m+1}(\Omega)$ and $u\varepsilon H^m(\Omega)$ then one has*

$$\| u - u_h \|_0 + \| \lambda - \lambda_h \|_1 \leq \text{ch}^{m-1} (\| u \|_m + \| \lambda \|_{m+1})$$
$$\textit{for} \quad m = 2, \ldots, k \quad \textit{and}$$
$$\| \lambda - \lambda_h \|_0 \leq \text{ch}^m (\| u \|_m + \| \lambda \|_{m+1})$$

The above result is a consequence of Theorems 2, 7, 8, 10. We note that the result is not optimal since with polynomials of degree k, we should get an error bound in h^k for $\| \lambda - \lambda_h \|_1$ provided that $\lambda \varepsilon H^{k+1}$.

123 These results have been recently improved by SCHOLTZ [41] who is able to give an error estimate in the case $k = 1$.

Note that the matrix A of the bilinear form on V_h *has to be computed exactly in this case*. (The use of a 1-point formula for the computation of $\int_K uv\,dx$ leads to a diagonal matrix A but there is no convergence).

REMARK 7. *Due to the inclusion $M_h \subset V_h$, the matrix B of the bilinear form $b(\cdot\,,\cdot)$ on $V_h \times M_h$ has the particular form $B = (B_1, B_2)$ where B_1 is the matrix of $b(\cdot\,,\cdot)$ on $M_h \times M_h$ and therefore B_1 is invertible. The linear system corresponding to the approximate problem can be written as*

$$\begin{pmatrix} A_1 & A_2 & B_1^T \\ A_2^T & A_3 & B_2^T \\ B_1 & B_2 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \phi \end{pmatrix}$$

One can eliminate u_1 from the last equation ($u_1 = -B_1^{-1}\ B_2\ u_2$) and λ from the first one. This gives a linear system of equations in u_2 which can be solved by any of the standard method.

The advantage is that the size of the linear system is $p \times p$ where p is the number of boundary points which is relatively small.

7.9 General Numerical Methods for the Solution of the Approximate Problem

124 As we have noticed in Exercise 1, the approximate problem is equivalent to solving

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ \phi \end{pmatrix} \tag{7.53}$$

where the matrix on the left is invertible (provided that B has a maximal rank, i.e. the approximate Brezzi condition holds at least for γ dependent on h) and symmetric if $a(\cdot\,,\cdot)$ is symmetric but not positive.

Alternatively one can use the matrix

$$\begin{pmatrix} A & B^T \\ -B & 0 \end{pmatrix}$$

which is positive but not symmetric.

a) **Solution of the Linear System** (7.53) **by Direct Methods.** The Gaussian elimination of (7.53) can be performed without pivoting. However, due to storage considerations, it is often much better to permute rows and columns to get a band matrix after a suitable ordering of the unknowns. (For finite elements, we order the unknowns following the ordering that we give to the associated nodes, with no distinction between λ_i's and u_i's). In this case a strategy of partial pivoting may be necessary.

b) **Solution of the Approximate Problem by Penalty Methods.** **125**
First solve

$$(A + 1/\in\ B^T B)\, u^{\in} = f + 1/\in\ B^T\ \phi,$$

and then find

$$\lambda^{\in} = 1/\in\ (Bu^{\in} - \phi).$$

If B has maximal rank the error is only $0(\in)$ (see proof of Theorem 1), which is certainly small compared to the discretization error if $\in= 10^{-4}$ or 10^{-6}.

However, the condition number of the matrix

$$A + 1/\in\ B^T B$$

might be quite big and it is wise in such a case to use direct methods and then to compute explicitly the matrix $B^T B$, which is easy only if

$$M_h = \underset{K}{\pi}(\mathbb{P}_k(K))^d$$

for some positive k and d; in otherwords, no continuity requirements has to be asked for λ_h between two elements. (Then $B^T B$

can be computed by assembling some stiffness matrices). Otherwise the computation of $B^T B$ is too costly. Note that this method is possible even if B has not a maximal rank (and the error is then only $0(\sqrt{\in})$).

c) **Solution of the Problem by Iterative Methods.** The conjugate gradient method has been successfully extended to the case of
126 matrices such as (7.53) by PAIGE **and** SAUNDERS [35].

Another way of applying the conjugate gradient method is to notice that u can be eliminated from (7.53):

$$u = A^{-1} f - A^{-1} B^T \lambda.$$

Then we get

$$C\lambda = b$$

where

$$C = BA^{-1}\ B^T,\ b = BA^{-1} f - \phi.$$

As C is symmetric and positive definite (if $a(\cdot\,,\cdot)$ is symmetric) then the conjugate gradient method can be applied to the matrix C, but *one has never to compute the matrix C explicitly*. (This is too costly *unless* A is block diagonal and therefore V_h is a finite element space with no continuity requirements between 2 elements. Let A_K denote the element matrix of $a(\cdot\,,\cdot)$ on K and B_K that of $b(\cdot\,,\cdot)$; the matrix C is computed by assembling the $B_K A_K^{-1} B_K^T$'s. This is the case for hybrid elements).

Indeed what one needs for the conjugate gradient method is to be able to compute $y = Cz$ for any column vector z and this is done in the following way:

$$\begin{aligned} \text{Compute} \quad z_1 &= B^T\ ; \\ \text{Solve} \quad Az_2 &= z_1\ ; \\ \text{Compute} \quad y &= Bz_2\ . \end{aligned}$$

127 Note that it is not necessary for B to have maximal rank.

7.10 Equilibrium Elements for the Dirichlet Problem

Let us consider the following problem:

Find $\{u, \lambda\}$ such that

$$\begin{cases} u - \nabla\lambda = 0 \quad \text{in} \quad \Omega \\ \operatorname{div} u = \phi \quad \text{in} \quad \Omega \\ \lambda = 0 \quad \text{in} \quad \partial\Omega \end{cases} \tag{7.54}$$

If $\{u, \lambda\}$ is a solution of (7.54) then λ is the solution of the Standard Dirichlet Problem:

$$\begin{cases} \Delta\lambda = \phi \quad \text{in} \quad \Omega \\ \lambda = 0 \quad \text{on} \quad \partial\Omega \end{cases} \tag{7.55}$$

Multiplying (7.54) by $v \in (L^2(\Omega))^2$ and using integration by parts we obtain an equivalent problem:

Find $\{u, \lambda\} \in (L^2(\Omega))^2 \times H^1_\circ(\Omega)$ such that

$$\begin{cases} a(u, v) + b(v, \lambda) = 0 \ \ \forall\, v \ \in \ (L^2(\Omega))^2, \\ b(u, \mu) = \int\limits_\Omega \phi\mu \ \ \forall\, \mu \ \in \ H^1_\circ(\Omega), \end{cases}$$

where

$$a(u, v) = \int\limits_\Omega u.v \, dx,$$

$$b(v, \mu) = -\int\limits_\Omega v.\nabla\mu \, dx.$$

If $M_h = C^\circ(\overline{\Omega}) \cap \underset{K}{\pi}\mathbb{P}_k(K)$ then a natural choice for V_h is **128**

$$V_h = \underset{K}{\pi}(\mathbb{P}_{k-1}(K))^2.$$

Note that the operator ∇ maps M_h into V_H. This implies that the approximate Brezzi condition holds.

Problem (7.54) can be formulated in another way also.
Find $\{u, \lambda\} \in H(\text{div}, \Omega) \times L^2(\Omega)$ such that

$$\int\limits_{\Omega} uv + \int\limits_{\Omega} \lambda \operatorname{div} v = 0 \ \ \forall\, v \in H\,(\text{div}, \Omega),$$
$$\int\limits_{\Omega} \mu \operatorname{div} u = (\mu, \phi) \ \ \forall\, \mu \in L^2(\Omega),$$

where
$$H\,(\text{div}, \Omega) = \{v \in (L^2(\Omega))^2 : \operatorname{div} v \in L^2(\Omega)\}.$$

We notice that
$$a(u, v) = \int\limits_{\Omega} u.v\,dx$$
is coercive on
$$H = (L^2(\Omega))^2.$$

To prove that $b(\cdot\,,\cdot)$ satisfies Brezzi condition, we use the fact that *if $\mu \varepsilon L^2(\Omega)$ then the solution ψ of the Dirichlet problem*:

$$\begin{aligned} \Delta\psi &= \mu \quad \text{in} \quad \Omega \\ \psi &= 0 \quad \text{on} \quad \partial\Omega, \end{aligned}$$

129 *satisfies* $\| \psi \|_1 \leq c \| \mu \|_0$.
Therefore,

$$\operatorname*{Sup}_{v\varepsilon(L^2(\Omega))^2} \frac{\int\limits_{\Omega} \mu. \operatorname{div} v}{\| v \|_V} \geq \frac{\int\limits_{\Omega} \mu \operatorname{div} (\nabla\psi)}{\| \nabla\psi \|_V} \geq \frac{\int\limits_{\Omega} \mu^2\,dx}{c \| \mu \|_0} = \frac{1}{c} \| \mu \|_0 \,.$$

As approximate spaces, we choose

$$\begin{aligned} V_h = \{v \,\varepsilon\, V : v|_K \,\varepsilon\, Q(K),\ v.n \quad &\text{is} \\ \text{continuous across the sides of} \quad &T_h\}. \end{aligned}$$

(See Section 7.5 for the definition of $Q(K)$).

$$M_h = \underset{K}{\pi}\, \mathbb{P}_0(K).$$

As div : $V_h \to M_h$, we see that $Z_h(0) \subset Z(0)$. (Hence the name *equilibrium* elements: u_h will satisfy equilibrium equations $\operatorname{div} u_h = \phi$ for ϕ piecewise constant).

We may then apply the error estimate derived in Section 7.2 and use the improvement given in Remark 4, since $Z_h(0) \subset Z(0)$. We shall choose $v_h = \pi'_h v$ where π'_h is the interpolation operator defined in Section 7.5.

Indeed, we have

$$\int_\gamma (\pi'_h\, v).n\, ds = \int_\gamma v.n\, ds \quad \text{for each edge} \quad \gamma \quad \text{of} \quad T_h.$$

Therefore, 130

$$\begin{aligned}\int_\Omega \mu_h \operatorname{div} \pi'_h\, v &= \sum_K \int_{\partial K} (\pi'_h v).n\mu_h\, d\Gamma \\ &= \sum_K \int_{\partial K} (v.n)\, \mu_h\, d\Gamma = \int_\Omega \mu_h \operatorname{div}\, v\, dx \ \ \forall\, \mu_h \,\varepsilon\, M_h.\end{aligned}$$

This shows that $v \varepsilon Z(\phi) \Rightarrow \pi'_h\, v\, \varepsilon\, Z_h(\phi)$.

Finally we get

$$\| u - u_h \|_0 \leq \ c \ \| u - \pi'_h\, u \|_0 \leq \text{ch},$$

where we have used the estimate given in Theorem 6.

To get an error estimate for $\| \lambda - \lambda_h \|_0$, we shall make use of the results in Section 7.6 and construct the operator π_h occurring in Lemma 9.

Let v in V be given and ϕ satisfy

$$\begin{aligned}\Delta\phi &= \operatorname{div}\, v \quad \text{in} \quad \Omega,\\ \phi &= 0 \quad \text{on} \quad \partial\Omega.\end{aligned}$$

Let

$$\Lambda : V \to (H^1(\Omega))^2$$

be defined by

$$\Lambda v = \nabla \phi.$$

We have (regularity result)

$$\| \Lambda v \|_1 \leq c \| \operatorname{div} \ v \|_0,$$

so that

$$\Lambda_h = \pi'_h \Lambda$$

131 satisfies

$$\| \Lambda_h v \|_M \leq c \| \Lambda v \|_1 \leq c \| \operatorname{div} \ v \|_0$$

and

$$b(\mu_h, \pi'_h \Lambda v) = b(\mu_h, \Lambda v) = b(\mu_h, v),$$

where we used the definitions of π'_h and Λv.

Thus Λ_h satisfies the conditions required in Lemma 9. Hence we have

$$\| \lambda - \lambda_h \|_0 \leq \text{ch},$$

by Theorem 8.

For further details about equilibrium elements the reader can refer the thesis of J.M. THOMAS, 1977.

REMARK 8. *If we replace $Q(K)$ by $(\mathbb{P}_1(K))^2$ we get a finite element with 6 degrees of freedom instead of 3 (2 values of v.n on each side). The interpolation operator π'_h is defined with the help of the degrees of freedom and has the same properties.*

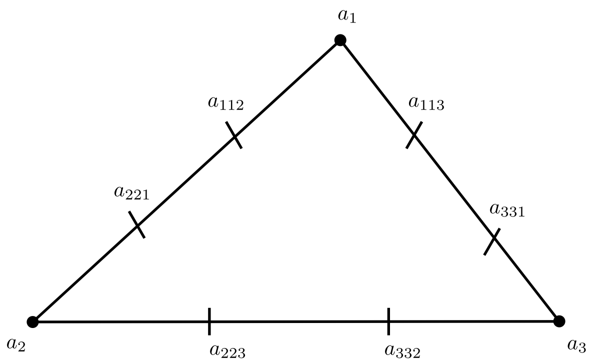

Figure 7.1:

In fact, π'_h is defined by **132**

$$\int_\gamma p(\pi'_h v).n\, ds = \int_\gamma p(v.n)\, ds \quad \forall\ p\ \varepsilon\ \mathbb{P}_1(\gamma)$$

However, the error estimates now become

$$\| u - u_h \|_0 \leq \text{ch}^2$$
$$\| \lambda - \lambda_h \|_0 \leq \text{ch}$$

REMARK 9. *The present finite element method can be extended to the elasticity equation where v represents the stress tensor σ_{ij}. The difficulty lies in the required symmetry of σ_{ij} but can be surmounted. See C. JOHNSON-B. MERCIER [25] and AMARA-THOMAS[2].*

REMARK 10. Aposteriori Error Estimate. *Let us consider the following optimization problem:*

$$\begin{aligned} &\text{Inf } J(v,\mu) \\ &v\ \varepsilon\ Z(\phi) \\ &\mu\ \varepsilon\ M \end{aligned}$$

where $J(v,\mu) = 1/2 \| v - \nabla\mu \|^2$. Clearly the optimal value is zero and corresponds to $v = u$ and $\mu = \lambda$ solution of (7.54). *Since $v \varepsilon Z(\phi)$, we* *also have*

$$J(v, \mu) = 1/2 \parallel \nabla\mu \parallel^2 + (\phi, \mu) + 1/2 \parallel v \parallel^2 .$$

Since $J(u, \lambda) \leq J(v_h, \lambda)\ \forall\ v_h\ \varepsilon\ Z(\phi)$, *we obtain*

$$1/2 \parallel u \parallel^2 \leq 1/2 \parallel v_h \parallel^2 .$$

133 *Adding*

$$1/2 \parallel \nabla\mu_h \parallel^2 + (\phi, \mu_h)$$

to both sides, where $\mu_h \varepsilon M$, *we get*

$$J(u, \mu_h) \leq J(v_h, \mu_h)\ \ \forall\ \mu_h\ \varepsilon\ M,\ v_h\ \varepsilon\ Z(\phi).$$

That is,

$$\parallel \nabla(\lambda - \mu_h) \parallel_0 \leq \parallel v_h - \nabla\mu_h \parallel_0 \quad \forall\ v_h\ \varepsilon\ Z(\phi).$$

Suppose that μ_h *is a solution of Dirichlet problem with a conforming finite element method, then an upper bound for the error in the energy norm is given by*

$$\parallel v_h - \nabla\mu_h \parallel_0$$

where

$$v_h\ \varepsilon\ Z_h(\phi)$$

is arbitrary. One can choose $v_h = u_h$, *a solution of the present equilibrium finite element approximation to Dirichlet problem.*

7.11 Equilibrium Elements for the Plate Problem

We recall that the equations of the plate problem are:

Find σ_{ij}, w such that

$$\sigma_{ij} = \lambda\Delta w\delta_{ij} + 2\mu\frac{\partial^2 w}{\partial x_i\, \partial x_j} \text{ in } \Omega \subset \mathbb{R}^2 \tag{7.57}$$

$$\frac{\partial^2\sigma_{ij}}{\partial x_i\, \partial x_j} = f \text{ in } \Omega \tag{7.57b}$$

$$w = 0 \text{ on } \partial\Omega \tag{7.57c}$$

134 and

$$\begin{cases} \frac{\partial w}{\partial n} = 0 \quad \text{on} \quad \partial\Omega \quad \text{clamped case} \\ \sigma_{ij}\, n_i\, n_j = 0 \quad \text{on} \quad \partial\Omega \quad \text{simply supported plate problem} \end{cases} \tag{7.57d}$$

The summation convention is used in the above equations.

Exercise 6. Show that (7.57) is equivalent to a biharmonic problem for w alone, and write the variational formulation of that biharmonic problem. Notice that in the clamped case several bilinear forms may be chosen unlike in the simply supported case.

It is easy to check that (7.57) is equivalent to

$$\alpha\sigma_{ij} + \beta(\sigma_{kk})\,\delta_{ij} = \frac{\partial^2 w}{\partial x_i\, \partial x_j}, \tag{7.58}$$

where

$$\alpha = 1/2\mu \quad \text{and} \quad \beta = \frac{-\lambda}{2\mu(2\lambda + 2\mu)}.$$

Let

$$a(\sigma, \tau) = \int_\Omega (\alpha\sigma_{ij}\tau_{ij} + \beta(\sigma_{kk})(\tau_{\ell\ell}))\, dx$$

and

$$Dw = \left(\frac{\partial^2 w}{\partial x_i\, \partial x_j}\right).$$

Then problem (7.57) is equivalent to: 135

$$\text{Find} \quad \sigma\varepsilon\ (L^2(\Omega))^4_s,\ w\ \varepsilon\ H^2_\circ(\Omega) \quad \text{such that}$$

$$a(\sigma, \tau) - \int_\Omega \tau\ Dw\, dx = 0 \ \ \forall\ \tau\varepsilon\ (L^2(\Omega))^4_s,$$

$$\int_\Omega \sigma\ Dv = \int_\Omega f v\, dx.\ \forall \begin{cases} v\ \varepsilon\ H^2_\circ(\Omega). \\ v\ \varepsilon\ H^2(\Omega) \cap\ H^1_\circ(\Omega). \end{cases}$$

$(L^2(\Omega))^4_s$ denotes the set of symmetric 2×2 tensors which are in $L^2(\Omega)$.

However, as noticed in Section 7.10, to approximate this problem in the usual way does not represent any progress on the usual conforming approximations since one has to approximate $H^2_\circ(\Omega)$.

We have, by Green's formula,

$$\begin{aligned}\int_K \tau\, Dv &= \int_K \tau_{ij}\frac{\partial^2 v}{\partial x_i \partial x_j}\,dx\\ &= \int_{\partial K}\tau_{ij}\; n_j\frac{\partial v}{\partial x_i} - \int_K \frac{\partial \tau_{ij}}{\partial x_j}\frac{\partial v}{\partial x_i}\,dx\end{aligned}$$

Since

$$\frac{\partial v}{\partial x_i} = \frac{\partial v}{\partial n}n_i + \frac{\partial v}{\partial s}s_i$$

where (s_i) are the components of the unit tangent vector, we obtain

$$\int_\Omega \tau\, Dv = \sum_K \int_{\partial K} M_n(\tau)\frac{\partial v}{\partial n} + \sum_K\left(\int_{\partial K} M_{ns}(\tau)\frac{\partial v}{\partial s} - \int_K \frac{\partial \tau_{ij}}{\partial x_j}\frac{\partial v}{\partial x_i}\,dx\right)$$

Here

$$M_n(\tau) = \tau_{ij}\; n_i\; n_j,\;\; M_{ns}(\tau) = \tau_{ij}\; n_j\; s_i.$$

136 We define

$$\begin{aligned}b(\tau, v) &= \sum_K\left(\int_K \frac{\partial \tau_{ij}}{\partial x_j}\;\frac{\partial v}{\partial x_i}\,dx - \int_{\partial K} M_{ns}(\tau)\frac{\partial v}{\partial s}\,ds\right),\\ V &= \{\tau : \tau|_K\; \varepsilon\; (H^1(K))^4_s,\;\; K\; \varepsilon\; T_h, M_n(\tau)\end{aligned}$$

is continuous across inter element boundaries}.

Then

$$\int_\Omega \tau\; Dv = b(v, \tau)\;\; \forall\; v\; \varepsilon\; H^2_\circ(\Omega),\;\; \tau\; \varepsilon\; V.$$

In fact, $b(\cdot\,,\cdot)$ is continuous over $V\times M$ where $M = W^{1,p}_\circ(\Omega)(p > 2)$ so that (7.57) (clamped case) is equivalent to:

Find $\{\sigma, w\} \; \varepsilon \; V \times M$ such that

$$\begin{aligned} &a(\sigma, \tau) + b(\tau, w) = 0 \;\; \forall \; \tau \; \varepsilon \; V, \\ &b(\sigma, v) = -(f, v) \;\; \forall \; v \; \varepsilon \; M. \end{aligned} \tag{7.59}$$

We take

$$H = (L^2(\Omega))^4_s.$$

The Brezzi condition holds only on $H^1_\circ(\Omega)$, since if v is smooth and $\tau_{ij} = v\delta_{ij}$, then

$$\begin{aligned} M_{ns}(\tau) &= v\;(\delta_{ij}\; n_j\; s_i) = 0, \\ M_n(\tau) &= v, \end{aligned}$$

and 137

$$b(v, \tau) = \int_\Omega |\nabla v|^2 \, dx \geq \alpha \parallel v \parallel_1^2 \geq c \parallel v \parallel_1 \parallel \tau \parallel .$$

For the proof of existence of solutions of (7.59) and modified error estimates see BREZZI-RAVIART [7].

We choose

$$V_h = \{\tau : \tau \; \varepsilon \; (\mathbb{P}_\circ(K))^4_s, \;\; K \; \varepsilon \; T_h, \;\; M_n(\tau) \quad \text{is continuous}\}.$$

Since

$$v \; \varepsilon \; W^{1,p}_\circ(\Omega) \subset \; C^\circ(\overline{\Omega})$$

we find that after integration by parts on each of ∂K, $b(\tau, v)$ involves only the values of v at the vertices of T_h:

$$b(\tau, v) = \sum_K \int_{\partial K} M_{ns}(\tau) \frac{\partial v}{\partial s} = \sum_N R(\tau, N)\; v(N) \;\; \forall \; \tau \; \varepsilon \; V_h. \tag{7.59b}$$

Notice that only the value of v at the vertices has to be taken; therefore, we choose

$$M_h = \left\{ v_h \; \varepsilon \; C^\circ(\overline{\Omega}) : \;\; v_h|_K \; \varepsilon \; \mathbb{P}_1(K), K \; \varepsilon \; T_h, \;\; v_h = 0 \quad \text{on} \quad \partial\Omega \right\}$$

so that, if $\tau \varepsilon V_h$, then

$$b(\tau, v_h) = 0 \ \ \forall \ v_h \ \varepsilon \ M_h \Rightarrow b(\tau, v) = 0 \ \ \forall \ v \ \varepsilon \ M.$$

Therefore, $Z_h(0) \subset Z(0)$. Hence

$$\| \ \sigma - \sigma_h \ \|_0 \leq \operatorname*{Inf}_{\tau_h \varepsilon Z_h(-f)} \| \ \sigma - \tau_h \ \| \ .$$

Here $\{\sigma_h, w_h\} \varepsilon V_h \times M_h$ is the solution of the approximate problem.

$$\begin{cases} a(\sigma_h, \tau_h) + \Sigma \ R(\tau_h, N) \ w_h(N) = 0 \ \ \forall \ \tau_h \ \varepsilon \ V_h, \\ \Sigma \ R(\sigma_h, N) \ v_h(N) = -(f, v_h) \ \ \forall \ v_h \ \varepsilon \ M_h \end{cases} \tag{7.60}$$

138

Interpolation Operator.

The interpolation operator

$$\pi_h : \ (H^1(\Omega))^4 \to V_h$$

is defined by

$$\int\limits_{\gamma} M_n(\pi_h, \sigma) \, ds = \int\limits_{\gamma} M_n(\sigma) \, ds,$$

for each edge γ of the triangulation.

We have the estimate

THEOREM 12. *There exists a constant c independent of h such that*

$$\| \ \pi_h v \ \|_V \leq c \ \| \ v \ \|_V \tag{7.61}$$

and

$$\| \ \pi_h v - v \ \|_{0,\Omega} \leq \text{ch} \ \| \ v \ \|_{1.\Omega} \quad \forall \ v \ \varepsilon \ (H^1(\Omega))^4. \tag{7.62}$$

The proof of this is found in C.JOHNSON [24].

Properties of T_h.

We have

$$
\begin{aligned}
b(\pi_h\sigma, v_h) &= \sum_K \int\limits_{\partial K} M_n(\pi_h\sigma)\frac{\partial v_h}{\partial n} - \int\limits_K (\pi_h\sigma)\ D\ v_h \\
&= \sum_K \int\limits_{\partial K} M_n(\sigma)\frac{\partial v_h}{\partial n},\ \forall\ v_h\ \varepsilon\ M_h \\
&= b(\sigma, v_h).
\end{aligned}
$$

Hence 139

$$b(\pi_h\sigma - \sigma, v_h) = 0\ \ \forall\ v_h\ \varepsilon\ M_h \tag{7.63}$$

Therefore π_h maps $Z(\phi)$ into $Z_h(\phi)$.

Equations (7.61) and (7.63) imply that the discrete Brezzi condition is satisfied.

We have the error estimate

$$\| w - w_h \|_1 \leq c\left(\| \sigma - \sigma_h \|_0 + \operatorname*{Inf}_{v_h \varepsilon M_h} \| w - v_h \|_{1,p}\right)$$

(See BREZZI-RAVIART [7]).

The above method is called Hermann-Johnson method.

Morley Nonconforming Method.

Let

$$
\begin{aligned}
W_h = \{v_h : v_h|_K\ \varepsilon\ \mathbb{P}_2(K),\ K\ \varepsilon\ T_h,& \\
& v_h \quad \text{continuous at the vertices,} \\
& \frac{\partial v_h}{\partial n} \quad \text{continuous at the mid side point,} \\
& v_h = 0 \quad \text{at the boundary vertices,} \\
& \frac{\partial v_h}{\partial n} = 0 \quad \text{at the mid point of boundary edges}\}
\end{aligned}
$$

The space W_h makes use of the Morley finite element which has 6 degrees of freedom, namely, values at the three vertices and the values of the normal derivatives at the three mid side points.

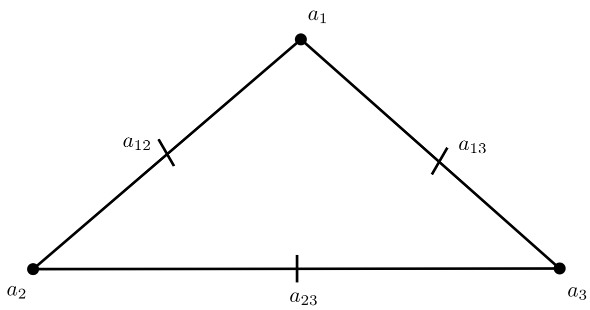

Figure 7.2:

140

We consider, for simplicity, the case $\lambda = 0$ and $\mu = 1/2$ so that

$$a(\sigma, \tau) = \int_{\Omega} \tau_{ij}\ \sigma_{ij}\, dx.$$

Let

$$L(v) = \sum_N f_N\ v(N);$$

that is L is a linear combination of Dirac masses (concentrated loads). Then

THEOREM 13. *The problem:*

Find $u_h \varepsilon W_h$ such that

$$\sum_K \int_K D\ u_h\ D\ v_h\, dx = L(v_h) \quad \forall\ v_h\ \varepsilon\ W_h, \tag{7.64}$$

is equivalent to (7.60) *when*

$$(f, v) = \sum_N f_N\ v(N)$$

in the sense that

$$u_h(N) = W_h(N) \quad \textit{at the vertices} \quad N,$$

and

$$\sigma_h|_K = Du_h|_K, \; K \; \varepsilon \; T_h.$$

Proof. Let u_h be a solution of (7.64). **141**

Define

$$\sigma_h^*|_K = Du_h|_K \quad \text{and} \quad w_h^* \varepsilon M_h \quad \text{by} \quad w_h^*(N) = u_h(N).$$

We will show that $\{\sigma_h^*(N), w_h^*\}$ is the solution of (7.60). Since u_h is a solution (7.64), we have

$$\sum_K \int_K \sigma_h^* \, Dv_h \, dx = L(v_h) \;\; \forall \; v_h \; \varepsilon \; W_h \tag{7.65}$$

Using Green's formula, we obtain

$$\sum_K \left(\int_{\partial K} M_n(\sigma_h^*) \frac{\partial v_h}{\partial n} + \int_{\partial K} M_{ns}(\sigma_h^*) \frac{\partial v_h}{\partial s} \right) = L(v_h) \;\; \forall \; v_h \; \varepsilon \; W_h \tag{7.66}$$

If b_i is one mid side point, then substituting v_h satisfying:

$$v_h = 0 \quad \text{at the vertices}$$

$$\frac{\partial v_h}{\partial n} = \begin{cases} 1 & \text{at} \quad b_i \\ 0 & \text{at the other nodes} \quad b_j, \;\; j \neq i \end{cases}$$

in the above equation we obtain that $M_n(\sigma_h^*)$ is continuous at b_i (by using 7.59b). This proves that ${}^*_h \varepsilon V_h$.

Since $\sigma_h^* \varepsilon V_h$, equation (7.66) gives

$$\sum_K \int_{\partial K} M_{ns}(\sigma_h^*) \frac{\partial v_h}{\partial s} = \sum_N f_N v_h(N) \;\; \forall \; v_h \; \varepsilon \; W_h$$

But
$$\sum_K \int_K M_{ns}(\sigma_h^*) \frac{\partial v_h}{\partial s} = -\sum_N R(\sigma_h^*,\ N)\ v_h(N).$$

Hence 142
$$\sum_K R(\sigma_h^*, N)\ v_h(N) = -\sum_N f_N\ v_h(N) \ \ \forall\ v_h\ \varepsilon\ W_h. \tag{7.67}$$

Let $v_h \varepsilon M_h$. Consider $\tilde{v}_h \varepsilon V_h$ defined by
$$\tilde{v}_h(N) = v_h(N), \frac{\partial}{\partial n}\ \tilde{v}_h(b_i) = 0.$$

Then (7.67) gives
$$\sum_K R(\sigma_h^*, N)\ \tilde{v}_h(N) = -\sum_N f_N\ \tilde{v}_h(N)$$

Therefore
$$\sum_N R(\sigma_h^*, N)\ v_h(N) = -\sum_N f_N\ v_h(N) \ \ \forall\ v_h\ \varepsilon\ M_h.$$

This is nothing but the second equation in (7.60) with σ_h replaced by σ_h^*. Now
$$\begin{aligned} a(\sigma_h^*, \tau) &= \sum_K \int_K (\sigma_h^*)_{ij}\ \tau_{ij} \\ &= \sum_K \int_K \frac{\partial^2 u_h}{\partial x_i\ \partial x_j}\ \tau_{ij} \\ &= \sum_K \int_K M_n(\tau) \frac{\partial u_h}{\partial n} + \int_K M_{ns}(\tau) \frac{\partial u_h}{\partial s} \ \ \forall\ \tau\ \varepsilon\ V_h \end{aligned}$$

by Green's formula.

The first term in the right side is zero since $u_h \varepsilon W_h$ and $\tau \varepsilon V_h$. The second term equals $-\sum\limits_N R(\tau, N) u_h(N)$. Hence we obtain
$$a(\sigma_h^*, \tau) + \sum_N R(\tau, N)\ w_h^*(N) = 0 \ \ \forall\ \tau\ \varepsilon\ V_h,$$

since $u_h(N) = w_h^*(N)$.

143 Thus $\{\sigma_h^*, w_h^*\}$ is a solution of (7.60). By uniqueness we have $\sigma_h = \sigma_h^*$ and $w_h^* = w_h$.

Thus we have proved that (7.64) $\Rightarrow$ (7.60). Let $\{\sigma_h, w_h\}$ be the solution of (7.60). We will show that u_h defined by

$$u_h(N) = w_h(N) \quad \text{for each vertex} \quad N \tag{7.68}$$

$$Du_h|_K = \sigma_h|_K \quad \text{for each} \quad K \,\varepsilon\, T_h \tag{7.69}$$

is the solution of (7.64).

It is easy to see that (7.68) and (7.69) define a unique u_h such that $u_h|_K \,\varepsilon\, \mathbb{P}_2(K)$ for each $K \varepsilon T_h$. We will prove that this $u_h \varepsilon W_h$.

From the first equation in (7.60), we obtain

$$\sum_K \int_{\partial K} M_n(\tau) \frac{\partial u_h}{\partial n} = 0 \;\; \forall \; \tau \,\varepsilon\, V_h.$$

This implies $\partial u_h / \partial n$ is continuous at mid side points and $\partial u_h / \partial n = 0$ at the boundary mid side points. Hence $u_h \varepsilon W_h$.

Let $v_h \varepsilon W_h$. Then there exists $v_h \varepsilon M_h$ such that $\tilde{v}_h(N) = v_h(N)$. Hence the second equation in (7.60) gives

$$\sum R(\sigma_h,\; N)\; v_h(N) = -\sum f_N \; v_h(N).$$

This shows

$$\sum_K \int_K Du_h \; Dv_h = \sum_N f_N \; v_h.$$

This proves (7.60) $\Rightarrow$ (7.64).

Thus the Hermann-Johnson method and the Morley nonconforming 144
method are equivalent, in this particular case where the load is a sum of concentrated loads. □

Exercise 7. Let K be a triangle. Let $\mathbb{P}_K = \mathbb{P}_2(K)$ and

$$\sum_K = \left\{ \delta_{a_i}, \frac{\partial}{\partial n} \delta_{a_{ij}}, 1 \le i < j \le 3 \right\},$$

where a_i's denote the vertices of K and a_{ij}'s denote the mid points of the sides of K. Show that $\sum_K$ is $\mathbb{P}_K$-unisolvent. The above finite element is called the Morley finite element.

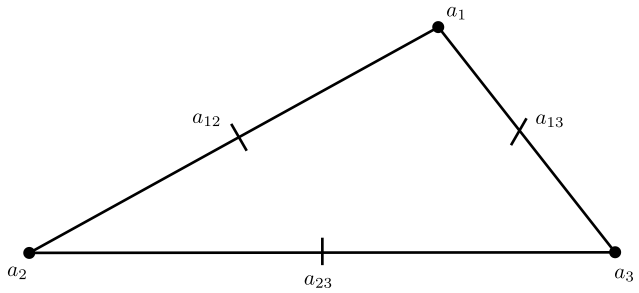

Figure 7.3:

REMARK 11. *We note that the Morley element has advantage over Herrmann-Johnson method, since in Morley's method we get a positive definite matrix and we have no constraints.*

Chapter 8

Spectral Approximation for Conforming Finite Element Method

8.1 The Eigen Value Problem

Let V and H be Hilbert spaces such that $V \hookrightarrow H$. We also assume that 145
this imbedding is compact. Let $\|\cdot\|_1$ denote the norm in V. The norm in H is denoted by $\|\cdot\|$ or $\|\cdot\|_0$ and the scalar product in H is $(\cdot\,,\cdot)$. We identify H with its dual H'.

Let $a(\cdot\,,\cdot) : V \times V \to \mathbb{R}$ be a continuous, symmetric bilinear form which is V-coercive with α as the coercive constant.

We shall consider the eigen value problem:

Find $u \varepsilon V, \mu \varepsilon \mathbb{R}$ such that

$$a(u,v) = \mu(u,v) \quad \forall\ v\ \varepsilon\ V \tag{8.1}$$

In the following, for an operator $\top : H \to H$, we write

$$\| \top \| = \operatorname*{Sup}_{f \varepsilon H, f \neq 0} \frac{\| \top f \|}{\| f \|}.$$

8.2 The Operator $\top$

The operator $\top : H \to V$ is defined as follows. If $f \varepsilon H$ then $\top f$ is defined to be the unique solution of the variational equation

$$a(\top f, v) = (f, v) \quad \forall\, v \,\varepsilon\, V.$$

By Lax-Milgram Lemma $\top f$ is well defined for all $f \varepsilon H$. As the
146 imbedding $V \hookrightarrow H$ is compact we obtain that $\top$, considered as an operator from H into H, is compact. The symmetry of $a(\cdot\,,\cdot)$ implies that $\top$ is symmetric. It is easy to see that (8.1) is equivalent to:

Find $u \varepsilon V$ and $\lambda \varepsilon \mathbb{R}$ such that

$$\top u = \lambda u \tag{8.2}$$

The μ and λ in (8.1) and (8.2) have the relation

$$\lambda\mu = 1.$$

From the Spectral Theorem for compact self-adjoint operators we have:

$\mathrm{Sp}(\top)$ is a countable set with no accumulation point other than zero. Every point in $\mathrm{Sp}(\top)$ other than zero is an eigenvalue of $\top$ with finite multiplicity.

8.3 Example

The model problem for (8.1) is

$$V = H^1_\circ(\Omega), \; H = L^2(\Omega)$$

where Ω is a smooth bounded open subset of $\mathbb{R}^n$,

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

The compactness of the imbedding $H^1(\Omega) \hookrightarrow L^2(\Omega)$ is well known. Problem (8.1) corresponds to:

Find $u\varepsilon H^1_\circ(\Omega), \mu\varepsilon\mathbb{R}$ such that

$$\begin{cases} -\Delta u = u \quad \text{in} \quad \Omega, \\ u = 0 \quad \text{on} \quad \partial\Omega \end{cases} \tag{8.3}$$

We note that $\top$ is the inverse of $-\Delta$. **147**

8.4 Approximate Problem

Let $V_h \hookrightarrow V$ be a finite element subspace of V. We consider the approximate eigen value problem:

Find $u_h \ \varepsilon \ V_h, \ \mu_h \ \varepsilon \ \mathbb{R}$ such that

$$a(u_h, v_h) = \mu_h(u_h, v_h) \ \ \forall \ v_h \ \varepsilon \ V_h \tag{8.4}$$

Here again, we introduce an operator $\top_h : H \to H$ where $\top_h f$ is the unique solution of

$$a(\top_h f, v_h) = (f, v_h) \ \ \forall \ v_h \ \varepsilon \ V_h. \tag{8.5}$$

As in the continuous case, we have $\| \top_h \| \leq c/\alpha$ which shows that $\top_h$ is uniformly bounded. Again (8.4) is equivalent to:

Find $u_h\varepsilon V_h$ and $\lambda_h = 1/\mu_h$ such that

$$\top_h \ u_h = \lambda_h \ u_h \tag{8.6}$$

It is obvious that $\top_h$ is a self-adjoint, compact operator.

We assume that

$$\| \top - \top_h \| \leq \varepsilon(h) \tag{8.7}$$

and

$$\| (\top - \top_h) \ f \| \leq e(h) \tag{8.8}$$

for all smooth f and $\top f$. Further, we assume **148**

$$0 \leq e(h) \leq \ \varepsilon(h) \quad \text{and} \quad \varepsilon(h) \to 0 \tag{8.9}$$

EXAMPLE. Let

$$V_h = \{v_h \,\varepsilon\, H^1_\circ(\Omega) : v_h|_K \,\varepsilon\, \mathbb{P}_k(K),\ K \,\varepsilon\, T_h\}$$

where T_h is a regular family of triangulations of Ω. We have (cf. Chapter 5)

$$\| \top f - \top_h f \|_{0,\Omega} \leq \text{ch}^{s+1} \| f \|_{s-1,\Omega}, 1 \leq s \leq k, \tag{8.10}$$

provided that Ω is a convex polygon and that

$$\| \top f \|_{s+1} \leq c \| f \|_{s-1,\Omega} \ . \tag{8.11}$$

From GRISVARD [22] this is atleast true for $s = 1$, which shows that $\varepsilon(h) = \text{ch}^2$ and $e(h) = 0(h^{k+1})$.

8.5 Convergence and Error Estimate for the Eigen Space.

Assumption (8.7) (8.9) show that $\top_h \to \top$ in norm.

From KATO [26] (Chapter 5. Section 4.3) we know that the spectrum of $\top_h$ converges to the spectrum of $\top$ in the following sense: For all non-zero $\lambda \varepsilon \operatorname{Sp}(\top)$ with multiplicity m and for each h such that $\varepsilon(h) < d/2$, where

$$d = \min_{\lambda' \varepsilon \operatorname{Sp}(\top)} |\lambda - \lambda'|,$$

there exist exactly m eigen values $\lambda_{ih} \varepsilon \operatorname{Sp}(\top_h)$ (counted according to multiplicity) such that

$$|\lambda - \lambda_{ih}| \leq \ \varepsilon(h).$$

149

Let $\Gamma = \{z \varepsilon C : |z - \lambda| = d/2\}$. We know that

$$P = -\frac{1}{2\pi i} \int_\Gamma R_z(\top)\, dz, \tag{8.12}$$

$$P_h = -\frac{1}{2\pi i} \int_\Gamma R_z(\top_h)\, dz,\ \varepsilon(h) < d/2, \tag{8.13}$$

where $R_z(\top) = (\top - z)^{-1}$, are the spectral projections on to the eigenspaces E and E_h associated with λ and λ_{ih}'s. The dimension of each of the spaces E and E_h is m (See KATO [26], Chapter 4, Section 4.3).

LEMMA 1. *For $u \varepsilon E$, we have*

$$\| u - P_h u \| \leq c_2 \| \ (\top - \top_h)\, u \| \,. \tag{8.14}$$

Proof. We consider

$$\begin{aligned} R_z(\top) - R_z(\top_h) &= R_z(\top_h)(\top_h - z)R_z(\top) - R_z(\top_h)(\top - z)R_z(\top) \\ &= R_z(\top_h)(\top_h - \top)R_z(\top). \end{aligned}$$

Hence

$$P - P_h = -\frac{1}{2\pi i}\int\limits_{\Gamma} R_z(\top_h)(\top_h - \top)R_z(\top)\, dz.$$

Let $u \varepsilon E$. Then we have

$$Pu = u, \top u = \lambda u \quad \text{and} \quad R_z(\top)u\frac{1}{\lambda - z}u.$$

Therefore 150

$$u - P_h u = -\frac{1}{2\pi i}\int\limits_{\Gamma} \frac{R_z(\top_h)}{\lambda - z}\, dz\ (\top - \top_h)\ u.$$

We show that the integral on the right is bounded. Indeed, for $z \varepsilon \Gamma$ and $\varepsilon(h) < d/2$ we have

$$\begin{aligned} \top_h - z &= \top_h - \top + \top - z \\ &= ((\top_h - \top)\ R_z(\top) + I)\ (\top - z), \end{aligned}$$

which implies

$$R_z(\top_h) = R_z(\top)\ (I + A_h)^{-1},$$

where

$$A_h = (\top_h - \top)\ R_z(\top).$$

If $P(\top)$ denotes the resolvent set of $\top$ then, as $R_z(\top)$ is continuous in $z \varepsilon P(\top)$ and Γ is a compact subset of $P(\top)$, we obtain

$$\| R_z(\top) \| \leq c_1 \quad \text{for all } z \ \varepsilon \Gamma,$$

and $\| A_h \| \leq c_1 \varepsilon(h)$, where c_1 is a constant. This implies

$$\| (I + A_h)^{-1} \| \leq 2 \quad \text{for} \quad \varepsilon(h) \leq \frac{1}{2c_1}.$$

Thus

$$\| u - P_h u \| \leq c_2 \| \ (\top - \top_h)\ u \| .$$

□

LEMMA 2. *If the eigen vectors in E are smooth enough, we have*

$$\delta(E,\ E_h) \leq c\ e(h), \tag{8.15}$$

151 *where*

$$\delta(E,\ E_h) = \text{Sup}\ \{d(u, E_h) :\ u \ \varepsilon\ E, \| u \| = 1\}.$$

Remark 1. In this chapter we follow closely OSBORN [34]. We will use the result: "In a Hilbert space $H, \delta(E, E_h) = \delta(E_h, E)$". Osborn, however, considers the more general case of a non-self-adjoint operator in a Banach space, which involves more complicated arguments.

8.6 Error Estimates for the Eigen Values.

Let $Q_h = P_h|_E$, the restriction of P_h to E. Then Q_h maps E into E_h. We prove that Q_h is invertible for small h.

Indeed, for h small enough we have $\dim E = \dim E_h$.

Let $f \varepsilon E$ be such that $Q_h f = 0$. Then

$$\| f \| = \| f - Q_h f \| = \| f - P_h f \| \leq c_2 \varepsilon(h) \| f \|,$$

where we have used Lemma 1. Therefore, for $c_2 \varepsilon(h) < 1$ we have $\| f \| = 0$. Hence Q_h is invertible for $\varepsilon(h) < \min(d/2, 1/c_2)$.

Let us evaluate $\| Q_h^{-1} \|$. If $f \varepsilon E$ with $\| f \| = 1$, then

$$1 - \| Q_h f \| \leq \| f - P_h f \| \leq c_2\ \varepsilon(h).$$

Therefore

$$\| Q_h f \| \geq 1/2, \quad \text{if} \quad \varepsilon(h) \leq 1/2c_2$$

and 152

$$\| Q_h^{-1} \| \leq 2, \quad \text{for} \quad \varepsilon(h) \leq 1/2c_2.$$

Let $\hat{\top}_h : E \to E$ be defined by

$$\hat{\top}_h = Q_h^{-1} \top_h\ Q_h.$$

The eigenvalues of $\hat{\top}_h$ are again $\lambda_{ih}, i = 1, 2, \ldots, m$ (but the eigenvectors of $\hat{\top}_h$ are different from those of $\top_h$).

Let $W_{jh} \varepsilon E, \| W_{jh} \| = 1$ be an eigen vector of $\hat{\top}_h$ associated with the eigen value λ_{jh}. Therefore,

$$\begin{aligned}
\lambda - \lambda_{jh} &= ((\lambda - \lambda_{jh})\ w_{jh},\ w_{jh}) \\
&= ((\top - \top_h)\ w_{jh},\ w_{jh}) \\
&\leq \sup_{\phi \varepsilon E, \|\phi\| = 1} \{((\top - \top_h)\ \phi, \phi)\}.
\end{aligned}$$

Now

$$\begin{aligned}
\top - \hat{\top}_h &= \top - Q_h^{-1} \top_h\ Q_h = \top - Q_h^{-1}\ \top_h\ P_h \\
&= \top - Q_h^{-1}\ P_h\ \top_h.
\end{aligned}$$

Hence

$$\top - \hat{\top}_h = Q_h^{-1}\ P_h\ (\top - \top_h), \tag{8.16}$$

since P_h commutes with $\top_h$ and $Q_h^{-1} P_h u = u$ for $u \varepsilon E$. Hence

$$\| (\top - \hat{\top}_h) \phi \| \leq 2 \| (\top - \top_h) \phi \| \quad \text{for all } \phi \varepsilon E$$

since

$$\| Q_h^{-1} \| \leq 2, \| P_h \| \leq 1 \quad \text{for} \quad \varepsilon(h) \leq 1/2c_1.$$

Therefore 153

$$|\lambda - \lambda_{jh}| \leq 2e(h) \quad \text{for} \quad \varepsilon(h) \leq \min(1/2c_1,\ d/2) \tag{8.17}$$

Application. In the example given in Section 8.3, where $\top$ is the inverse of the negative Laplace operator, we get

$$|\lambda - \lambda_{ih}| \leq \mathrm{ch}^{k+1},\ 1 \leq i \leq m, \tag{8.18}$$

$$d(E, E_h) \leq \mathrm{ch}^{k+1}, \tag{8.19}$$

since $e(h) \leq \mathrm{ch}^{k+1}$ provided that the eigen functions in E are in H^{k+1} (This may happen even though Ω is a polygon: If $\Omega =]0,1[^2$ the eigen functions are known explicitly and they are C^∞. In fact, the eigen functions are

$$u_{nm} = \mathrm{Sin}\ n\pi x.\,\mathrm{Sin}\ m\pi x.$$

However, the error estimate for $|\lambda - \lambda_{jh}|$ can be improved as will be shown in the following section. One indeed has

$$|\lambda - \lambda_{jh}| \leq \mathrm{ch}^{2k}.$$

8.7 Improvement of the Error Estimate for the Eigen Values.

We denote by S_h the projection on E_h along $E^\perp$. We notice that

$$R_h = Q_h^{-1} P_h$$

is the projection on E along $E_h^\perp$.

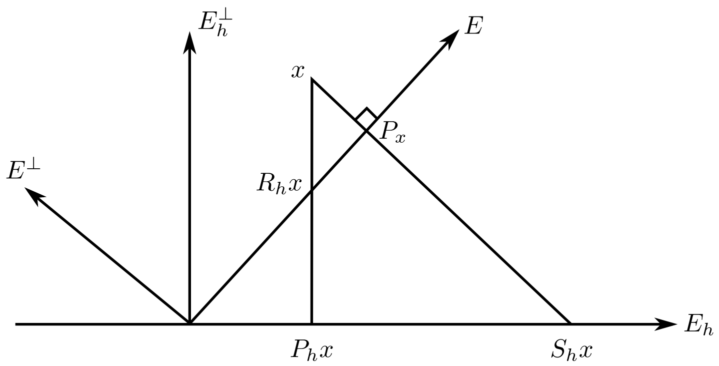

Figure 8.1:

154

S_h and R_h are related by

LEMMA 3. $S_h = R_h^*$.

Proof. For $u, v\varepsilon H$, we have

$$
\begin{aligned}
(u, S_h\phi) &= (P_hu, S_h\phi), \quad \text{since} \quad P_hS_h = S_h \quad \text{and} \quad P_h^* = P_h,\\
&= (P_hR_hu, S_h\phi), \text{ since } \quad R_h \text{ is the projection on } E \text{ along } E_h^\perp;\\
&= (R_hu, S_h\phi)\\
&= (R_hu, PS_h\phi), \quad \text{since} \quad PR_h = R_h \quad \text{and} \quad P = P^*,\\
&= (R_hu, P\phi), \quad \text{since} \quad PS_h = P,\\
&= (R_hu, \phi).
\end{aligned}
$$

□

We will now prove a Lemma which will give an upper bound for $\| P - S_h \|$.

LEMMA 4. *We have*

$$\| P - S_h \| \leq \frac{\delta(E_h, E)}{1 - \delta(E_h, E)} \tag{8.20}$$

155

Proof. Since S_h is the projection on E_h along $E^{\perp}$ we have $P = PS_h$. If $x \varepsilon E_h$ and $\| x \| = 1$ then

$$\| x - Px \| = d(x, E) \leq \sup_{x\varepsilon E_h, \|x\|=1} d(x, E) = \delta(E_h, E).$$

Therefore,

$$\| y - Py \| \leq \delta(E_h, E) \| y \| \quad \text{for all } y \, \varepsilon E_h \tag{8.21}$$

Now, for any $x \varepsilon H$,

$$\begin{aligned} \| S_h x \| &\leq \| S_h x - PS_h x \| + \| PS_h x \| \\ &\leq \delta(E_h, E) \| S_h x \| + \| Px \| . \end{aligned}$$

Thus

$$\| S_h x \| \leq \frac{1}{1 - \delta(E_h, E)} \| x \| \tag{8.22}$$

Hence we obtain

$$\begin{aligned} \| (P - S_h)x \| &= \| PS_h x - S_h x \| \\ &\leq \delta(E_h, E) \| S_h x \|, \quad \text{by (8.21)}, \\ &\leq \frac{\delta(E_h, E)}{1 - \delta(E_h, E)} \| x \|, \quad \text{using (8.22)} \end{aligned}$$

Therefore

$$\| P - S_h \| \leq \frac{\delta(E_h, E)}{1 - \delta(E_h, E)}.$$

□

Finally we have from (8.16),

$$\begin{aligned} ((\top - \hat{\top}_h)\, \phi, \phi) &= (R_h\, (\top - \top_h)\phi, \phi) \\ &= ((\top - \top_h)\, \phi,\ S_h \phi) \\ &= ((\top - \top_h)\phi, \phi) + ((\top - \top_h)\phi, S_h\phi - \phi), \end{aligned}$$

156 where $\phi \varepsilon E$ and $\| \phi \| = 1$.

For $\phi \varepsilon E$, using Lemma 4 and Lemma 2, we obtain

$$\| \phi - S_h \| \leq \frac{\delta(E_h, E)}{1 - \delta(E_h, E)} \leq k \; e(h),$$

for sufficiently small h, where k is a constant.

We know that for all $\phi \varepsilon E$ with $\| \phi \| = 1$

$$\lambda - \lambda_{ih} = ((\top - \hat{\top}_h) \; \phi, \phi).$$

Hence

$$|\lambda - \lambda_{ih}| \leq \sup_{\phi \varepsilon E, \|\phi\|=1} ((\top - \top_h)\phi, \phi) + k(e(h))^2.$$

Thus we have proved

THEOREM 5. *When E is a smooth subset of H and h is sufficiently small we have*

$$|\lambda - \lambda_{ih}| \leq \sup_{\phi \varepsilon E, \|\phi\|=1} ((\top - \top_h)\phi, \phi) + k(e(h))^2, \tag{8.23}$$

where k is a constant.

Application. In the case of the example in Section 8.3, we give an estimate of 157

$$\alpha_h = \sup_{\phi \varepsilon E, \|\phi\|=1} ((\top - \top_h)\phi, \phi).$$

Let w be the solution of

$$a(v, w) = (\phi, v) \quad \forall \; v \; \varepsilon \; V, \tag{8.24}$$

where $\phi \varepsilon E$ and $\| \phi \| = 1$.

We have

$$\begin{aligned} ((\top - \top_h)\phi, \phi) &= a((\top - \top_h)\phi, \; w) \\ &= a((\top - \top_h)\phi, w - v_h) \quad \text{for all } v_h \; \varepsilon \; V_h, \end{aligned}$$

since

$$a((\top - \top_h)\phi, \; v_h) = 0 \quad \text{for all } v_h \; \varepsilon \; V_h.$$

If $w \varepsilon H^{k+1}(\Omega)$, we get

$$\| w - v_h \|_1 \leq \mathrm{ch}^k \| w \|_{k+1} \leq \mathrm{ch}^k \| \phi \|_{k-1},$$

using regularity theorem and the error estimates in Chapter 5.

Since E is finite-dimensional there exists a constant c such that

$$\| \phi \|_{k-1} \leq c \| \phi \| = c.$$

Finally, we have

$$\begin{aligned} \alpha_h &\leq \mathrm{ch}^k \ \| (\top - \top_h)\phi \|_1 \\ &\leq \mathrm{ch}^{2k}, \end{aligned}$$

158 provided that $E \subset H^{k+1}$.

Thus we have proved

THEOREM 6. *For the model problem (See Section 8.3) we have*

$$|\lambda - \lambda_{ih}| \leq \mathrm{ch}^{2k},$$

provided $E \subset H^{k+1}(\Omega)$, the solution of (8.24) *is in $H^{k+1}(\Omega)$ and h is small.*

REMARK 2. *Error estimates for the semi-discrete approximation to parabolic equation of the form*

$$\left(\frac{du}{dt}, v\right) + a(u, v) = (f, v) \quad \textit{for all } v \ \varepsilon \ V,$$
$$u(0) = v_\circ, \ v_\circ \ \varepsilon \ V,$$

can be obtained using spectral approximation. The reader is referred to THOMEE [44], [45].

Chapter 9

Nonlinear Problems

Introduction.

We consider here problems of the following type: **159**

Find $u \varepsilon C$ such that

$$J(u) \leq J(v) \quad \text{for all } v \;\varepsilon\; V, \tag{9.1}$$

where C is a closed, convex subset of a Banach space V and $J : V \to \mathbb{R}$ is a convex, lower semi-continuous ($l.s.c.$) function.

We denote by $(\cdot\,,\cdot\,)$ the duality pairing $V' - V$ and $\|\cdot\|$ the norm in V. We write (9.1) in the form:

Find $u \varepsilon C$ such that

$$J(u) = \operatorname*{Inf}_{v \varepsilon C} J(v). \tag{9.2}$$

The existence and uniqueness of the solutions of (9.2) is given by

THEOREM 1. *Assume that J is coercive on C, that is*

$$J(v) \to \infty \quad if \quad \|v\| \to \infty, v \;\varepsilon\; C. \tag{9.3}$$

Then problem (9.2) *has atleast one solution provided that V is reflexive.*

If J is strictly convex, then (9.2) *has almost one solution.*

The proof of theorem 1 can be found in EMELAND-TEMAN [15].

Exercise 1. If J is Gateaux-differentiable, then u is a solution of (9.2) iff 160

$$(J'(u), v - u) \geq 0 \quad \text{for all } v \,\varepsilon\, C.$$

If C is affine linear, (9.1) implies

$$(J'(u), v - u) = 0 \;\forall\; v \,\varepsilon\, C.$$

Approximation.

Let V_h be a finite-dimensional subspace of V and C_h a closed, convex subset of V_h. Then the approximate problem corresponding to (9.2) is:

Find $u_h \,\varepsilon\, C_h$ such that

$$J(u_h) = \operatorname*{Inf}_{v_h \varepsilon C_h} J(v_h). \tag{9.4}$$

We assume that J is strictly convex and coercive. Moreover, we assume that C_h approximates C.

$$\left.\begin{array}{l} \text{For all } v \,\varepsilon\, C \quad \text{there exists} \quad v_h \,\varepsilon\, C_h \\ \text{such that} \quad v_h \to v \quad \text{(strongly) as} \quad h \to 0; \end{array}\right\} \tag{9.5}$$

$$\left.\begin{array}{l} \text{If} \quad w_h \rightharpoonup w \,\varepsilon\, V \quad \text{as} \quad h \to 0 \quad \text{and} \quad w_h \,\varepsilon\, C_h \\ \text{then} \quad w \,\varepsilon\, C. \end{array}\right\} \tag{9.6}$$

Then one can easily prove that the solution u_h of (9.4) converges weakly to u, the solution of (9.2).

Note that (9.5) implies that C_h has to be sufficiently big and (9.6) demands C_h to be sufficiently small.

161 To get strong convergence of u_h to u, one needs some strong monotonicity; there exists $\alpha, \gamma > 0$ such that

$$(J'(u) - J'(v), u - v) \geq \alpha \parallel u - v \parallel^{\gamma}, \forall\, u,\; v \,\varepsilon\, C \tag{9.7}$$

Case 1. *We obtain an error estimate when $C_h = C \cap V_h$. From Exercise 1, we have*

$$(J'(u), v - u) \geq 0 \ \forall\ v \ \varepsilon\ C, \tag{9.8}$$

$$(J'(u_h),\ v_h - u_h) \geq 0 \ \ \forall\ v_h \ \varepsilon\ C_h. \tag{9.9}$$

As $C_h = C \cap V_h$, choosing $v = u_h$ in (9.8) *and adding to* (9.9)*, we get*

$$(J'(u) - J'(u_h), u_h - u) + (J'(u_h),\ v_h - u) \geq 0.$$

Therefore (assuming J to be continuously differentiable)

$$\alpha \parallel u_h - u \parallel^{\gamma} \leq (J'(u_h), v_h - u) \leq c \parallel v_h - u \parallel,$$

since u_h is bounded. Finally,

$$\parallel u_h - u \parallel \leq c \operatorname*{Inf}_{v_h \varepsilon C_h} \parallel v_h - u \parallel^{1/\gamma}, \tag{9.10}$$

provided J' is weakly continuous.

Note that

$$\operatorname*{Inf}_{v_h \varepsilon C_h} \parallel v_h - u \parallel$$

measures how good is the approximation C_h to C. Note also that, as u_h is bounded, it is enough if (9.7) *holds on bounded subsets of C.*

Exercise 2. Let $\phi : V \to \mathbb{R}$ be Lipschitz but not differentiable. Then **162**

$$J(u) = \operatorname*{Inf}_{v \varepsilon C}[J(v) + \phi(v)]$$

is equivalent to

$$(J'(u), v - u) + \phi(v) - \phi(u) \geq 0 \ \ \forall\ v \ \varepsilon\ C.$$

Derive an error estimate similar to (9.10).

Case 2. *When C is of the form*

$$C = \{v \ \varepsilon\ V : b(v, \mu) = (\phi, \mu) \ \ \forall\ \mu \ \varepsilon\ M\}, \tag{9.11}$$

where M is a Hilbert space, $b(\cdot\,,\cdot)$ is a continuous, bilinear form on $V \times M$ and $\phi\varepsilon M$ satisfying Brezzi's condition (See Chapter 7), Problem (9.1) *is equivalent to:*

Find $\{u, \lambda\}\ \varepsilon\ V \times M$ such that

$$< J'(u), v > +b(v, \lambda) = 0 \ \ \forall\ v\ \varepsilon\ V, \tag{9.12}$$

$$b(u, \mu) = (\phi, \mu) \ \ \forall\ \mu\ \varepsilon\ M. \tag{9.13}$$

We notice that (9.11) *is affine linear and from Exercise 1 we obtain that* (9.1) *is equivalent to*

$$(J'(u), v - u) = 0 \ \ \forall\ v\ \varepsilon\ C. \tag{9.14}$$

Let $B : V \to M$ be defined by

$$(Bv, \mu) = b(v, \mu) \ \ \forall\ v\ \varepsilon\ V,\ \mu\ \varepsilon\ M.$$

163 *Clearly $C = v + \operatorname{Ker} B$, where $v\varepsilon C$. This together with* (9.14) *implies*

$$J'(u)\ \varepsilon\ (\operatorname{Ker} B)^{\perp} = \operatorname{Im} B^*,$$

which is closed from Brezzi's condition. Hence there exists $\lambda\varepsilon M$ such that

$$J'(u) = -B^*\lambda.$$

Thus

$$(J'(u),\ v) + b(v, \lambda) = 0 \ \ \forall\ v\ \varepsilon\ V.$$

$u\varepsilon C$ implies $b(u, \mu) = (\phi, \mu) \forall \mu\varepsilon M$.

So we have proved that (9.1) *implies* (9.12) *and* (9.13).

If (9.12) *and* (9.13) *hold, then $u\varepsilon C$ and*

$$(J'(u), u) = -b(u, \lambda) = -b(v, \lambda) \ \ \forall\ v\ \varepsilon\ C$$

Hence $\langle J'(u), v - u\rangle = 0 \ \ \forall v\varepsilon C$, which is equivalent to (9.1).

Thus we proved the equivalence of (9.1) *and* (9.12) (9.13).

A natural approximation C_h to C will be

$$C_h = \{v\ \varepsilon\ V : b(v, \mu_h) = (\phi, \mu_h) \ \ \forall\ \mu_h\ \varepsilon\ M_h\},$$

where M_h approximates M. In this case

$$C_h \not\subset C.$$

164 **EXAMPLE 1.** Nonlinear Dirichlet Problem.

$$V = W^{1,p}(\Omega), C = V$$
$$J(v) = \frac{1}{p}\int_{\Omega} |\nabla v|^p \, dx - \int_{\Omega} fv\, dx,$$

where

$$f \varepsilon L^q, \ 1/p + 1/q = 1$$

For $1 < p < \infty, W^{1.p}(\Omega)$ is reflexive and $J(v) \to \infty$ as $\| v \| \to \infty$. One has

$$(J'(u), v) = \int_{\Omega} |\nabla u|^{p-2} \nabla u . \nabla v \ \ dx - \int_{\Omega} fv\, dx,$$

and some strong monotonicity results of the type (9.7) are proved in GLOWINSKI-MARROCCO [19].

EXAMPLE 2. The Obstacle Problem.

$$V = H^1_\circ(\Omega),$$
$$C = \{v \ \varepsilon \ H^1_\circ(\Omega) : v \geq 0 \ \text{ a.e.} \quad \text{on} \quad \Omega\},$$
$$J(v) = \frac{1}{2}\int_{\Omega} |\nabla v|^2 \, dx - \int_{\Omega} fv\, dx.$$

Existence and uniqueness of the solution of the minimization problem are straightforward.

Let V_h be the standard Lagrange finite element space of degree 1 and $C_h = C \cap V_h$. One has (9.10) with $\gamma = 2$; therefore it seems that one gets

$$\| u - u_h \| = 0(\sqrt{h})$$

since the interpolate $\pi_h u \varepsilon C_h$ as long as $u \varepsilon C$. However, one has 165

$$(J'(u_h), v_h - u) = (J'(u), v_h - u) + (J'(u_h) - J'(u), v_h - u)$$
$$\leq (-\Delta u - f, v_h - u) + \frac{M\varepsilon}{2} \| u_h - u \|^2_1 + \frac{M}{2\varepsilon} \| v_h - u \|^2_1 .$$

Hence

$$\| u - u_h \|_1^2 \leq c\left(\| v_h - u \|_0^2 + \| v_h - u \|_1^2\right)$$

Therefore

$$\| u - u_h \|_1 = 0(h).$$

EXAMPLE 3. Elasto - Plastic Torsion. J and V as is Example 2,

$$C = \{v \; \varepsilon \; H_\circ^1(\Omega) : |\nabla v| \leq 1 \text{ a. e.} \quad \text{on} \quad \Omega\},$$

V_h same as in Example 2 and $C_h = C \cap V_h$.

In this case the interpolate $\pi_h u$ is not in C_h whereas u is in C. One gets $0(h^{1/2-\varepsilon})$.

EXAMPLE 4. The Flow of a Bingham Fluid in a Cylindrical Pipe. This is a particular case of Exercise 2 with J, V as above and

$$\phi(v) = \int_\Omega |\nabla v| \, dx.$$

9.2 Generalization

Note that $J' : V \to V'$ satisfies

$$(J'(u) - J'(v), u - v) \geq 0 \;\; \forall \; u, v \; \varepsilon \; V.$$

166

An operator $A : V \to V'$ is said to be *monotone* if

$$(Au - Av, u - v) \geq 0 \;\; \forall \; u, \;\; v \; \varepsilon \; V. \tag{9.15}$$

A is *bounded* if A maps bounded sets of V into bounded sets of V'.

A is *hemi-continuous* if

$$\lim_{\lambda \to 0} (A(u + \lambda w), v) = (A(u), v) \;\; \forall \; u, v, w \; \varepsilon \; V. \tag{9.16}$$

A is *coercive* if

$$\frac{(A(v), v)}{\| v \|} \to \infty \text{ if } \| v \| \to \infty \text{ for } v \; \varepsilon \; C. \tag{9.17}$$

We have

THEOREM 2. *If A is a monotone, bounded hemi continuous and coercive operator then the problem:*

Find $u \varepsilon C$ such that

$$(A(u), v - u) \geq 0 \quad \forall \, v \, \varepsilon \, C \tag{9.18}$$

has atleast one solution.

For a proof of this Theorem see LIONS [28]. The problem (9.18) has at most one solution if A is *strongly monotone*, i.e. there exists $\alpha, \gamma > 0$ such that

$$\alpha \parallel u - v \parallel^{\gamma} \leq (A(u) - A(v), u - v) \quad \forall \, u, v \, \varepsilon \, C \tag{9.19}$$

The error analysis can be carried out in the same way.

9.3 Contractive Operators.

Let $T : C \to C$ be a mapping, where C is a closed, convex subset of a Hilbert space H. The scalar product in H is denoted by $(\cdot\, , \cdot)$. 167

We call T *contractive* iff

$$\parallel Tx - Ty \parallel \leq \parallel x - y \parallel, \quad \forall \; x, \; y \, \varepsilon \, C. \tag{9.20}$$

T is *strictly contractive* iff there exists a θ with $0 < \theta < 1$ such that

$$\parallel Tx - Ty \parallel \leq \theta \parallel x - y \parallel \quad \forall \; x, \; y \, \varepsilon \, C. \tag{9.21}$$

We say that T is *firmly contractive* iff (cf. BROWDER-PETRYSHN [8])

$$\parallel Tx - Ty \parallel^2 \leq (Tx - Ty, x - y) \quad \forall \; x, \; y \, \varepsilon \, C \tag{9.22}$$

T is *quasi firmly contractive* iff there exists a $\theta, 0 < \theta < 1$ such that

$$\parallel Tx - Ty \parallel^2 \leq \theta(Tx - Ty, x - y) + (1 - \theta) \parallel x - y \parallel^2 \tag{9.23}$$

Note that (9.22)$\Rightarrow$ (9.23) $\Rightarrow$ (9.20) and (9.21)$\Rightarrow$ (9.20).

Geometrical Interpretation of the Above Definitions Let $y \varepsilon C$ be a fixed point of T, i.e. $Ty = y$. If T is contractive and $x \varepsilon C$ then Tx lies in the closed ball with y as centre and $\| y - x \|$ as radius. If T is strictly contractive, then Tx lies in the open ball with y as centre and $\| y - x \|$ as radius, for all $x \varepsilon C$. If T is firmly contractive then from (9.22) we obtain

$$(Tx - y, x - Tx) \geq 0 \ \ \forall \ x \ \varepsilon \ C.$$

168 This means that the angle between $y - Tx$ and $x - Tx$ is obtuse.

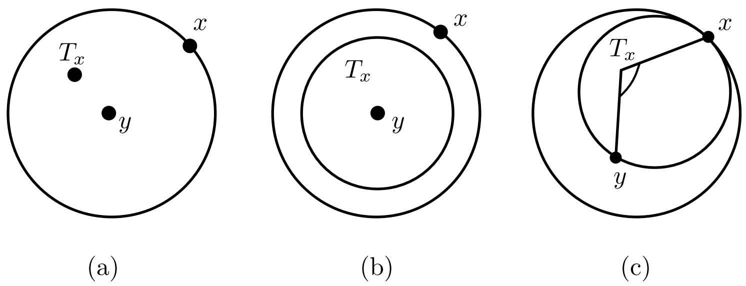

Figure 9.1:

Note that if T_1 and T_2 are contractive then $T = T_1 T_2$ is contractive and T is strictly contractive if any one of T_1 and T_2 is. However, T_1, T_2 firmly (or quasi-firmly) contractive implies *only* $T = T_1 T_2$ is *contractive*.

Fixed Points. We recall that x is a fixed point of T iff $Tx = x$. Let $F(T)$ denote the set of all fixed points of T. If T is strictly contractive then T has a unique fixed point and $F(T)$ is singleton. We have

THEOREM 3. *If T is contractive, then $F(T)$ is closed and convex.*

Proof. Let $x_n \varepsilon F(T), x_n \to x$. Then

$$\| x_n - Tx \| \leq \| x_n - x \|$$

169 Taking the limit as $n \to \infty$ we get $\| x - Tx \| = 0$, *i.e.* $x \varepsilon F(T)$. Hence $F(T)$ is closed.

Let $x, y \varepsilon F(T)$ and $u = \theta x + (1 - \theta)$ where $0 < \theta < 1$. We have

$$\| x - Tu \| \leq \| x - u \| = (1 - \theta) \| x - y \|, \tag{9.24}$$

$$\| y - Tu \| \leq \| y - u \| = \theta \| x - y \|, \tag{9.25}$$
$$\| x - y \| \leq \| x - Tu \| + \| y - Tu \| \leq \| x - y \|, \tag{9.26}$$

Since H is strictly convex, we obtain using (9.24) and (9.25),

$$x - Tu = c(y - Tu) \tag{9.27}$$
$$\| x - Tu \|^2 = c(y - Tu, x - Tu), \tag{9.28}$$

$$\| x - y \|^2 = \| x - Tu \|^2 + \| y - Tu \|^2 + 2 \| x - Tu \| \| y - Tu \|,$$

by (9.26),

$$\begin{aligned} \| x - y \|^2 &= \| x - Tu + Tu - y \|^2 \\ &= \| x - Tu \|^2 + \| Tu - y \|^2 + 2(x - Tu, Tu - y) \end{aligned}$$

So

$$(x - Tu, Tu - y) = \| x - Tu \| \| y - Tu \| > 0. \tag{9.29}$$

From (9.28) and (9.29) we obtain $c < 0$.

Equations (9.26) and (9.27) imply

$$\| y - Tu \| = \frac{1}{1 + |c|} \| x - y \| .$$

This with (9.25) gives $|c| \geq (1 - \theta)\theta^{-1}$. Similarly using (9.24), (9.26) and (9.27) we obtain $|c| \leq (1 - \theta)\theta^{-1}$. Thus

$$|c| = (1 - \theta)\, \theta^{-1}.$$

Hence **170**

$$\theta(x - Tu) = -(1 - \theta)(y - Tu).$$

Therefore

$$Tu = \theta x + (1 - \theta)y = u,$$

that is

$$u \; \varepsilon \; F(T).$$

□

REMARK 1. *Theorem 3 can be proved geometrically.*

Let $u = \theta x + (1-\theta)y$, $x, y \varepsilon F$. Since $x \varepsilon F(T)$ and T is contractive Tu lies in the closed ball C_x with x as centre and $\| x - u \|$ as radius. Similarly Tu lies in the closed ball C_y with y as centre and $\| y - u \|$ as radius. But $C_x \cap C_y = u$. Hence $Tu = u$. Thus $u \varepsilon F(T)$ and $F(T)$ is convex.

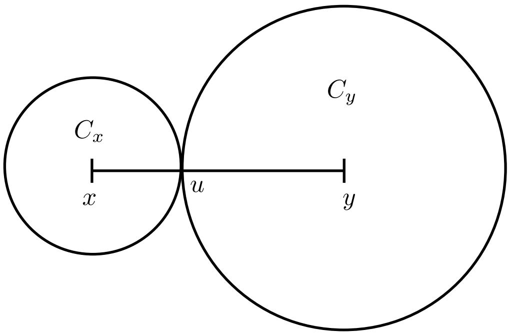

Figure 9.2:

If C is bounded and T is contractive, then

$$F(T) \neq \phi.$$

171 In the following we assume $F(T) \neq \phi$ and study the convergence of the iterative method

$$x^{n+1} = Tx^n.$$

which is known to be strongly convergent to the unique fixed point of T if T is strictly contractive. One has

THEOREM 4. *If T is firmly contractive and $F(T) \neq \phi$ then*

$$x^n \rightharpoonup \xi \; \varepsilon \; F(T) \quad as \quad n \to \infty$$

i.e. x^n converges weakly to a fixed point.

Proof. Let $y\varepsilon F(T)$. We have

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

But

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

Therefore

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

which proves that $\| x^{n+1} - x^n \| \to 0$ and $\{x^n\}$ is a bounded sequence. Let $x^{n'} \rightharpoonup x$ be a weakly convergent subsequence.

Since 172

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

choosing $y = x^{n'-1}$ we obtain

$$\left(Tx - x^{n'},\ Tx - x^{n'} + x^{n'-1} - x\right) \leq 0$$

As $n' \to \infty$, we get

$$(Tx - x,\ Tx - x) \leq 0,$$

and hence $x\varepsilon F(T)$.

As $\| x^n - y \|^2$ is a decreasing sequence for any $y\varepsilon F(T)$ it converges to some number $P(y)$, and we conclude from the following Lemma that the whole sequence x^n converges. ◻

OPAL'S LEMMA 5. *Let $F \subset H$ be a subset of a Hilbert space H and $\{x_n\}$ a sequence such that*

(i) $\| x^n - y \|^2 \to P(y)$ *as $n \to \infty$ for any $y\varepsilon F$*

(ii) *any weakly converging subsequence $x_{n'} \rightharpoonup z$ is such that z belongs actually to F.*

Then $x^n \rightharpoonup \xi \varepsilon F$.

Proof. Let $x_{m'} \rightharpoonup y, x_{n'} \rightharpoonup z$ be two converging subsequences, we have

$$\begin{aligned} \| x_n - y \|^2 &= \| x_n - z + z - y \|^2 \\ &= \| x_n - z \|^2 + 2(x_n - z, z - y) + \| z - y \|^2 \end{aligned}$$

hence taking the limit following m

$$P(y) = P(z) + 2(y - z, z - y) + \| z - y \|^2 = P(z) - \| z - y \|^2$$

173 and taking the limit following n'

$$P(y) = P(z) + 0 + \| z - y \|^2$$

hence $\| z - y \|^2 = 0 \Rightarrow z = y$. ◻

Exercise 3. Prove Theorem 4 when T is quasi firmly contractive.

THEOREM 6. *Let $T = QS$ where S is quasi-firmly contractive and Q is firmly contractive. Then*

$$x^n \rightharpoonup x \; \varepsilon \; F(T),$$

provided that $F(T)$ is non-empty.

Proof. Let $y \varepsilon F(T)$. We have

$$\begin{aligned} \| Sx^n - Sy \|^2 &\leq \theta(Sx^n - Sy, x^n - y) + (1 - \theta) \; \| x^n - y \|^2, \\ (Sx^n - Sy, x^n - y) &= \frac{1}{2} \| Sx^n - Sy \|^2 + \frac{1}{2} \| x^n - y \|^2 \\ &\quad - \frac{1}{2} \| Sx^n - Sy + y - x^n \|^2 . \end{aligned}$$

Therefore,

$$(1 - \frac{\theta}{2}) \| Sx^n - Sy \|^2 + \frac{\theta}{2} \| Sx^n - Sy + y - x^n \|^2 \leq (1 - \frac{\theta}{2}) \| x^n - y \|^2 \quad (9.30)$$

In the same way,

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

i.e. **174**

$$\frac{1}{2} \| x^{n+1} - y \|^2 + \frac{1}{2} \| x^{n+1} - y + S y - S x^n \|^2 \leq \frac{1}{2} \| S x^n - S y \|^2 \quad (9.31)$$

From (9.30) and (9.31) we obtain

$$\| x^{n+1} - y \|^2 + \| x^{n+1} - y + S y - S x^n \|^2 + \alpha \| S x^n - S y + y - x^n \|^2 \leq \| x^n - y \|^2$$

where

$$\alpha = \frac{\theta}{2 - \theta}$$

This implies

$$\| x^{N+1} - y \|^2 + \sum_{n=0}^{N} \Big(\| x^{n+1} - y + S y - S x^n \|^2 +$$
$$+ \alpha \| S x^n - S y + y - x^n \|^2 \Big) \leq \| x^\circ - y \|^2 .$$

Therefore,

$$x^{n+1} - S x^n \to y - S y,$$
$$S x^n - x^n \to S y - y,$$
$$x^{n+1} - x^n \to 0.$$

Let $x^{n'} \rightharpoonup x$, T being contractive, we have

$$\| T x^{n'} - T x \|^2 \leq \| x^{n'} - x \|^2$$

that is

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

and to the limit

$$(x - Tx,\ x + x - Tx - x) \leq 0$$
$$x = Tx.$$

175 Once again we apply Opial's Lemma to get the convergence of the whole sequence x^n to a fixed point of T. ◻

Exercise 4. Let $C \subset V$ be a closed convex subset of a Hilbert space V, then show that the projection map $P_c : V \to C$ is firmly contractive.

9.4 Application to Unconstrained Problem

We shall apply the previous results to the solution of

$$A(u) = 0.$$

where A is a monotone operator from $D(A)$ into H; i.e.

$$(Au - Av,\ u - v) \geq 0 \ \ \forall\ u,\ v\ \varepsilon\ D(A).$$

A is said to be *maximal monotone* if $E \subset H \times H$, Graph $A \subset E$,

$$(x_1 - x_2, y_1 - y_2) \geq 0 \ \ \forall \{x_i, y_i\}\ \varepsilon\ E,\ \ i = 1, 2$$

implies

$$\text{Graph} \quad A = E.$$

It is proved in *BRÉZIS* [4] that

THEOREM 7. *A maximal monotone* iff

$$R(I + \lambda A) = H \quad \textit{for} \quad \lambda \geq 0.$$

176 **EXAMPLE 5.** Let $A : V \to V'$ satisfy (9.15) - (9.17) with

$$\underset{dense}{V \hookrightarrow H} \hookrightarrow V'$$

Then the restriction of A to

$$D(A) = \{v\ \varepsilon\ V :\ Av\ \varepsilon\ H\}$$

is a maximal monotone operator.

Exercise 5. Use Theorem 6 to prove that the operator defined in Example 5 is monotone.

We have

LEMMA 8. *If A is maximal monotone then*

$$T = (I + \lambda A)^{-1}$$

is firmly contractive.

Proof. Let

$$(I + \lambda A)x = (I + \lambda A)y.$$

Then

$$\lambda(A(x) - A(y)) = -(x - y)$$

Therefore

$$- \| x - y \|^2 = \lambda(A(x) - A(y), x - y) \geq 0.$$

Hence $x = y$. This proves $(I + \lambda A)$ is one-one.

From Theorem 6, we obtain $R(I + \lambda A) = H$. Hence $(I + \lambda A)^{-1}$ is well defined on H. **177**

Let

$$u_i = Tx_i, \;\; x_i \varepsilon H, \;\; i = 1, 2.$$

Then

$$u_i + \lambda A u_i = x_i.$$

We have to prove that

$$\| Tx_1 - Tx_2 \|^2 \leq (Tx_1 - Tx_2, x_1 - x_2),$$

i.e.

$$(Tx_1 - Tx_2, Tx_1 - Tx_2 + x_2 - x_1) \leq 0,$$

i.e.

$$(u_1 - u_2, (u_1 - u_2) - (u_1 - u_2) - \lambda(Au_1 - Au_2)) \leq 0,$$

i.e.

$$-\lambda(u_1 - u_2, Au_1 - Au_2) \leq 0,$$

which is true since A is monotone. ◻

COROLLARY 1. *The algorithm*

$$x^{n+1} = (I + \lambda A)^{-1} \; x^n \tag{9.34}$$

converges weakly to a solution of

$$A(u) = 0 \tag{9.35}$$

Note that algorithm (9.34) *can be written as*

$$\frac{x^{n+1} - x^n}{\lambda} + A(x^{n+1}) = 0 \tag{9.36}$$

178 *and corresponds to an implicit scheme for*

$$\frac{\partial u}{\partial t} + A(u) = 0. \tag{9.37}$$

Proof. Since $T = (I + \lambda A)^{-1}$ is firmly contractive, algorithm (9.34) converges weakly to a fixed point of T which is a solution of (9.35). □

REMARK 2. *Algorithm* (9.34) *is called a* proximal point algorithm. *Note that computing x^{n+1} at each step might be as difficult as the original problem except in some special cases.*

REMARK 3. *If $A : V \to V'$ where V is a Hilbert space, then it is better to choose $H = V$. Let $J : V' \to V$ be the Riesz isometry. Then one has to replace A by JA. Then algorithm* (9.34) *is an implicit scheme for*

$$\frac{\partial u}{\partial t} + JA(u) = 0.$$

9.5 Application to Problems with Constraint.

We want to solve the problem

$$(A(u), v - u) \geq 0 \quad \forall\, v \in C. \tag{9.38}$$

If u is a solution of (9.38) then for any $\lambda > 0$ we have

$$(u - \lambda A(u) - u, v - u) \leq 0 \quad \forall\, v \in C$$

179 which implies $u = P_C S u$, where

$$S u = u - \lambda A(u).$$

Conversely if u is a fixed point of $P_C S$, then u is a solution of (9.38). We like to solve (9.38) via the algorithm

$$x^{n+1} = P_C S x^n = P_C(x^n - \lambda A(x^n)). \tag{9.39}$$

Note that if J is a convex, *l.s.c.*, Gateaux differentiable function and $A = J'$ then (9.38) is the gradient algorithm with projection for solving

$$\operatorname*{Inf}_{v \varepsilon C} J(v).$$

We will now give some conditions on A and λ which will ensure the convergence of the algorithm (9.39) to a solution of (9.38).

THEOREM 9. *If A is strongly monotone, i.e.*

$$(A(u) - A(v), u - v) \geq \alpha \parallel u - v \parallel^2 \quad \forall\, u,\ v\ \varepsilon\ C \tag{9.40}$$

and Lipshitzian,

$$\parallel A(u) - A(v) \parallel \leq c \parallel u - v \parallel \quad \forall\, u, v\ \varepsilon\ C \tag{9.41}$$

then the algorithm (9.39) *converges strongly to the solution* (9.38) *for all* $0 < \lambda < 2\alpha/c$.

Proof. S is strictly contractive for $0 < \lambda < 2\alpha/c$. Indeed 180

$$\begin{aligned}\parallel S u - S v \parallel^2 = \parallel u - v \parallel^2 & -2\lambda(A(u) - A(v), u - v) + \\ +\lambda^2 \parallel A(u) - A(v) \parallel^2 \leq \parallel u - v \parallel^2 & (1 - 2\lambda\alpha + \lambda^2 c^2)\end{aligned}$$

by (9.40) and (9.41) and

$$1 - 2\alpha\lambda + \lambda^2 c^2 < 1 \quad \text{for} \quad 0 < \lambda < \frac{2\alpha}{c^2}$$

From exercise 4, we know that P_C is firmly contractive. Therefore $P_C S$ is strictly contractive for $0 < \lambda < 2\alpha/c^2$. Thus the algorithm (9.39) converges strongly to the solution of (9.38). $\square$

We will now give a condition on A which will imply weak convergence of the algorithm (9.39).

THEOREM 10. *If A^{-1} is coercive, namely*

$$(A(u) - A(v), u - v) \geq \alpha \parallel A(u) - A(v) \parallel^2 \quad \forall\ u, v\ \varepsilon\ C \tag{9.42}$$

then for $0 < \lambda < 2\alpha$ *the algorithm* (9.39) *converges weakly to a solution of* (9.38).

Proof. We claim that S is quasi firmly contractive for $\alpha < \lambda < 2\alpha$. In fact,

$$\begin{aligned}\parallel Su - Sv \parallel^2 &\leq \parallel u - v \parallel^2 + \left(\frac{\lambda^2}{\alpha} - 2\lambda\right)(A(u) - A(v), u - v) \quad \text{by (9.42)}\\ &= (1 - \theta)\ \parallel u - v \parallel^2 + \theta(Su - Sv, u - v)\end{aligned} \tag{9.43}$$

181 where

$$\theta = 2 - \lambda/\alpha.$$

When $\alpha < \lambda < 2\alpha$, we have $0 < \theta < 1$.

When $0 < \lambda < \alpha$ we obtain S to be firmly contractive. To prove this use (9.43), the Schwarz inequality and the fact that $\theta \varepsilon [1, 2]$ when $0 < \lambda < \alpha$. Thus S is quasi firmly contractive for $0 < \lambda < 2\alpha$. Using Theorem 5, we obtain the conclusion of the Theorem. □

REMARK 4. *When A satisfies* (9.42) *and* $0 < \lambda < 2\alpha$, *we obtain from the proof of Theorem 5 that*

$$\lambda A(x^n) = x^n - Sx^n \to x - Sx = \lambda A(x),$$

i.e. $\qquad A(x^n) \to A(x), \qquad$ *(Strong convergence)*

whereas

$$x^n \to x. \qquad \text{(Weak convergence)}$$

We also notice that $x - Sx$ is unique and therefore $A(x)$ (x, the solution of (9.38)*, need not be unique).*

EXAMPLE 6. Let

$f : H \to \mathbb{R}$ be convex, *l.s.c.* differentiable and

$A : V \to H$ be a linear operator.

We want to solve

$$\operatorname*{Inf}_{v \varepsilon V} f(Av). \tag{9.44}$$

Note that (9.44) is equivalent to **182**

$$\operatorname*{Inf}_{y \varepsilon C} f(y), \tag{9.45}$$

where $C = R(A)$, the range of A.

Now apply the algorithm (9.39). The projection on C is easy to compute. In fact,

$$P_C = A(A^* A)^{-1} A^*.$$

The nonlinear Dirichlet problem and the Minimal surface problem are particular cases of the abstract problem.

EXAMPLE 7. Let us consider

$$\begin{aligned} \frac{\partial u}{\partial t} + Au &= 0, \\ u(0) &= u^\circ, \end{aligned} \tag{9.46}$$

which has a solution provided A is maximal monotone.

We like to solve this problem via the algorithm

$$u^{n+1} = F(\lambda)\, u^n. \tag{9.47}$$

In BREZIS [4] one can find the proof of

THEOREM 11. *If $F(\lambda)$ is a contraction and if*

$$\lim_{\lambda \to 0} \frac{x - F(\lambda)x}{\lambda} = A(x) \quad \textit{exists}, \tag{9.48}$$

then **183**

$$(F(t/n))^n\, u^\circ \to u(t) \quad \textit{uniformly}.$$

Applications:

1. If $F(\lambda) = I - \lambda A$, where A satisfies (9.40), then $F(\lambda)$ is a contraction for $0 < \lambda < 2\alpha$. The limit in (9.48) exists and hence the algorithm (9.47) converges.

2. Let $F(\lambda) = (I + \lambda A)^{-1}$, where A is maximal monotone. Then by Lemma 7, $F(\lambda)$ is firmly contractive and hence contractive. Existence of the limit (9.48) is proved in *BRÉZIS* [4] In this case also algorithm (9.47) converges.

REMARK 5. *Theorem 11 can be used to prove that $F(\lambda) = P_C(I - \lambda A)$ gives a sequence converging to the solution of*

$$\left(\frac{du}{dt} + Au,\ v - u\right) \geq 0 \quad \forall\ v\ \varepsilon\ C,$$
$$u(0) = u^{\circ}.$$

REMARK 6. *If A is linear, monotone and closed then A is maximal monotone.*

EXAMPLE 8. The Flow of a Bingham Fluid: Consider the problem:

Find $\sigma\varepsilon(L^2(\Omega)), u\varepsilon H^1_{\circ}(\Omega)$ such that

$$J'(\sigma) - \nabla u = 0 \tag{9.49}$$

$$(\sigma, \nabla v) = (f, v) \quad \forall\ v\ \varepsilon\ H^1_{\circ}(\Omega),$$

184 where

$$J(\sigma) = \frac{1}{2} \parallel \sigma - P_K\sigma \parallel^2\ (L^2(\Omega))^n,$$
$$K = \{\sigma\ \varepsilon\ (L^2(\Omega))^n : |\sigma(x)| \leq 1 \text{ a. e.} \quad \text{in} \quad \Omega\}.$$

It is possible to prove that (9.49) is equivalent to the Bingham flow given in Example 4. It can be proved that

$$J'(\sigma) = \sigma - P_K\sigma.$$

Let

$$Z(f) = \{\sigma \,\varepsilon\, (L^2(\Omega))^2 : (\sigma, \nabla v) = (f, v) \;\; \forall\, v \,\varepsilon\, H^1_\circ(\Omega)\}$$

If σ is a solution of (9.49), then σ is also a solution of

$$J(\sigma) = \operatorname*{Inf}_{\tau \varepsilon Z(f)} J(\tau). \tag{9.50}$$

Therefore, we can apply previous results. Note that J' satisfies (9.42) with $\alpha = 1$, so that previous results can be applied. Note also that the projection on $Z(f)$ is easy to compute:

$$P_{Z(f)}(\sigma) = \sigma + \nabla(-\Delta)^{-1} \operatorname{div} \sigma + \nabla(-\Delta)^{-1} f.$$

EXAMPLE 9. Transonic Flows.

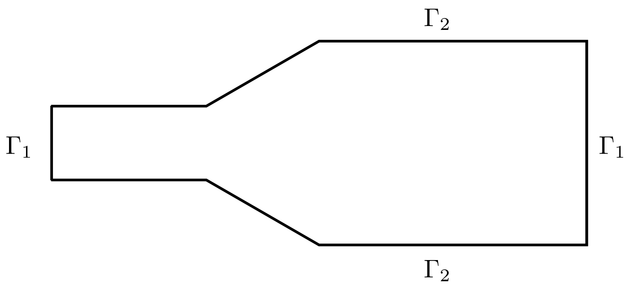

Figure 9.3:

The potential flow at transonic speed in a nozzle is governed by the 185
equations

$$\begin{aligned} \operatorname{div}(\rho \vec{q}) &= 0, \\ \operatorname{rot} \vec{q} &= 0, \vec{q} = \nabla\phi, \\ \rho &= \left(1 - \frac{\gamma - 1}{2} M_\infty^2 (1 - q^2)\right)^{1/\gamma - 1}. \end{aligned}$$

(Practical value of $\gamma = 1.4$) where $q = |\vec{q}|$.

Finally we solve

$$\operatorname{div}(\rho(|\nabla\phi|)\nabla\phi) = 0,$$

$$\phi|_{\Gamma_1} = \phi_\circ,$$
$$\frac{\partial \phi}{\partial n}|_{\Gamma_2} = 0$$

Let $M = q/a$, where $a = \frac{\rho^{\gamma-1}}{M_\infty^2}$. M is called the *Mach number*.

The equation is elliptic for $M < 1$ and hyperbolic for $M > 1$.

When $M < 1$ continuous piecewise linear finite element can be used. For $M > 1$, we do not know much (see COURANT- FRIEDRICHS [13]).

The reader can refer to GLOWINSKI-PIRONNEAU [20],[21], RAVIART [36], CIAVALDINI-POGU-TOURNEMINE [12], J. ROUX [40].

Bibliography

[1] ADAMS, R.A.: *Sobolev Spaces*, Academic Press, 1976. **186**

[2] AMARA, M. and J.M. THOMAS: *Une méthode d'éléments finis équilibre pour le problème de l'élasticité linéaire*, C.R. Acad. Sci. Paris, Série A, Séance du 3 Avril 1978.

[3] BERCOVIER, M. and O. PIRONNEAU: *Estimations d'erreur pour la résolution du problème de STokes en éléments finis conformes de Lagrange*, C.R. Acad. Sci. Paris, t 285 Série A (1977), 557 – 559.

[4] BRÉZIS, H.: *Operateurs Maximaux Monotones*, Lecture Notes # 5, North Holland, 1975, RAIRO, 1974.

[5] BREZZI, F.: *On the existence, Uniqueness and Approximation of Saddle point problems arising from lagrangian multipliers*, RAIRO, Vol. 8, 1974, 129 – 151.

[6] BREZZI, F.; J. RAPPAZ and P.A. RAVIART: *Finite element approximation of bifurcation problem (to appear)*.

[7] BREZZI, F. and P.A. RAVIART: *Mixed finite element methods for* 4^{th} *order elliptic equations*, Topics in Numerical Analysis III, ed. by John, J.H. Miller, Academic Press, 33 – 57.

[8] BROWDER, F.E. and W.V. PETRYSHN: *Construction of fixed points of nonlinear mappings in Hilbert Spaces*, J. Math. Anal. Appl., 20 (1977), 197 – 228.

187 [9] CIARLET, P.G.: *The Finite Element Method for elliptic problems*, North Holland, 1978.

[10] CIARLET, P.G. and P.A. RAVIART: *A Finite Element Method for the Biharmonic Equation*, Mathematical Aspects of Finite Elements in Partial Differential Equations, by Carl de Boor, Academic Press, 1974, 125 – 146.

[11] CIAVALDINI, J.F. and J.C. NEDELEC: *Sur l'element de Fraeijs de Veubeke et Sander*, Serie Analyse Numerique, RAIRO, 1974, 29 – 46.

[12] CIAVALDINI, J.F.: M.POGU and G. TOURMINE: *Une nouvelle approche daus le plan physique pour le calcul d'elements subcritiques et Stationnaires autour d'un profil portant*, Journal de mécanique, 16 (1977) 257 – 288.

[13] COURANT, R and K.O. FRIEDRICHS: *Supersonic flow and Shock Waves*, J. Wiley, 1948.

[14] CROUZEIX, M. and P.A. RAVIART: *Conforming and Nonconforming Finite Elements Methods for solving the Stationary Stokes Equations*, RAIRO, 1974, 1 – 53.

[15] EKELAND, I. and R. THEMAN: *Convex Analysis and Variational Problems*, North-Holland, 1976.

[16] FORTIN, M.: *Approximation des Fonctions a Divergence Nulle par la Methode des Elements Fini*, Springer-Verlag Lecture Notes in Physics, # 18, 1972, 99 – 103.

188 [17] FORTIN, M.: *Résolution Numerique des Equations de Navier-Stokes par des Eléments Fini de type Mixte*, IRIA Report, # 184, 1976.

[18] FORTIN, M: *Analysis of the convergence of mixed finite element methods*, Serie Numerical Analysis, RAIRO, 1977 341 – 354.

[19] GLOWINSKI, R. and MARROCCO: *Approximation par éléments Finis d'ordre un et resolution par pénalisation dualite d'une classe de problemes nonlineariès*, Serie Analyse Numerique, RAIRO, 1975, 41 – 76.

[20] GLOWINSKI, R.; J. PERIAUX and O. PIRONNEAU: *Use of Optimal Control Theory for the Numerical Simulation of Transonic Flow by the Method of Finite Elements*, Springer-Verlag, Lecture Notes in Physics, # 59, 1976, 205 – 211.

[21] GLOWINSKI, R. and O. PIRONNEAU: *Calcul d ecoulements transoniques par des méthodes d'éléments finis et de controle optimal*, Springer-Verlag Lecture Notes in Economics and Mathematical systems, # 134, 1975, 276 – 296.

[22] GRISVARD, P.: *Behaviour of the solutions of an Elliptic Boundary Value Problem in a polygonal or polyhedral Domain*, Numerical Solution of Partial Differential Equations III, Synspade, 1975, ed. by Vert Hubbard, Academic Press, 1976, 207 – 274.

[23] JAMET, P. and P.A. RAVIART: *Numerical solution of the station-* **189**
ary Navier-Stokes Equations by finite element methods, Springer-Verlag Lecture Notes in Computer Science, # 10, 1973, 193 – 223.

[24] JOHNSON, C: *On the convergence of some mixed finite element methods in plate bending problems*, Num. Math. 21 (1973), 43 – 62.

[25] JOHNSON, C. and B. MERCIER: *Some mixed finite element methods for elasticity problems*, Num. Math., 30 (1978) 103 – 116.

[26] KATO, T.: *Perturbation Theory for linear operators*, Springer-Verlag 1976.

[27] LADYZHENSKAYA, O.A.: *The Mathematical Theory of Viscous Incompressible Flow*, Gorden and Breach, 1969.

[28] LIONS, J.L.: *Quelques méthodes de résolution des problèms aux limites nonlinéaries*, Paris, Dunod, 1969.

[29] LIONS, J.L. and E. MAGENES: *Non Homogeneous Boundary Value Problems and Applications*, Vol. I, Springer-Verlag, 1973.

[30] LUENBERGER, D.G.: *Introduction to linear and nonlinear programming*, Addison-Wesley, 1973.

[31] LUENBERGER, D.G.: *Optimization by Vector Space Methods*, John Wiley, 1969.

[32] MERCIER, B. and O. PIRONNEAU: *Some Examples of implementation and of application of the finite element method,*

190 [33] NEČAS, J: *Les méthodes directes en theorie des equations elliptiques*, Masson, 1967.

[34] OSBORN, J.E.: *Spectral approximation for compact operators*, Mathematics of Computations, 29 (1975) 712 – 725.

[35] PAIGE, C.C. and M.A. SAUNDERS: *Solution of sparse indefinite systems equations*, SIAM, J. Num. Analysis, 12 (1975), 617 – 629.

[36] RAVIART, P.A.: *Journees elements finis*, Conference, Rennes, 1978.

[37] RAVIART, P.A. and J.M. THOMAS: *Primal hybrid finite element methods for* 2^{nd} *order Elliptic equations*, Mathematics of Computations, 31 (1977), 391 – 413.

[38] RAVIART, P.A. and J.M. THOMAS: *A mixed finite element method for second order elliptic problems*, Springer-Verlag Lecture Notes, # 606, ed. by Galligani, I. and E. Magenes (1975), 292 – 315.

[39] de RHAM, G.: *Varietes differentiables*, Hermann, 1950.

[40] ROUX, J.: *Resolution Numerique d'un probleme d Ecoulement Subsonique de Fluides Compressibles*, Springer-Verlag Lecture Notes in Physics, # 59, 1976, 360 – 369.

[41] SCHOLZ, R.: *A mixed method for* 4^{th} *order problems using linear finite elements*, Serie Numerical Analysis, RAIRO, Vol. 12, 1978, 85 – 90.

[42] TAYLOR, C. and P. HOOD: *A numerical solution of the Navier-* **191**
Stokes Equation using the finite elements technique, Computers and Fluids, Vol. I, 1973, 73 – 100.

[43] TEMAN, R.: *On the theory and numerical analysis of Navier-Stokes equations*, North-Holland, 1977.

[44] THOMEE, V. : *Some convergence results for Galerkin Methods for Parabolic Boundary Value Problems*, Mathematical Aspects of Finite Elements in Partial Differential Equations, ed. by Carl de Boor, Academic Press, 1974, 55 – 88.

[45] THOMEE, V.: *Some Error Estimates in Galerkin Methods for Parabolic Equations*, Springer-Verlag Lecture Notes in Mathematics, # 606, 1975, 343 – 352.

[46] YOSIDA, K.: *Functional Analysis*, Springer-Verlag, 1974.

[47] SIENKIEWICZ, O.C.: *Why Finite Elements?* Finite elements in Fluids, Vol. I, ed. by Gallagher, John Wiley, 1975, 1 – 24.