Lectures on
The Finite Element Method

By

Ph. Ciarlet

Tata Institute of Fundamental Research

Bombay

1975

Lectures on
The Finite Element Method

By

Ph. Ciarlet

Notes by

S. Kesavan, Akhil Ranjan

M. Vanninathan

Tata Institute of Fundamental Research

Bombay

1975

Printed In India

By

Anil D. Ved At Prabhat Printers, Bombay 400004

And Published By

The Tata Institute of Fundamental Research, Bombay.

Acknowledgements

THE AUTHOR EXPRESSES his deep gratitude to Professor Ramanathan for his very kind invitation to deliver these lectures at Bangalore, as part of the T.I.F.R.–I.I.Sc. Mathematics Programme. Also, the author tanks the T.I.F.R. for financial support.

Ph. Ciarlet

Preface

OUR BASIC AIM has been to present *some* of the mathematical aspects of the finite element method, as well as *some* applications of the finite element method for solving problems in Elasticity. This is why important topics, such as curved boundaries, mixed and hybrid methods, time-dependent problems, etc..., are not covered here. No attempt has been made to give an exhaustive bibliography.

Contents

Chapter 1

The Abstract Problem

SEVERAL PROBLEMS IN the theory of Elasticity boil down to the solution of a problem described, in an abstract manner, as follows: 1

Let V be a normed linear space over $\mathbb{R}$. Let $J : V \to \mathbb{R}$ be a functional which can be written in the form

$$J(v) = \frac{1}{2}a(v, v) - f(v) \quad \text{for all} \quad v \in V, \tag{1.1}$$

where $a(\cdot, \cdot)$ is a continuous, symmetric bilinear form on V and f is an element of V', the dual of V. Then the problem consists in finding an element $u \in V$ such that

$$J(u) = \operatorname*{Min}_{v \in V} J(v). \tag{1.2}$$

Usually J represents the *energy* of some physical system.

More often, instead of minimising J over the entire space V, we do so over a non-empty convex subset K of V and find a element $u \in K$ such that

$$J(u) = \operatorname*{Min}_{v \in K} J(v). \tag{1.3}$$

Henceforth we shall denote this abstract problem by the symbol (P). One can ask immediately whether this problem admits of a solution and if so, is the solution unique? We present in this section the essential results regarding existence and uniqueness.

Definition 1.1. Let V be a normed linear space. A bilinear form $a(\cdot,\cdot)$ on V is said to be V-elliptic if there exists a constant $\alpha > 0$ such that for all $v \in V$.

$$a(v, v) \geq \alpha \|v\|^2. \tag{1.4}$$

2

Theorem 1.1. *Let V be a Banach space and K a closed convex subset of V. Let $a(\cdot,\cdot)$ be V-elliptic. Then there exists a unique solution for the problem (P).*

Further this solution is characterised by the property:

$$a(u, v-u) \geq f(v-u) \quad \textit{for all} \quad v \in K. \tag{1.5}$$

Remark 1.1. The inequalities (1.5) are known as *variational inequalities*.

Proof. The V-ellipticity of $a(\cdot,\cdot)$ clearly implies that if $a(v, v) = 0$ then $v = 0$. This together with the symmetry and bilinearity of $a(\cdot,\cdot)$ shows that $a(\cdot,\cdot)$ defines an inner-product on V. Further the continuity and the V-ellipticity of $a(\cdot,\cdot)$ shows that the norm

$$v \in V \to a(v, v)^{\frac{1}{2}} \tag{1.6}$$

defined by the inner-product is equivalent to the existing norm on V. Thus V acquires the structure of a Hilbert space and we apply the Riesz representation theorem to obtain the following: for all $f \in V'$, there exists $\sigma f \in V$ such that

$$f(v) = a(\sigma f, v) \quad \text{for all} \quad v \in V. \tag{1.7}$$

The map $\sigma : V' \to V$ given by $f \mapsto \sigma f$ is linear. Now,

$$\begin{aligned} J(v) &= \frac{1}{2} a(v, v) - f(v) \\ &= \frac{1}{2} a(v, v) - a(\sigma f, v) \end{aligned}$$

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

The symmetry of $a(\cdot, \cdot)$ is essential in obtaining the last equality.
3 For a given f, since σf is fixed, J is minimised if and only if $a(v - \sigma f, v - \sigma f)$ is minimised. But this being the distance between v and σf, our knowledge of Hilbert space theory tells us that since K is a closed convex subset, there exists a unique element $u \in K$ such that this minimum is obtained. This proves the existence and uniqueness of the solution, which is merely the projection of σf over K.

We know that this projection is characterised by the inequalities:

$$a(\sigma f - u, v - u) \leq 0 \quad \text{for all} \quad v \in K. \tag{1.8}$$

Geometrically, this means that the angle between the vectors $(\sigma f - u)$ and $(v - u)$ is obtuse. See Fig. 1.1.

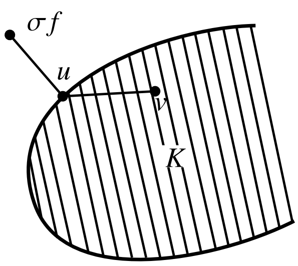

Figure 1.1:

Thus, $a(\sigma f, v - u) \leq a(u, v - u)$ which by virtue of (1.7) is precisely the relation (1.5). This completes the proof. $\square$

We can state the following

Corollary 1.1. (a) *If K is a non-empty closed convex cone with vertex at origin 0, then the solution of (P) is characterised by:*

$$\begin{cases} a(u, v) \geq f(v) & \textit{for all} \quad v \in K \\ a(u, u) = f(u). \end{cases} \tag{1.9}$$

(b) *If K is a subspace of V then the solution is characterised by* 4

$$a(u, v) = f(v) \quad \textit{for all} \quad v \in K. \tag{1.10}$$

Remark 1.2. The relations (1.5), (1.9) and (1.10) are all called *variational formulations* of the problem (P).

Proof. (a) If K is a cone with vertex at 0, then for $u, v \in K$, $u + v \in K$. (cf. Fig. 1.2). If u is the solution to (P), then for all $v \in K$ applying (1.5) to $(u + v)$ we get $a(u, v) \geq f(v)$ for all $v \in K$. In particular this applies to u itself. Setting $v = 0$ in (1.5) we get $-a(u, u) \geq -f(u)$ which gives the reverse inequality necessary to complete the proof of (1.9). Conversely, if (1.9) holds, we get (1.5) by just subtracting one inequality from the other.

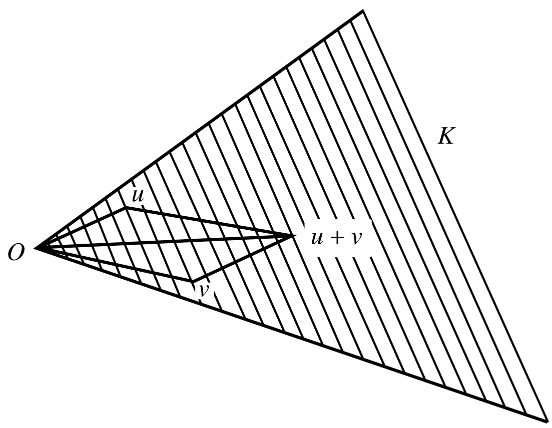

Figure 1.2:

(b) Applying (a) to K, since any subspace is a cone with vertex at 0, we get (b) immediately. For if $v \in K$, then $-v \in K$ and applying (1.9) both to v and $-v$ we get (1.10).

This completes the proof. □

Remark 1.3. The solution u of (P) corresponding to $f \in V'$ (for a fixed $a(\cdot, \cdot)$) defines a map $V' \to V$. Since this solution is the projection of

σf on K, it follows that *the above map is linear if and only if K is a subspace.* The problems associated with variational inequalities are therefore nonlinear in general.

Exercise 1.1. Let V be as in Theorem 1.1. For f_1, $f_2 \in V'$, let u_1, u_2 5
be the corresponding solutions of (P). If $\|\cdot\|^*$ denotes the norm in V', prove that

$$\|u_1 - u_2\| \leq \frac{1}{\alpha}\|f_1 - f_2\|^*.$$

Remark 1.4. The above exercise shows, in particular, the continuous dependence of the solution of f, in the sense described above. This together with the existence and uniqueness establishes that the problem (P) is "well-posed" in the terminology of partial differential equations.

Exercise 1.2. If V is a normed linear space, K a given convex subset of V and $J : V \to \mathbb{R}$ any functional which is once differentiable everywhere, then (i) if $u \in K$ is such that $J(u) = \operatorname*{Min}_{v \in K} J(v)$, u satisfies, $J'(u)(v-u) \geq 0$ for all $v \in K$. (ii) Conversely, if $u \in K$ such that $J'(u)(v-u) \geq 0$ for all $v \in K$, and J is everywhere twice differentiable with J'' satisfying $J''(v)(w,w) \geq \alpha\|w\|^2$, for all v, $w \in K$ and some $\alpha \geq 0$, then $J(u) = \operatorname*{Min}_{v \in K} J(v)$.

Exercise 1.3 ([1])**.** Apply the previous exercise to the functional

$$J(v) = \frac{1}{2}a(v,v) = f(v)$$

with $a(\cdot,\cdot)$ and f as in Theorem 1.1. If K is a subspace of V, show that $J'(u)(v) = 0$ for all $v \in K$. In particular if $K = V$, $J'(u) = 0$.

It was essentially the symmetry of the bilinear form which provided the Hilbert space structure in Theorem 1.1. We now drop the symmetry assumption on $a(\cdot,\cdot)$ but we assume V to be a Hilbert space. In addition we assume that $K = V$.

Theorem 1.2 (LAX-MILGRAM LEMMA)**.** *Let V be a Hilbert space.* 6

[1]Exercises (1.2)(i) and 1.3 together give relations (1.5)

$a(\cdot,\cdot)$ a continuous, bilinear, V-elliptic form, $f \in V'$. If (P) is the problem: to find $u \in V$ such that for all $v \in V$,

$$a(u,v) = f(v), \tag{1.11}$$

then (P) has a unique solution in V[1].

Proof. Since $a(\cdot,\cdot)$ is continuous and V-elliptic, there are constants M, $\alpha > 0$ such that

$$\begin{aligned} |a(u,v)| &\leq M\|u\|\,\|v\|, \\ a(v,v) &\geq \alpha\|v\|^2, \end{aligned} \tag{1.12}$$

for all $u, v \in V$. Fix any $u \in V$. Then the map $v \mapsto a(u,v)$ is continuous and linear. Let us denote it by $\mathrm{Au} \in V'$. Thus we have a map $A : V \to V'$ defined by $u \mapsto \mathrm{Au}$.

$$\|\,\mathrm{Au}\,\|^* = \sup_{\substack{v\in V\\ v\neq 0}} \frac{|\,\mathrm{Au}(v)|}{\|v\|} = \sup_{\substack{v\in V\\ v\neq 0}} \frac{|a(u,v)|}{\|v\|} \leq M\|u\|. \tag{1.13}$$

Thus A is continuous and $\|A\| \leq M$.

We are required to solve the equation

$$\mathrm{Au} = f. \tag{1.14}$$

Let τ be the Riesz isometry, $\tau : V' \to V$ so that

$$f(v) = ((\tau f, v)), \tag{1.15}$$

where $((\cdot,\cdot))$ denotes the inner product in V. Then, $\mathrm{Au} = f$ if and only if $\tau\,\mathrm{Au} = \tau f$ or equivalently.

$$u = u - \rho(\tau\,\mathrm{Au} - \tau f), \tag{1.16}$$

7 where $\rho > 0$ is a constant to be specified. We choose ρ such that $g : V \to V$ is a contraction map, where g is defined by

$$g(v) = v - \rho(\tau Av - \tau f) \quad \text{for} \quad v \in V. \tag{1.17}$$

[1] cf. Corollary 1.1(b).

Then the solution to (P) will be the unique fixed point of this contraction map, which exists by the contraction mapping theorem.

Let $v_1, v_2 \in V$. Set $v = v_1 - v_2$. Then

$$\begin{aligned} ||g(v_1) - g(v_2)|| &= ||(v_1 - v_2) - \rho\tau A(v_1 - v_2)|| \\ &= ||v - \rho\tau Av||. \end{aligned}$$

But,

$$\begin{aligned} ||v - \rho\tau Av||^2 &= ((v - \rho\tau Av, v - \rho\tau Av)) \\ &= ||v||^2 - 2\rho((\tau Av, v)) + \rho^2||\tau Av||^2 \\ &= ||v||^2 - 2\rho Av(v) + \rho^2||Av||^{*2} \\ &\le ||v||^2 - 2\rho\alpha||v||^2 + \rho^2 M^2||v||^2 \\ &= (1 - 2\rho\alpha + \rho^2 M^2)||v||^2 \end{aligned}$$

since $Av(v) = a(v, v) \ge \alpha||v||^2$ and $||A|| \le M$. Choosing $\rho \in]0, \dfrac{2\alpha}{M^2}[$, we get that

$$1 - 2\rho\alpha + \rho^2 M^2 < 1 \tag{1.18}$$

and hence g is a contraction, thus completing the proof. ◻

Remark 1.5. The problem (P) of Theorem 1.2 is well-posed. The existence and uniqueness were proved in the theorem. For the continuous dependence of u on f, we have

$$\alpha||u||^2 \le a(u, u) = f(u) \le ||f||^* \cdot ||u||. \tag{1.19}$$

8

REFERENCE. For Variational Inequalities, see Lions and Stampacchia [18].

Chapter 2

Examples

WE GIVE IN this section several examples of the abstract problem formulated in Sec. 1. We interpret the solutions of these problems as solutions of classical boundary value problems which often occur in the theory of Elasticity. 9

Before we proceed with the examples, we summarize briefly the results (without proofs) on Sobolev spaces which will prove to be very useful in our discussion.

Henceforth $\Omega \subset \mathbb{R}^n$ will denote an open set (more often Ω will be a bounded open set with a specific type of boundary which will be described presently). A *multi-index* α will denote an n-tuple $(\alpha_1, \ldots, \alpha_n)$ of non-negative integers, and we denote

(2.1) $$|\alpha| = \alpha_1 + \cdots + \alpha_n,$$

and call it the *length* of the multi-index. If v is a real-valued function on Ω for which all derivatives upto order m exist, for a multi-index α with $|\alpha| \leq m$ we define

(2.2) $$\partial^{\alpha_v} = \frac{\partial^{|\alpha|_v}}{\partial^{\alpha_1}_{x_1} \ldots \partial^{\alpha_n}_{x_n}}.$$

The space of *test functions* on Ω is given by

(2.3) $$\mathscr{D}(\Omega) = \{v \in C^\infty(\Omega); \operatorname{supp}(v) \text{ is a compact subset of } \Omega\}.$$

where

$$\operatorname{supp}(v) = \overline{\{x \in \Omega; v(x) \neq 0\}}. \tag{2.4}$$

10 **Definition 2.1.** Let $m \geq 0$ be an integer. Then the Sobolev space $H^m(\Omega)$ is given by

$$H^m(\Omega) = \left\{ v \in L^2(\Omega); \partial^\alpha v \in L^2(\Omega) \text{ for all } |\alpha| \leq m \right\}, \tag{2.5}$$

where all derivatives are understood in the sense of distributions.

On $H^m(\Omega)$ one can define a norm by means of the formula

$$\|v\|_{m,\Omega} = \left(\sum_{|\alpha| \leq m} \int_\Omega |\partial^\alpha v|^2 dx \right)^{\frac{1}{2}}, \quad v \in H^m(\Omega). \tag{2.6}$$

It is easy to check that $\|\cdot\|_{m,\Omega}$ defines a norm on $H^m(\Omega)$, which makes it a Hilbert space. One can also define a semi-norm by

$$|v|_{m,\Omega} = \left(\sum_{|\alpha| = m} \int_\Omega |\partial^\alpha v|^2 dx \right)^{\frac{1}{2}}, \quad v \in H^m(\Omega). \tag{2.7}$$

Note that since for all $m \geq 0$, $\mathscr{D}(\Omega) \subset H^m(\Omega)$, we may define,

$$H_0^m(\Omega) = \overline{\mathscr{D}(\Omega)}, \tag{2.8}$$

the closure being taken with respect to the topology of $H^m(\Omega)$. Since $H_0^m(\Omega)$ is a closed subspace of $H^m(\Omega)$, it is also a Hilbert space under the restriction of the norm $\|\cdot\|_{m,\Omega}$. We also have a stronger result:

Theorem 2.1. *Assume that Ω is a bounded open set. Then over $H_0^m(\Omega)$ the semi-norm $|\cdot|_{m,\Omega}$ is a norm equivalent to the norm $\|\cdot\|_{m,\Omega}$.*

This result is a consequence of the following:

Theorem 2.2 (POINCARÉ-FRIEDRICHS' INEQUALITY)**.** *If Ω is a bounded open set, there exists a constant $C = C(\Omega)$ such that, for all $v \in H_0^1(\Omega)$,*

$$|v|_{0,\Omega} \leq C|v|_{1,\Omega}. \tag{2.9}$$

11 Henceforth, unless specified to the contrary, the following will be our standing assumptions: Ω *is a bounded open subset of* $\mathbb{R}^n$. *If* Γ *is the boundary of* Ω*, then* Γ *is Lipschitz continuous in the sense of Nečas* [20]. (Essentially, Γ can be covered by a finite number of local coordinate systems, such that in each, the corresponding portion of Γ is described by a Lipschitz continuous function).

If $L^2(\Gamma)$ is defined in the usual fashion using the Lipschitz continuity of Γ, one has the following result:

Theorem 2.3. *There exists a constant* $C = C(\Omega)$ *such that, for all* $v \in C^\infty(\Omega)$*,*

$$\|v\|_{L^2(\Omega)} \leq C\|v\|_{1,\Omega}. \tag{2.10}$$

By virtue of Theorem 2.3 we get that if $v \in C^\infty(\overline{\Omega})$, then its restriction to Γ is an element of $L^2(\Gamma)$. Thus we have a map from the space $C^\infty(\overline{\Omega})$ equipped with the norm $\|\cdot\|_{1,\Omega}$ into the space $L^2(\Gamma)$ which is continuous. We also have:

Theorem 2.4. *The space* $C^\infty(\overline{\Omega})$ *is dense in* $H^1(\Omega)$*, for domains with Lipschitz continuous boundaries.*

Consequently, the above map may be extended to a continuous map $H^1(\Omega) \to L^2(\Gamma)$ which we denote by tr_Γ. It is called the *trace operator*. An important result on the trace is the characterization:

$$H_0^1(\Omega) = \left\{v \in H^1(\Omega);\ \mathrm{tr}_\Gamma\, v = 0\right\}. \tag{2.11}$$

When no confusion is likely to occur we will merely write v instead of $\mathrm{tr}_\Gamma\, v$. In fact if v is a "smooth" function then $\mathrm{tr}_\Gamma\, v$ is the restriction of v to Γ.

Retaining our assumption on Ω and Γ, the unit outer normal $\vec{\nu}$ is 12
defined a.e. on Γ. Let $\vec{\nu} = (\nu_1, \ldots, \nu_n)$. If v is smooth then we may define the *outer normal derivative* $\dfrac{\partial v}{\partial \nu}$ by

$$\frac{\partial v}{\partial \nu} = \sum_{i=1}^{n} \nu_i \frac{\partial v}{\partial x_i}. \tag{2.12}$$

We extend this definition to $v \in H^2(\Omega)$. If $v \in H^2(\Omega)$, then $\dfrac{\partial v}{\partial x_i} \in H^1(\Omega)$ and hence $\mathrm{tr}_\Gamma \dfrac{\partial v}{\partial x_i} \in L^2(\Gamma)$. We now define

$$\frac{\partial v}{\partial \nu} = \sum_{i=1}^{n} \nu_i \,\mathrm{tr}_\Gamma \frac{\partial v}{\partial x_i}. \tag{2.13}$$

However when there is no confusion we write it in the form of (2.12). Then one has the following characterization:

$$H_0^2(\Omega) = \left\{ v \in H^2(\Omega);\ v = \frac{\partial v}{\partial \nu} = 0 \ \text{ on } \ \Gamma \right\}. \tag{2.14}$$

Theorem 2.5 (GREEN'S FORMULA IN SOBOLEV SPACES)**.** *Let* u, $v \in H^1(\Omega)$. *Then we have*

$$\int_\Omega u \frac{\partial v}{\partial x_i} dx = -\int_\Omega \frac{\partial u}{\partial x_i} v\, dx + \int_\Gamma u\, v \nu_i\, d\gamma, \tag{2.15}$$

for all $1 \leq i \leq n$.

If we assume $u \in H^2(\Omega)$, we may replace u in (2.15) by $\dfrac{\partial u}{\partial x_i}$; summing over all $1 \leq i \leq n$, we get for $u \in H^2(\Omega)$, $v \in H^1(\Omega)$,

$$\int_\Omega \sum_{i=1}^{n} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} dx = -\int_\Omega \Delta u\, v\, dx + \int_\Omega \frac{\partial u}{\partial \nu} v\, d\gamma, \tag{2.16}$$

where $\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$ is the Laplacian.

If both u and v are in $H^2(\Omega)$, we may interchange the roles of u and v in (2.16). Subtracting one formula from the other, we get

$$\int_\Omega (u\Delta v - \Delta u v) dx = \int_\Gamma \left(u \frac{\partial v}{\partial \nu} - \frac{\partial u}{\partial \nu} v \right) d\gamma, \tag{2.17}$$

13 for $u, v \in H^2(\Omega)$.

Finally replacing u by Δu in (2.17) we get, for $u \in H^4(\Omega)$, $v \in H^2(\Omega)$,

$$\int_\Omega \Delta u \Delta v\, dx = \int_\Omega \Delta^2 u\, v\, dx + \int_\Gamma \Delta u \frac{\partial v}{\partial \nu} d\gamma - \int_\Gamma \frac{\partial(\Delta u)}{\partial \nu} v d\gamma. \tag{2.18}$$

The formulae (2.15) through (2.18) are all known as *Green's formulae* in Sobolev spaces.

We derive two results from these formulae. These results will be useful later.

Lemma 2.1. *For all $v \in H_0^2(\Omega)$,*

$$|\Delta v|_{0,\Omega} = |v|_{2,\Omega}. \tag{2.19}$$

Consequently over $H_0^2(\Omega)$, the mapping $v \mapsto |\Delta v|_{0,\Omega}$ is a norm equivalent to the norm $\|\cdot\|_{2,\Omega}$.

Proof. Since $\mathscr{D}(\Omega)$ is dense in $H_0^2(\Omega)$, it suffices to prove (2.19) for $v \in (\Omega)$. Let $v \in \mathscr{D}(\Omega)$. Then

$$|\Delta v|^2 = \sum_{i=1}^n \left(\frac{\partial^2 v}{\partial x_i^2}\right)^2 + 2 \sum_{1 \le i < j \le n} \frac{\partial^2 v}{\partial x_i^2} \frac{\partial^2 v}{\partial x_j^2}.$$

By Green's formula (2.15),

$$\int_\Omega \frac{\partial^2 v}{\partial x_i^2} \frac{\partial^2 v}{\partial x_j^2} dx = -\int_\Omega \frac{\partial v}{\partial x_i} \frac{\partial^3 v}{\partial x_i \partial x_j^2} dx = \int_\Omega \left(\frac{\partial^2 v}{\partial x_i \partial x_j}\right)^2 dx \tag{2.20}$$

for, the integrals over Γ vanish for $v \in \mathscr{D}(\Omega)$. (cf. (2.11)). Now (2.19) follows directly from (2.20). This proves the lemma. □

Lemma 2.2. *Let $\Omega \subset \mathbb{R}^2$. Then for $u \in H^3(\Omega)$, $v \in H^2(\Omega)$,* **14**

$$\begin{aligned}&\int_\Omega 2\frac{\partial^2 u}{\partial x_1 \partial x_2} \frac{\partial^2 v}{\partial x_1 \partial x_2} - \frac{\partial^2 u}{\partial x_1^2} \frac{\partial^2 v}{\partial x_2^2} - \frac{\partial^2 u}{\partial x_2^2} \frac{\partial^2 v}{\partial x_1^2} dx \\ &= \int_\Gamma \left(-\frac{\partial^2 u}{\partial \tau^2} \frac{\partial v}{\partial \nu} + \frac{\partial^2 u}{\partial \tau \partial \nu} \frac{\partial v}{\partial \tau}\right) d\gamma,\end{aligned} \tag{2.21}$$

where $\dfrac{\partial}{\partial \tau}$ denotes the tangential derivative.

Proof. Let $\overrightarrow{\nu} = (\nu_1, \nu_2)$, $\overrightarrow{\tau} = (\tau_1, \tau_2)$ be the unit vectors along the outer normal and the tangent respectively. Without loss in generality we may assume $\tau_1 = -\nu_2$, $\tau_2 = \nu_1$. Also note that $\nu_1^2 + \nu_2^2 = 1$. The second derivatives occurring in the right hand side are defined by

$$\begin{cases} \dfrac{\partial^2 u}{\partial \tau^2} = \dfrac{\partial^2 u}{\partial x_1^2}\tau_1^2 + 2\dfrac{\partial^2 u}{\partial x_1 \partial x_2}\tau_1\tau_2 + \dfrac{\partial^2 u}{\partial x_2^2}\tau_2^2 \\ \dfrac{\partial^2 u}{\partial \tau \partial \nu} = \dfrac{\partial^2 u}{\partial x_1^2}\nu_1\tau_1 + \dfrac{\partial^2 u}{\partial x_1 \partial x_2}(\nu_1\tau_2 + \nu_2\tau_1) + \dfrac{\partial^2 u}{\partial x_2^2}\nu_2\tau_2. \end{cases} \tag{2.22}$$

Using all these relations we get

$$\begin{aligned} -\frac{\partial^2 u}{\partial \tau^2}\frac{\partial v}{\partial \nu} + \frac{\partial^2 u}{\partial \tau \partial \nu}\frac{\partial v}{\partial \tau} &= \left(\frac{\partial^2 u}{\partial x_1 \partial x_2}\frac{\partial v}{\partial x_2} - \frac{\partial^2 u}{\partial x_2^2}\frac{\partial v}{\partial x_1}\right)\nu_1 \\ &\quad + \left(\frac{\partial^2 u}{\partial x_1 \partial x_2}\frac{\partial v}{\partial x_1} - \frac{\partial^2 u}{\partial x_1^2}\frac{\partial v}{\partial x_2}\right)\nu_2 \\ &= \overrightarrow{X} \cdot \overrightarrow{\nu}, \end{aligned} \tag{2.23}$$

where, $\overrightarrow{X} = (X_1, X_2)$ and

$$\begin{cases} X_1 = \dfrac{\partial^2 u}{\partial x_1 \partial x_2}\dfrac{\partial v}{\partial x_2} - \dfrac{\partial^2 u}{\partial x_2^2}\dfrac{\partial v}{\partial x_1} \\ X_2 = \dfrac{\partial^2 u}{\partial x_1 \partial x_2}\dfrac{\partial v}{\partial x_1} - \dfrac{\partial^2 u}{\partial x_1^2}\dfrac{\partial v}{\partial x_2} \end{cases} \tag{2.24}$$

Also note that,

$$\operatorname{div} X = \frac{\partial X_1}{\partial x_1} + \frac{\partial X_2}{\partial x_2} = 2\frac{\partial^2 u}{\partial x_1 \partial x_2}\frac{\partial^2 v}{\partial x_1 \partial x_2} - \frac{\partial^2 u}{\partial x_1^2}\frac{\partial^2 v}{\partial x_2^2} - \frac{\partial^2 u}{\partial x_2^2}\frac{\partial^2 v}{\partial x_1^2}. \tag{2.25}$$

15

Now by Green's formula (2.15) applied to functions $v_i \in H^1(\Omega)$ and to the constant function 1,

$$\int_\Omega \frac{\partial v_i}{\partial x_i} dx = \int_\Gamma v_i \nu_i d\gamma.$$

Hence summing over all i, if $\overrightarrow{v} = (v_1, \ldots, v_n)$,

$$\int_\Gamma \operatorname{div} \overrightarrow{v}\, dx = \int_\Gamma \overrightarrow{v} \cdot \overrightarrow{v}\, d\gamma. \tag{2.26}$$

(This is known as the *Gauss' Divergence Theorem* and also as the *Ostrogradsky's formula*). From (2.23), (2.25) and (2.26) the result follows. □

With this background, we proceed to examples of the abstract problem of Sec. 1.

Example 2.1. Let $K = V = H_0^1(\Omega)$. Let $a \in L^\infty(\Omega)$ such that $a \geq 0$ a.e. in Ω. Let $f \in L^2(\Omega)$. Define the bilinear form $a(\cdot,\cdot)$ and the functional $f(\cdot)$, by

$$\begin{cases} a(u,v) = \int_\Omega \left(\sum_{i=1}^n \dfrac{\partial u}{\partial x_i} \dfrac{\partial v}{\partial x_i} + auv \right) dx, \\ f(v) = \int_\Omega fv\, dx. \end{cases} \tag{2.27}$$

The continuity of $a(\cdot,\cdot)$ and $f(\cdot)$ follows from the Cauchy-Schwarz inequality. For instance

$$|f(v)| \leq |f|_{0,\Omega} |v|_{0,\Omega} \leq |f|_{0,\Omega} \|v\|_{1,\Omega}. \tag{2.28}$$

We now show that $a(\cdot,\cdot)$ is V-elliptic.

$$\begin{aligned} a(v,v) &= \int_\Omega \left(\sum_{i=1}^n \left(\frac{\partial v}{\partial x_i} \right)^2 + av^2 \right) dx \\ &\geq \int_\Omega \sum_{i=1}^n \left(\frac{\partial v}{\partial x_i} \right)^2 dx \quad \text{(since } a \geq 0 \text{ a.e. in } \Omega) \\ &= |v|_{1,\Omega}^2. \end{aligned}$$

16

Since $|\cdot|_{1,\Omega}$ is equivalent to $\|\cdot\|_{1,\Omega}$ over V, this proves the V-ellipticity. Hence by our results in Sec. 1 there exists a unique function $u \in V$ such that $a(u,v) = f(v)$ for all $v \in V$.

Interpretation of this problem: Using the above equation satisfied by u, we get

$$\int_\Omega \left(\sum_{i=1}^{n} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} + auv \right) dx = \int_\Omega fv\, dx, \quad \text{for all} \quad v \in H_0^1(\Omega). \tag{2.29}$$

From the inclusion $\mathscr{D}(\Omega) \subset H_0^1(\Omega)$, we get that u satisfies the equation $-\Delta u + au = f$ *in the sense of distributions*.

If we assume that u is "sufficiently smooth", for example $u \in H^2(\Omega)$, then we may apply Green's formula (2.16), which gives

$$\int_\Omega auv\, dx - \int_\Omega \Delta uv\, dx + \int_\Gamma \frac{\partial u}{\partial \nu} v\, d\gamma = \int_\Omega fv\, dx. \tag{2.30}$$

Since $v \in H_0^1(\Omega)$, $\operatorname{tr}_\Gamma v = 0$. Hence the integral over Γ vanishes. Thus we get

$$\int_\Omega (-\Delta u + au - f)v\, dx = 0 \quad \text{for all} \quad v \in H_0^1(\Omega). \tag{2.31}$$

Varying v over $H_0^1(\Omega)$, we get that u satisfies the equation $-\Delta u + au = f$ in Ω. Further since $u \in H_0^1(\Omega)$, we get the boundary condition $u = 0$ on Γ. Thus we may interpret u as the solution of the "classical" boundary value problem:

$$\begin{cases} -\Delta u + au = f \quad \text{in} \quad \Omega, \\ u = 0 \ \text{ on } \ \Gamma. \end{cases} \tag{2.32}$$

17

This is known as the *homogeneous Dirichlet problem* for the operator $-\Delta u + au$.

A particular case of this equation arises in the theory of Elasticity, for which $\Omega \subset \mathbb{R}^2$ and $a = 0$. Thus $-\Delta u = f$ in Ω and $u = 0$ on Γ. This corresponds to the *membrane problem*:

Consider an elastic membrane stretched over Ω and kept fixed along Γ. Let $F dx$ be the density of force acting on an element dx of Ω. Let $u(x)$ be the vertical displacement of the point $x \in \Omega \subset \mathbb{R}^2$, measured in the x_3-direction from the (x_1, x_2)-plane. If t is the 'tension' of the membrane, then u is the solution of the problem

$$\begin{cases} -\Delta u = f \text{ in } \Omega \\ u = 0 \text{ on } \Gamma \end{cases} \tag{2.33}$$

where $f = F/t$.

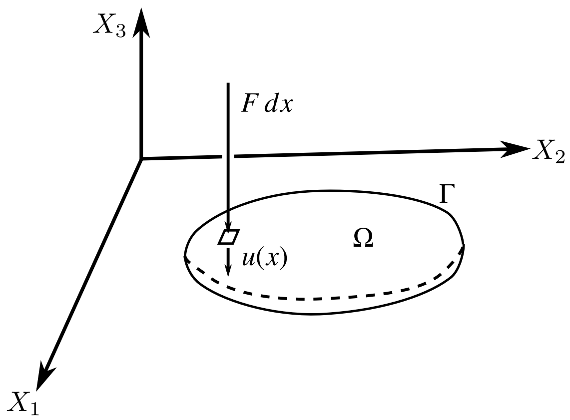

Figure 2.1:

Remark 2.1. To solve the problem (2.32) by the classical approach, one needs hard analysis involving Schauder's estimates. By the above procedure viz. the variational method, we have got through with it more easily.

The above problem is a typical example of a second-order problem. **18**

Exercise 2.1. *The obstacle problem*. Let Ω, Γ be as in example (2.1). Let $\mathcal{X}$ be an "obstacle" in this region. Let $\mathcal{X} \leq 0$ on Γ. Let $F\,dx$ be the density of the force acting on a membrane stretched over Ω, fixed along Γ. The displacement $u(x)$ at x in the vertical direction is the solution of the following problem:

$$\text{If} \quad V = H_0^1(\Omega),\ K = \left\{ v \in H_0^1(\Omega);\ v \geq \mathcal{X} \text{ a.e. in } \Omega \right\},$$

$$\text{let} \quad a(u,v) = \int_\Omega \sum_{i=1}^n \frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_i} dx, \;\; f(v) = \int_\Omega fv\,dx, \;\; f \in L^2(\Omega),$$
$$\mathcal{X} \in H^2(\Omega), \;\; \mathcal{X} \le 0 \;\text{ on }\; \Gamma.$$

Show that K is a closed convex set and hence that this problem admits of a unique solution. Assuming the regularity result $u \in H^2(\Omega) \cap H_0^1(\Omega)$, show that this problem solves the classical problem,

$$\begin{cases} u \ge \mathcal{X} \;\text{ in }\; \Omega, \\ -\Delta u = f \quad \text{when} \quad u > \mathcal{X}, \quad (f = F/t) \\ u = 0 \cdot \Gamma \end{cases}$$

(We will discuss the Obstacle Problem in Sec. 9).

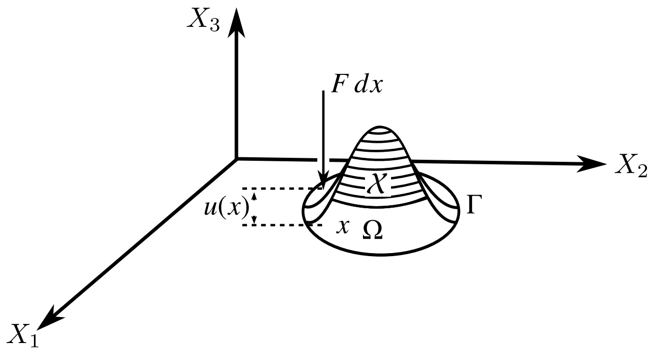

Figure 2.2:

19 **Exercise 2.2.** Let $V = H^1(\Omega)$. Define $a(\cdot,\cdot)$, $f(\cdot)$ as in example 2.1. Assume further that there exists a constant a_0 such that $a \ge a_0 > 0$ in Ω. If u_0 is a given function in $H^1(\Omega)$, define

$$\begin{aligned} K &= \left\{ v \in H^1(\Omega);\; v - u_0 \in H_0^1(\Omega) \right\} \\ &= \left\{ v \in H^1(\Omega);\; \mathrm{tr}_\Gamma\, v = \mathrm{tr}_\Gamma\, u_0 \right\}. \end{aligned}$$

Check that K is a closed convex subset. Interpret the solution to be that of the *Non-homogeneous Dirichlet problem*,

$$\begin{cases} -\Delta u + au = f & \text{in} \quad \Omega. \\ u = u_0 & \text{on} \quad \Gamma. \end{cases}$$

Example 2.2. Let $K = V = H^1(\Omega)$. Let $a \in L^\infty(\Omega)$ such that $a \geq a_0 > 0$, $f \in L^2(\Omega)$. Define $a(\cdot,\cdot)$ and $f(\cdot)$ as in example (2.1). The continuity of $a(\cdot,\cdot)$ and $f(\cdot)$ follow as usual. For the V-ellipticity, we can no longer prove it with the semi-norm $|\cdot|_{1,\Omega}$ as we did earlier. It is here we use the additional assumption on a, since

$$\begin{aligned} a(v,v) &= \int_\Omega \left(\sum_{i=1}^n \left(\frac{\partial v}{\partial x_i} \right)^2 + av^2 \right) dx \\ &\geq \min(1, a_0) \|v\|^2_{1,\Omega}. \end{aligned}$$

Thus we have a unique solution u to the abstract problem satisfying $a(u,v) = f(v)$. If we assume again that u is "sufficiently smooth" to apply the Green's formula (2.16), we get

$$\int_\Omega (-\Delta u + au) v \, dx + \int_\Gamma \frac{\partial u}{\partial \nu} v \, d\gamma = \int_\Gamma f v \, dx. \tag{2.34}$$

If $v \in \mathscr{D}(\Omega)$, then the integral over Γ will vanish. Thus u satisfies the equation $-\Delta u + au = f$ as in Example 2.1[1]. However we now get a **20**
different boundary condition. In example (2.1) the boundary condition was built in with the assumption $u \in V = H^1_0(\Omega)$. Now from (2.34), we may write:

$$\int_\Omega (-\Delta u + au - f) v \, dx = -\int_\Gamma \frac{\partial u}{\partial \nu} v \, d\gamma \tag{2.35}$$

for all $v \in H^1(\Omega)$.

But the left hand side of (2.35) is zero since u satisfies the differential equation as above so that for all $v \in H^1(\Omega)$, $\int_\Gamma \frac{\partial u}{\partial \nu} v \, d\gamma = 0$. Thus

[1]As in Example 2.1, the equation $-\Delta u + au = f$ is always satisfied in the sense of distributions since $\mathscr{D}(\Omega) \subset H^1(\Omega)$.

$\dfrac{\partial u}{\partial \nu} = 0$ on Γ, and we may interpret this problem as the classical problem:

$$\begin{cases} -\Delta u + au = f \text{ in } \Omega \\ \frac{\partial u}{\partial \nu} = 0 \text{ on } \Gamma. \end{cases} \tag{2.36}$$

This is a *homogeneous Neumann problem.*

Exercise 2.3. With K, V, $a(\cdot,\cdot)$ as in example 2.2., define

$$f(v) = \int_\Omega fv\,dx + \int_\Gamma gv\,d\gamma, \begin{cases} f \in L^2(\Omega), \\ g \in L^2(\Omega). \end{cases}$$

Show that the abstract problem leads to a solution of the *non-homogeneous Neumann problem*

$$\begin{cases} -\Delta u + au = f \text{ in } \Omega \\ \dfrac{\partial u}{\partial \nu} = g \text{ on } \Gamma. \end{cases}$$

21 **Remark 2.2.** In these examples one may use the more general bilinear from defined by

$$a(u,v) = \int_\Omega \left(\sum_{i,j=1}^n a_{ij} \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_i} + auv \right) dx, \tag{2.37}$$

where the functions $a_{ij} \in L^\infty(\Omega)$ satisfy the condition that for some $\nu > 0$,

$$\sum_{i,j=1}^n a_{ij}\xi_i\xi_j \geq \nu \sum_{i=1}^n \xi_i^2 \tag{2.38}$$

for all $\xi \in \mathbb{R}^n$ and a.e. in Ω. This is the classical ellipticity condition for second order partial differential operators. One should check (exercise!) in this case that the abstract problem leads to a solution of the boundary value problem

$$-\sum_{i,j=1}^n \frac{\partial}{\partial x_i}\left(a_{ij} \frac{\partial u}{\partial x_j} \right) + au = f \text{ in } \Omega \tag{2.39}$$

with the boundary condition

$$
(2.40) \qquad \begin{cases} u = 0 \ \text{ on } \ \Gamma \ \text{ if } \ K = V = H_0^1(\Omega), \\ \sum\limits_{i,j=1}^{n} a_{ij} \frac{\partial u}{\partial x_j} \nu_i = 0 \ \text{ on } \ \Gamma \ \text{ if } \ K = V = H^1(\Omega). \end{cases}
$$

The latter boundary operator in (2.40) is called the *conormal derivative* associated with the partial differential operator,

$$
-\sum_{i,j=1}^{n} \frac{\partial}{\partial x_i} \left(a_{ij} \frac{\partial}{\partial x_j} \right).
$$

Notice that the term (+ au) contributes nothing.

Example 2.3. *System of Elasticity.* Let $\Omega \subset \mathbb{R}^3$, with Lipschitz continuous boundary Γ. Further assume that Γ can be partitioned into two portions Γ_0 and Γ_1 such that the $d\gamma$-measure of Γ_0 is > 0. Let

$$
K = V = \left\{ \overrightarrow{v} = (v_1, v_2, v_3); \ v_i \in H^1(\Omega), \ 1 \leq i \leq 3 \text{ and } \overrightarrow{v} = \overrightarrow{0} \text{ on } \Gamma_0 \right\}.
$$

Define 22

$$
(2.41) \qquad \begin{cases} \epsilon_{ij}(\overrightarrow{v}) = \dfrac{1}{2} \left(\dfrac{\partial v_i}{\partial x_j} + \dfrac{\partial v_j}{\partial x_i} \right), \\ \sigma_{ij}(\overrightarrow{v}) = \lambda \left(\sum_{k=1}^{3} \epsilon_{kk}(\overrightarrow{v}) \right) \delta_{ij} + 2\mu \epsilon_{ij}(\overrightarrow{v}), \end{cases}
$$

for $1 \leq i, \ j \leq 3$. The latter relation is usually known as *Hooke's law*. The constants $\lambda(\geq 0)$ and $\mu(> 0)$ are known as *Lame's coefficients*. We define the bilinear form $a(\cdot, \cdot)$ by,

$$
(2.42) \qquad \begin{aligned} a(\overrightarrow{u}, \overrightarrow{v}) &= \int_\Omega \sum_{i,j=1}^{3} \sigma_{ij}(\overrightarrow{u} \epsilon_{ij}(\overrightarrow{v}) dx \\ &= \int_\Omega (\lambda \operatorname{div} \overrightarrow{v} + 2\mu \sum_{i,j=1}^{3} \epsilon_{ij}(\overrightarrow{u}) \epsilon_{ij}(\overrightarrow{v}) dx. \end{aligned}
$$

Let $\overrightarrow{f} = (f_1, f_2, f_3), \ f_i \in L^2(\Omega)$, and $\overrightarrow{g} = (g_1, g_2, g_3), \ g_i \in L^2(\Gamma)$, be given.

Define the linear functional $f(\cdot)$, by,

$$f(\vec{v}) = \int_{\Omega} \vec{f} \cdot \vec{v}\, dx + \int_{\Gamma_1} \vec{g} \cdot \vec{v}\, d\gamma. \tag{2.43}$$

The continuity of $a(\cdot,\cdot)$ and $f(\cdot)$ follow from the Cauchy-Schwarz inequality. For the V-ellipticity of $a(\cdot,\cdot)$ one uses the inequality

$$a(\vec{v},\vec{v}) \geq 2\mu \int_{\Omega} \sum_{i,j=1}^{3} (\epsilon_{ij}(\vec{v}))^2 dx$$

and the fact that the square root of the integral appearing in the right hand side of the above inequality is a norm over the space V, equivalent to the norm $\vec{v} = (v_1, v_2, v_3) \mapsto (\sum_{i=1}^{3} \|v_i\|_{1,\Omega}^2)^{\frac{1}{2}}$. This is a nontrivial fact which uses essentially the fact that Γ_0 has measure > 0 and an inequality known as *Körn's inequality*. We omit the proof here.

Again the problem $a(\vec{u},\vec{v}) = f(\vec{v})$ admits a unique solution. As-
23 suming sufficient smoothness, we may apply Green's formula:

$$\begin{aligned}
\int_{\Omega} \sum_{i,j=1}^{3} \sigma_{ij}(\vec{u})\epsilon_{ij}(\vec{v})dx &= \frac{1}{2}\int_{\Omega} \sum_{i,j=1}^{3} \sigma_{ij}(\vec{u})\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right)dx \\
&= \int_{\Omega} \sum_{i,j=1}^{3} \sigma_{ij}(\vec{u})\frac{\partial v_i}{\partial x_j}dx
\end{aligned}$$

(since $\epsilon_{ij}(v)$ is symmetric in i and j)

$$= -\int_{\Omega} \sum_{i,j=1}^{3} \frac{\partial}{\partial x_j}(\sigma_{ij}(\vec{u}))v_i dx + \int_{\Gamma_1} \sum_{i,j=1}^{3} \sigma_{ij} v_i \nu_j d\gamma.$$

Thus, the abstract problem leads to a solution of

$$\begin{cases}
-\sum\limits_{j=1}^{3} \frac{\partial}{\partial x_j}(\sigma_{ij}(\vec{u})) = f_i (1 \leq i \leq 3) \ \text{ in } \ \Omega \\
\qquad \vec{u} = \vec{0} \ \text{ on } \ \Gamma_0 \ \text{ and} \\
\sum\limits_{j=1}^{3} \sigma_{ij}(\vec{u})\nu_j = g_i \ \text{ on } \ \Gamma_1 (1 \leq i \leq 3).
\end{cases} \tag{2.44}$$

Note also that

$$
\begin{aligned}
-\sum_{j=1}^{3} \frac{\partial}{\partial x_j}(\sigma_{ij}(\vec{u})) &= -\sum_{j=1}^{3} \frac{\partial}{\partial x_j}\left(\lambda \sum_{k=1}^{3} \epsilon_{kk}(\vec{u})\delta_{ij} + 2\mu\epsilon_{ij}(\vec{u})\right)\\
&= -\sum_{j=1}^{3} \frac{\partial}{\partial x_j}\left(\lambda \sum_{k=1}^{3} \frac{\partial u_k}{\partial x_k}\right)\delta_{ij} - \sum_{j=1}^{3} \frac{\partial}{\partial x_j}\left(2\mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)\right)\\
&= -(\lambda+\mu)(\text{grad}\ \ \text{div}\,\vec{v})_i - \mu\Delta u_i.
\end{aligned}
$$

Thus the first equation of (2.44) is equivalent to

(2.45) $$-\mu\Delta\vec{u} - (\lambda+\mu)\,\text{grad}\,\text{div}\,\vec{u} = \vec{f} \quad \text{in} \ \ \Omega.$$

The equations (2.44) constitute the *system of linear Elasticity.*

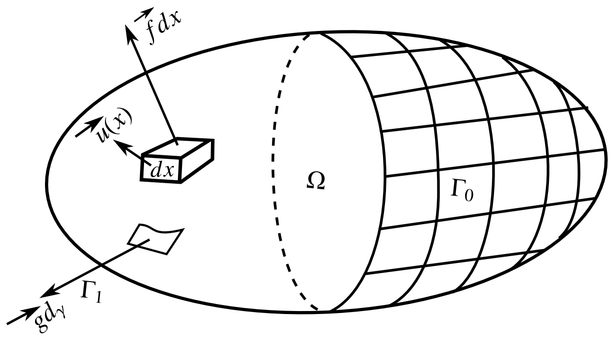

Figure 2.3:

24

If we have an elastic three-dimensional body fixed along Γ_0, acted on by an exterior force of density $f\,dx$ and force of density $g\,d\gamma$ along Γ_1 and if σ_{ij} is the stress tensor, the displacement u satisfies (2.44); cf. Fig. 2.3.

The relation $a(\vec{u}, \vec{v}) = f(\vec{v})$, viz.,

(2.46) $$\int_\Omega \sum_{i,j=1}^{3} \sigma_{ij}(\vec{u})\epsilon_{ij}(\vec{v})dx = \int_\Omega \vec{f}\cdot\vec{v}\,dx + \int_{\Gamma_1} \vec{g}\cdot\vec{v}\,d\gamma$$

for all $\overrightarrow{v} \in V$ is known as the *principle of virtual work*. The tensor ϵ_{ij} is *the strain tensor* and the tensor σ_{ij} *the stress tensor*. The expression $\frac{1}{2}a(\overrightarrow{v},\overrightarrow{v})$ is the *strain energy*, and the functional $f(\overrightarrow{v})$ is the *potential energy of exterior forces.* This example is of fundamental importance in that the finite element method has been essentially developed for solving this particular problem or some of its special cases (membranes, plates, shells, etc.,) and generalizations (nonlinear elasticity, etc...).

Remark 2.3. The above problems are all examples of *linear problems*[2]: **25**
The map from the right-hand side of the equation and of the boundary conditions to the solution u is linear. The non-linearity may occur in three ways:

(i) When K is not a subspace of V. (e.g. Exercises 2.1 and 2.4);

(ii) If in Example 2.3 we have, instead of the first equality in (2.41):

$$\epsilon_{ij}(\overrightarrow{v}) = \frac{1}{2}\left[\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right) + \sum_{k=1}^{3} \frac{\partial v_k}{\partial x_i}\frac{\partial v_k}{\partial x_j}\right].$$

This is the case for instance when one derives the so-called Von Karmann's equations of a clamped plate;

(iii) We may replace Hooke's law (the second relations in (2.41)) by non-linear equations connecting ϵ_{ij} and σ_{ij}, which are known as *non-linear constitutive equations*. (e.g. *Hencky's law*).

Exercise 2.4. Let $V = H^1(\Omega)$, and $a(\cdot,\cdot)$ and $f(\cdot)$ be as in example 2.2, and let

$$K = \left\{v \in H^1(\Omega);\ v \geq 0 \ \text{ a.e. on } \ \Gamma\right\}.$$

Show that K is a closed convex cone with vertex 0. Using the results of Sec. 1 show that the interpretation is

$$-\Delta u + au = f \ \text{ in } \ \Omega,$$
$$u \geq 0,\ \frac{\partial u}{\partial \nu} \geq 0,\ u\frac{\partial u}{\partial \nu} = 0 \ \text{ on } \ \Gamma.$$

(This is called the *SIGNORINI problem*).

We now examine fourth-order problems.

Example 2.4. Let $K = V = H_0^2(\Omega)$. Define

$$\text{(2.47)} \qquad \begin{cases} a(u,v) = \int_\Omega \Delta u \Delta v\, dx, \\ f(v) = \int_\Omega fv\, dx,\ \ f \in L^2(\Omega), \end{cases}$$

[2] Except in Exercise 2.1.

26 for all $u, v \in V$. The continuity follows as usual. For the V-ellipticity of $a(\cdot,\cdot)$ we have,

$$a(v,v) = \int_\Omega (\Delta v)^2 dx = |\Delta v|^2_{0,\Omega} = |v|^2_{2,\Omega}. \tag{2.48}$$

(by Lemma 2.1. Since $|\cdot|_{2,\Omega}$ and $\|\cdot\|_{2,\Omega}$ are equivalent on $H^2_0(\Omega)$, the V-ellipticity follows from (2.48).

Hence there exists a unique function $u \in H^2_0(\Omega)$ such that

$$\int_\Omega \Delta u \Delta v\, dx = \int_\Omega fv\, dx \quad \text{for all} \quad v \in H^2_0(\Omega).$$

Assuming u to be sufficiently smooth (say, $u \in H^4(\Omega)$), then by Green's formula (2.18),

$$\int_\Omega (\Delta^2 u - f)v\, dx = \int_\Gamma \frac{\partial(\Delta u)}{\partial \nu} v\, d\gamma - \int_\Gamma \Delta u \frac{\partial v}{\partial \nu} d\gamma, \tag{2.50}$$

for all $v \in H^2_0(\Omega)$. Hence by varying v over $H^2_0(\Omega)$, we get that u satisfies $\Delta^2 u = f$ in Ω. Since $u \in H^2_0(\Omega)$, the boundary conditions are given by (2.14). Thus we interpret this problem as the classical problem

$$\begin{cases} \Delta^2 u = f \quad \text{in} \quad \Omega, \\ u = \dfrac{\partial u}{\partial \nu} = 0 \quad \text{on} \quad \Gamma. \end{cases} \tag{2.51}$$

This is the *homogeneous Dirichlet problem* for the operator Δ^2.

When $n = 2$, this is an important problem in Hydrodynamics. Here u is known as the stream function and $-\Delta u$ is the *vorticity*.

27 A slight modification of $a(\cdot,\cdot)$ leads to an important problem in Elasticity. Again let $n = 2$. Let $f \in L^2(\Omega)$ and if $K = V = H^2_0(\Omega)$, define

$$\begin{cases} f(v) = \int_\Omega fv\, dx, \\ a(u,v) = \int_\Omega \left[\Delta u\, \Delta v + (1-\sigma)\left(2\dfrac{\partial^2 u}{\partial x_1 \partial x_2}\dfrac{\partial^2 v}{\partial x_1 \partial x_2} - \dfrac{\partial^2 u}{\partial x_1^2}\dfrac{\partial^2 v}{\partial x_2^2} - \dfrac{\partial^2 u}{\partial x_2^2}\dfrac{\partial^2 v}{\partial x_1^2}\right)\right] dx. \end{cases} \tag{2.52}$$

The integrand occurring in the definition of $a(u,v)$ may also be written as

$$\sigma \Delta u \Delta v + (1-\sigma)\left(\frac{\partial^2 u}{\partial x_1^2}\frac{\partial^2 v}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2}\frac{\partial^2 v}{\partial x_2^2} + \frac{2\partial^2 u}{\partial x_1 \partial x_2}\frac{\partial^2 v}{\partial x_1 \partial x_2}\right). \tag{2.53}$$

Usually, from physical considerations, $0 < \sigma < \frac{1}{2}$.

Note that

$$a(v, v) = \sigma|\Delta v|^2_{0,\Omega} + (1 - \sigma)|v|^2_{2,\Omega} \tag{2.54}$$

by (2.53) and this leads to the V-ellipticity of $a(\cdot, \cdot)$. By virtue of (2.21) in Lemma 2.2, we get that the relations $a(u, v) = f(v)$ read as

$$\int_\Omega \Delta^2 uv\, dx = \int_\Omega fv\, dx, \tag{2.55}$$

assuming sufficient smoothness of u. Thus again we get the same equation as in (2.51). Notice that the additional term in the definition of $a(\cdot, \cdot)$ has contributed nothing towards the differential equation.

This latter problem is known as the *clamped plate problem*:

Consider a plate of "small" thickness e lying on the x_1x_2-plane. Let E be its Young's modulus and σ its Poisson coefficient. Let there be a load F acting on the plate. The displacement u is the solution of (2.51), where f is given by (cf. Fig. 2.4):

$$F = \frac{Ee^3 f}{12(1 - \sigma^2)}. \tag{2.56}$$

28

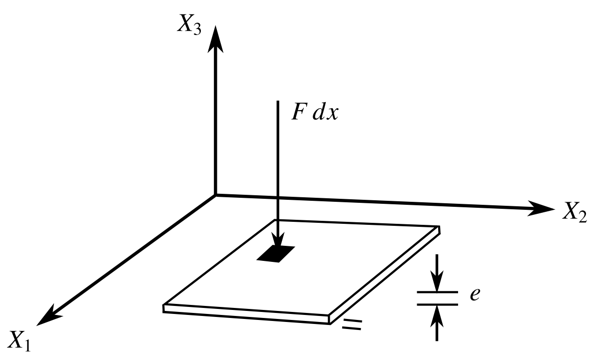

Figure 2.4:

We will return to this problem in Sections 10 and 11.

Exercise 2.5. Let $K = V = \{v \in H^2(\Omega);\ v = 0 \text{ on } \Gamma\} = H^2(\Omega) \cap H_0^1(\Omega)$, and define $a(\cdot,\cdot)$ and $f(\cdot)$ as in the case of the clamped plate. Assuming the V-ellipticity of $a(\cdot,\cdot)$ show that the solution of the abstract problem satisfies $\Delta^2 u = f$ in Ω and $u = 0$ on Γ. What is the other boundary condition? This is known as the problem of the *simply supported plate.*

Exercise 2.6. Let $K = V = H^2(\Omega) \cap H_0^1(\Omega)$, and

$$
\begin{aligned}
a(u, v) &= \int_\Omega \Delta u \Delta v \, dx; \\
f(v) &= \int_\Omega f v \, dx - \int_\Gamma \lambda \frac{\partial v}{\partial \nu} d\gamma, \quad \text{where } f \in L^2(\Omega),\ \lambda \in L^2(\Gamma).
\end{aligned}
$$

Show that we may apply the result of Sec. 1 and give an interpretation of this problem.

REFERENCES. For details on Sobolev spaces, see Nečas [20] and Lions and Magenes [17]. For the theory of Elasticity, one may refer to Duvaut and Lions [10] and Landau and Lipschitz [14].

Chapter 3

The Finite Element Method in its Simplest Form

MAINTAINING OUR ASSUMPTIONS as in the Lax-Milgram Lemma **29**
(Theorem 1.2 we concentrate our attention on the following problem (P):

(P): To find $u \in V$ such that $a(u, v) = f(v)$ for all $v \in V$.

Let V_h be a finite-dimensional subspace of V. Then we may state the following problem:

(P_h): To find $u_h \in V_h$ such that $a(u_h, v_h) = f(v_h)$ for all $v_h \in V_h$.

V_h, being a finite-dimensional subspace, is a Hilbert space for the norm of V. Hence by Theorem 1.2, u_h exists and is unique. We try to approximate the solution u of (P) by means of solutions u_h of the problem (P_h) for various subspaces V_h. This is known as the *internal approximation method.*

As a first step in this direction, we prove a most fundamental result:

Theorem 3.1. *There exists a constant C, which is independent of V_h, such that*

$$\|u - u_h\| \leq C \inf_{v_h \in V_h} \|u - v_h\|. \tag{3.1}$$

Proof. If $w_h \in V_h$, then

$$a(u, w_h) = f(w_h) = a(u_h, w_h).$$

Thus for all $w_h \in V_h$

$$a(u - u_h, w_h) = 0. \tag{3.2}$$

Using this and the V-ellipticity of $a(\cdot, \cdot)$, we get, for all $v_h \in V_h$,

$$\begin{aligned}
\alpha||u - u_h||^2 &\leq a(u - u_h, u - u_h) \\
&= a(u - u_h, u - v_h) + a(u - u_h, v_h - u_h) \\
&= a(u - u_h, u - v_h) \text{ by } (3.2) \\
&\leq M||u - u_h||\,||u - v_h||.
\end{aligned}$$

30 Hence, $||u - u_h|| \leq M/\alpha\, ||u - v_h||$ for all $v_h \in V_h$. Setting $C = M/\alpha$ and taking infimum on the right-hand side the result follows. □

The above result estimates the 'error' in the solution of (P) when instead we solve (P_h). To get an upper bound for the error, we only need to compute $\inf\limits_{v_h \in V_h} ||u - v_h||$ which is the distance of u from the subspace V_h. This is a problem in *approximation theory*.

Remark 3.1. If $a(\cdot, \cdot)$ is also symmetric then we observe the following:

(i) $J(u_h) = \inf\limits_{v_h \in V_h} J(v_h)$ by Corollary 1.1 (b)

(ii) We saw that $a(u - u_h, w_h) = 0$ for all $w_h \in V_h$. Since $a(\cdot, \cdot)$ is now an inner product, we get that u_h is the projection of u to the closed subspace V_h in the sense of this inner-product. Therefore,

$$\sqrt{\alpha}||u - u_h|| \leq \sqrt{a(u - u_h, u - u_h)} \leq \sqrt{a(u - v_h, u - v_h)} \leq \sqrt{M}||u - v_h||$$

for all $v_h \in V_h$. Hence the constant C in theorem 3.1 can be taken to be here $\sqrt{M/\alpha} \leq M/\alpha$, since the continuity and V-ellipticity imply jointly that $M \geq \alpha$.

We may now describe the *finite element method* (f.e.m.) in its simplest terms. The method consists in making *special choices for the subspaces* V_h such that the solutions u_h of the problems (P_h) converge to u.

We will outline the procedure for obtaining the spaces V_h by considering, for example, a second-order problem. **31**

Let $V = H_0^1(\Omega)$ or $H^1(\Omega)$. Let us assume Ω to be a *polygonal domain* in $\mathbb{R}^n$. That is, $\overline{\Omega}$ is a polygon in $\mathbb{R}^n$. We then have the following step-by-step procedure:

(i) We first establish a finite triangulation $\mathfrak{t}_h$ of the domain Ω such that $\overline{\Omega} = \bigcup\limits_{K \in \mathfrak{t}_h} K$. The sets K are called *finite elements*. If $n = 2$, they will be, in general, triangles. They will be tetrahedral in $n = 3$ and 'n-simplices' in any $\mathbb{R}^n$. These have the further property that any side of a finite element K is either a portion of the boundary or the side of an adjacent finite element. (See Fig. 3.1).

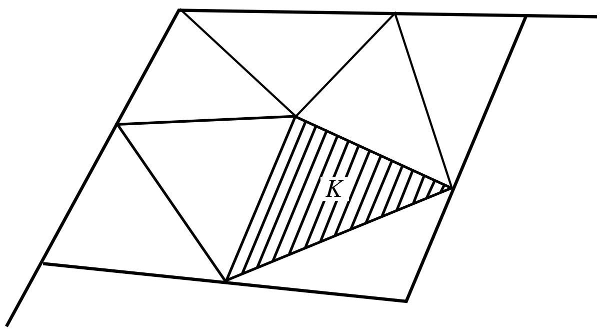

Figure 3.1:

(ii) The space V_h is such that for each $v_h \in V_h$, its restriction $v_h|K$ to each K belongs to some finite-dimensional space P_k of real valued functions over K which are preassigned. In practice we choose P_K to be a space of *polynomials*.

(iii) We then need inclusions such as $V_h \subset H_0^1(\Omega)$ or $H^1(\Omega)$. We establish a simple criterion to realise this.

32 **Theorem 3.2.** *If for every $K = \mathfrak{t}_h$, $P_K \subset H^1(K)$, and $V_h \subset C^0(\overline{\Omega})$, then $V_h \subset H^1(\Omega)$. If in addition $v = 0$ on Γ for all $v \in V_h$, then $V_h \subset H^1_0(\Omega)$.*

Proof. Let $v \in V_h$. Since $v|K \in L^2(K)$ for every $K \in \mathfrak{t}_h$, it follows that $v \in L^2(\Omega)$. Hence to complete the proof it only remains to show that for $1 \leq i \leq n$, there exist $v_i \in L^2(\Omega)$ such that for each $\varphi = \mathscr{D}(\Omega)$, we have,

$$\int_\Omega \varphi v_i dx = -\int_\Omega \frac{\partial \varphi}{\partial x_i} v\, dx. \quad (1 \leq i \leq n). \tag{3.3}$$

Then it will follow that $\dfrac{\partial v}{\partial x_i} = v_i$ and hence $v \in H^1(\Omega)$.

However, $v|K \in P_K \subset H^1(K)$ implies that $\dfrac{\partial(v|K)}{\partial x_i} \in L^2(K)$ for $1 \leq i \leq n$. Let $\varphi \in \mathscr{D}(\Omega)$. Since the boundary ∂K of any K of the triangulation is Lipschitz continuous, we apply the Green's formula (2.15) to get

$$\int_K \frac{\partial(v|K)}{\partial x_i} \varphi dx = -\int_K (v|K) \frac{\partial \varphi}{\partial x_i} dx + \int_{\partial K} (v|K) \varphi v_{i,K} d\gamma_K, \tag{3.4}$$

where $d\gamma_K$ is the measure on ∂K and $\vec{\nu}_K = (\nu_{1,K}, \ldots, \nu_{n,K})$ is the outer normal on ∂K. Summing over all the finite elements K, we get

$$\begin{aligned} \int_\Omega \varphi v_i dx &= \sum_{K \in \mathfrak{t}_h} \int_K \varphi \frac{\partial(v|K)}{\partial x_i} dx \\ &= -\int_\Omega \frac{\partial \varphi}{\partial x_i} v\, dx + \sum_{K \in \mathfrak{t}_h} \int_{\partial K} \varphi (v|K)^{\nu}_{i,K} d\gamma_K, \end{aligned} \tag{3.5}$$

where v_i is the function whose restriction to each K is $\dfrac{\partial(v|K)}{\partial x_i}$.

The summation on the right-hand side of the above equation is zero for the following reasons:

On the boundary Γ, since $\varphi \in \mathscr{D}(\Omega)$, the integral corresponding to $\partial K \cap \Gamma$ is zero. So the problem, if any, is only on the other portions of the boundary of each K. However, these always occur as common
33 boundaries of adjacent finite elements. The value of $v|K$ on the common boundary of two adjacent finite elements is the same ($V_h \subset C^0(\overline{\Omega})$). But

the outer normals are equal and opposite from orientation considerations. (See Fig. 3.2).

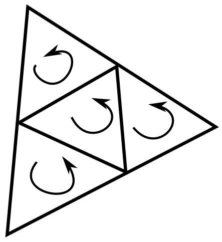

Figure 3.2:

Hence the contributions from each K along the common boundaries cancel one another. Thus the summation yields only zero. Hence v_i satisfies (3.3) for $1 \le i \le n$, and clearly $v_i \in L^2(\Omega)$. The last part of the theorem follows the characterization (2.11). $\square$

Exercise 3.1. If for all $K \in \mathscr{T}_h$, $P_K \subset H^2(K)$ and $V_h \subset C^1(\overline{\Omega})$, then show that $V_h \subset H^2(\Omega)$. Also if $v = \dfrac{\partial v}{\partial \nu} = 0$ on Γ, for all $v \in V_h$, then $V_h \subset H_0^2(\Omega)$.

We finally describe the system of linear equations associated with the space V_h. Suppose $\{w_j; 1 \le j \le M\}$ is a basis for V_h. Let u_h be the solution of (P_h). If u_h is given by

$$u_h = \sum_{j=1}^{M} u_j w_j. \tag{3.6}$$

then we have, since $a(u_h, w_i) = f(w_i)$ for $1 \le i \le M$,

$$\sum_{j=1}^{M} a(w_j, w_i) u_j = f(w_i), \quad 1 \le i \le M. \tag{3.7}$$

To find u_h, the above system of linear equations must be solved. The matrix for this system has for its (i, j)-coefficient the value $a(w_j, w_i)$. **34**

Note that the symmetry of $a(\cdot,\cdot)$ implies the symmetry of the matrix and the V-ellipticity says that the matrix is positive definite. In practical computations these observations are important.

Since we have to handle the matrix of the system, it would be ideal of course to have a diagonal matrix. We could in principle achieve this through a Gram-Schmidt orthogonalisation procedure applied to the basis functions. However such a process is not feasible since it is highly "numerically unstable". So the best we may hope for is a matrix with "a lot of" zeros in it - what is known as a *sparse matrix*.

For example in the problem given by

$$\begin{cases} -\Delta u + au = f \text{ in } \Omega \\ u = 0 \text{ on } \Gamma \end{cases}$$

the (i, j)-coefficient of the matrix is

$$a(w_j, w_i) = \int_\Omega \left(\sum_{k=1}^{n} \frac{\partial w_j}{\partial x_k} \frac{\partial w_i}{\partial x_k} + a w_j w_i \right) dx. \tag{3.8}$$

The matrix will be sparse if the supports of the basis functions are as "small" as possible so that their inner-products will be most often zero. We will study subsequently methods to achieve this. This trivial criterion extends, of course, to all types of problems.

Chapter 4

Examples of Finite Elements

WE SUMMARIZE BELOW *our requirements regarding the "finite element subspaces"* V_h **35**

(i) Let $\Omega \subset \mathbb{R}^n$ be a polygonal domain. Let $ŧ_h$ be a triangulation of Ω as in Sec. 3. Then V_h is a finite-dimensional vector space such that for all $v \in V_h$, $v|K \in P_K$ for every finite element K, where P_K is a vector space of finite dimension. Usually, *P_K is a space of polynomials*. This is of practical importance in computing the matrix of the system. We shall see later that it is of theoretical importance as well. Observe for the moment that if P_K consists of polynomials, then we automatically have that $P_K \subset H^1(K)$ or $P_K \subset H^2(K)$.

(ii) By Theorem 3.2, $V_h \subset C^0(\overline{\Omega})$ implies that $V_h \subset H^1(\Omega)$ and by Exercise 3.1 $V_h \subset C^1(\overline{\Omega})$ implies that $V_h \subset H^2(\Omega)$. Thus we must choose a *proper basis for the "local" spaces P_K such that these "global" inclusions hold.*

(iii) *There must exist at least one basis $\{w_j\}$ of V_h which consists of functions with "small" support.*

We bear these points in mind when constructing examples of finite elements. Before we proceed we need a few definitions.

Definition 4.1. An n-simplex is the convex hull in $\mathbb{R}^n$ of $(n+1)$ points $\{a_j\}_{j=1}^{n+1}$ such that if $a_j = (a_{kj})_{k=1}^{n}$ and A is the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1,n+1} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{n,n+1} \\ 1 & 1 & \dots & 1 \end{pmatrix} \tag{4.1}$$

36 then $\det A \neq 0$.

The above definition generalises the notion of a triangle to n dimensions. Geometrically the condition $\det A \neq 0$ simply means that the points $\{a_j\}_{j=1}^{n+1}$ do not lie in the same hyperplane. For $\det A$ is equal to, by elementary column operations, the determinant of the matrix

$$\begin{pmatrix} (a_{11} - a_{1,n+1}) & \dots & (a_{1,n} - a_{1,n+1}) \\ \vdots & & \vdots \\ (a_{n1} - a_{n,n+1}) & \dots & (a_{n,n} - a_{n,n+1}) \end{pmatrix}$$

and that this is non-zero means that $(a_1 - a_{n+1}), \dots, (a_n - a_{n+1})$ are linearly independent vectors in $\mathbb{R}^n$, which is the same as saying that $a_1, \dots, a_{n+1}$ do not lie in the same hyperplane.

Definition 4.2. Let $\{a_j\}_{j=1}^{n+1}$ be $(n+1)$-points in $\mathbb{R}^n$ satisfying the conditions of definition 4.1. The barycentric coordinates of any $x \in \mathbb{R}^n$ with respect to these points are numbers $\{\lambda_j\}_{j=1}^{n+1}$ such that

$$\begin{cases} x = \sum_{j=1}^{n+1} \lambda_j a_j, \\ 1 = \sum_{j=1}^{n+1} \lambda_j. \end{cases} \tag{4.2}$$

The barycentric coordinates exist because they are merely the components of the unique solution vector $\overrightarrow{\lambda}$ of the system of $(n+1)$ linear equations in $(n+1)$ unknowns given by

$$A\overrightarrow{\lambda} = \begin{pmatrix} \overrightarrow{x} \\ 1 \end{pmatrix} \quad \text{where} \quad \overrightarrow{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

The functions $\lambda_j = \lambda_j(x)$ are all affine functions of x. Also $\lambda_j(a_i) = \delta_{ij}$, where δ is the Kronecker symbol. 37

Remark 4.1. Given points $\{a_j\}_{j=1}^{n+1}$ as in the definition (4.1), the corresponding n-simplex is given by

$$K = \left\{ x = \sum_{j=1}^{n+1} \lambda_j a_j;\ 0 \le \lambda_j \le 1;\ \sum_{j=1}^{n+1} \lambda_j = 1 \right\}.$$

Definition 4.3. Let $k \ge 0$ be an integer. Then, P_k is the space of all polynomials of degree $\le k$ in $x_1, \ldots, x_n$.

We now proceed with *examples of finite elements.*

Example 4.1. The n-simplex of type (1).

Let K be an n-simplex. Let $P_K = P_1$. We define a set $\sum_K = \{p(a_i); 1 \le i \le n+1\}$ *of degrees of freedom* for $p \in P_K$, where $\{a_i\}_{i=1}^{n+1}$ are the vertices of K: *The set* $\sum_K$ *determines every polynomial* $p \in P_K$ *uniquely.* For, note that $\dim P_K = \dim P_1 = n+1$. Consider $\lambda_1, \ldots, \lambda_{n+1} \in P_1$, the barycentric coordinate functions. These are linearly independent since $\sum \alpha_k \lambda_k = 0$ implies that its value at each vertex is zero. Since $\lambda_k(a_j) = \delta_{kj}$ we get that $\alpha_j = 0$ for all j. Thus these functions form a basis for P_1. Let us write

$$p = \sum_{i=1}^{n+1} \alpha_i \lambda_i.$$

Then

$$p(a_j) = \sum_{i=1}^{n+1} \alpha_i \lambda_i(a_j) = \sum_{i=1}^{n+1} \alpha_i \delta_{ij} = \alpha_j.$$

Thus,

$$p = \sum_{i=1}^{n+1} p(a_i) \lambda_i. \tag{4.4}$$

Example 4.2. The n-simplex of type (2).

Let K be an n-simplex with vertices $\{a_i\}_{i=1}^{n+1}$. Let $a_{ij} (i < j)$ be the mid-points of the line joining a_i and a_j, i.e. $a_{ij} = \frac{1}{2}(a_i + a_j)$. **38**

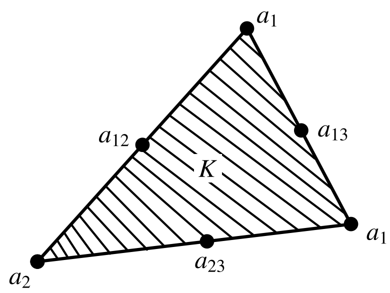

Figure 4.1:

Let $P_K = P_2$. We define for $p \in P_2$, *the set* $\sum_K = \{p(a_i), 1 \leq i \leq n+1; p(a_{ij}), 1 \leq i < j \leq n+1\}$ *of degrees of freedom.* Again $\sum_K$ determines $p \in P_2$ completely. To see this note that $\dim P_K = \binom{n+2}{2}$ and there are as many functions in the set $\{\lambda_i(2\lambda_i - 1), 1 \leq i \leq n+1; \lambda_i\lambda_j, 1 \leq i \leq j \leq n+1\}$. There are all functions in P_2. Further since

$$\lambda_i(a_j) = \delta_{ij},\ \lambda_i(a_{kj}) = \begin{cases} \frac{1}{2} & \text{if } i = k \text{ or } j, \\ 0 & \text{otherwise,} \end{cases}$$

we see again that these are linearly independent in P_2. Let us write

$$p = \sum_{i=1}^{n+1} \alpha_i\lambda_i(2\lambda_i - 1) + \sum_{1\leq i<j\leq n+1} \beta_{ij}\lambda_i\lambda_j.$$

Then

$$p(a_k) = \sum_{i=1}^{n+1} \alpha_i\delta_{ik}(2\delta_{ik} - 1) = \alpha_k.$$

Further,

$$p(a_{kl}) = \sum_{i=1}^{n} \alpha_i(2\lambda_i^2(a_{kl}) - \lambda_i(a_{kl}))$$

$$+ \sum_{1\le i<j\le n+1} \beta_{ij}\lambda_i(a_{kl})\lambda_j(a_{kl}).$$

But since $\lambda_i(a_{kl}) = 0$ or $\frac{1}{2}$, the first sum $\sum\limits_{i=1}^{n}$ is zero. Further, 39

$$\lambda_i(a_{kl})\lambda_j(a_{kl}) = \begin{cases} \frac{1}{4} & \text{if } (i,j) = (k,l) \text{ or } (l,k), \\ 0 & \text{otherwise.} \end{cases}$$

Hence $\beta_{kl} = 4p(a_{kl})$. Thus we have

$$p = \sum_{i=1}^{n+1} \lambda_i(2\lambda_i - 1)p(a_i) + \sum_{1\le i<j\le n+1} 4\lambda_i\lambda_j p(a_{ij}). \tag{4.5}$$

Example 4.3. The n-simplex of type (3).

Let K be an n-simplex with vertices $\{a_i\}_{i=1}^{n+1}$. Let $a_{iij} = \dfrac{2a_i + a_j}{3}$, $i \neq j$. Let $a_{ijk} = \dfrac{a_i + a_j + a_k}{3}$ for $i < j < k$.

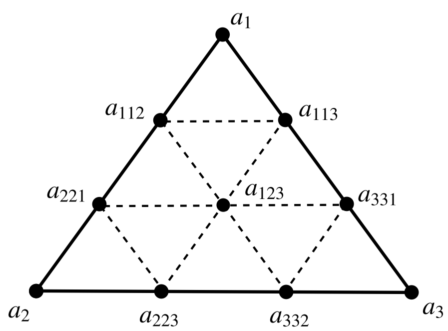

Figure 4.2:

Set $P_K = P_3$. Define *the set of degrees of freedom*

$$\sum_K = \Big\{p(a_i), 1 \le i \le n+1;\ p(a_{iij}),\ 1 \le i \neq j \le n+1;$$
$$p(a_{ijk}),\ 1 \le i < j < k \le a+1\Big\}.$$

Note that $\lambda_i(a_{iij}) = \dfrac{2}{3}$; $\lambda_j(a_{iij}) = \dfrac{1}{3}$; $\lambda_1(a_{iij}) = 0$ if $1 \neq i$, $1 \neq j$; $\lambda_1(a_{ijk}) = \dfrac{1}{3}$ if $1 = i, j$ or k and 0 otherwise, etc. Using these, one checks the linear independence of the functions

$$\Big\{\lambda_i(3\lambda_i-1)(3\lambda_i-2),\ 1 \leq i \leq n+1;\ \lambda_i\lambda_j(3\lambda_i-1),\ 1 \leq i \neq j \leq n+1;\\ \lambda_i\lambda_j\lambda_k 1 \leq i < j < k \leq n+1\Big\}.$$

These then form a basis for P_3, for there are as many functions in the above collection as $\dim P_K$. Using the values of λ_i at the special points described above, we get

$$\begin{aligned} p = {} & \sum_{i=1}^{n+1} \frac{\lambda_i(3\lambda_i-1)(3\lambda_i-2)}{2} p(a_i) \\ & + \sum_{1\leq i\neq j\leq n+1} \frac{9}{2}\lambda_i\lambda_j(3\lambda_i-1)p(a_{iij}) \\ & \sum_{1\leq i<j<k\leq n+1} 27\lambda_i\lambda_j\lambda_k p(a_{ijk}). \end{aligned} \tag{4.6}$$

40

Thus $\sum_K$ completely determines $p \in P_3$.

The points of K at which the polynomials are evaluated to get $\sum_K$ are known as the *nodes of the finite element*. The set $\sum_K$ is the set of *degrees of freedom of the finite element.*

Exercise 4.1. Generalize these ideas and describe the *n-simplex of type* (k) for any integer $k \geq 1$.

We now show how these finite elements may be used to define the space V_h.

First of all we show the inclusion $V_h \subset C^0(\overline{\Omega})$. Consider for instance a triangulation by n-simplices of type (1). Number all the nodes of the triangulation by $\{b_j\}$. Let us define $\sum_h = \{p(b_j);\ b_j$ is a node.$\}$: This is the *set of degrees of freedom of the space* V_h : *A* function v in the space V_h is, by definition, determined over each $K \in \mathfrak{t}_h$ by the values $v(b_j)$ for those nodes b_j which belong to K.

Let us examine the two-dimensional case, for simplicity. If K_1 and K_2 are two adjacent triangles with common side K' (cf. Fig. 4.3), we need to show that $v|K_1 = v|K_2$ along K' for any $v \in V_h$. Let t be an abscissa along K'.

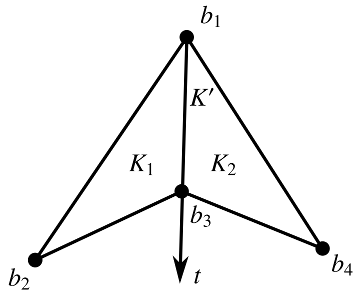

Figure 4.3:

41

Now $v|K_1$ along K' is a polynomial of degree 1 in t. So is $v|K_2$ along K'. But *these two agree at the nodes* b_1 *and* b_3. Therefore, they must be identical and hence the continuity of v follows.

This argument can be extended to any simplex of type (k). These simplices, by Theorem 3.2 yield the inclusion $V_h \subset H^1(\Omega)$ and hence we may use them for second order problems.

Exercise 4.2. The triangle of type $(3')$.

Let K be a triangle in $\mathbb{R}^2$. Define $\sum_K$ to be the values of p at the points $\{a_i, 1 \leq i \leq 3\}$, and the points $\{a_{iij}, 1 \leq i \neq j \leq 3\}$. If we define $P'_3 = \{p \in P_3; 12p(a_{123}) + 2\sum_{i=1}^{3} p(a_i) - 3\sum_{i \neq j} p(a_{iij}) = 0\}$, then show that $\sum_K$ uniquely determines $p \in P'_3 = P_K$. Further show that $P_2 \subset P'_3$.

We now relax our terminological rules about "triangulations" and admit rectangles (and in higher dimensions, hyper rectangles or hypercubes) in triangulations. We describe below some finite elements which are rectangles.

We need another space of polynomials.

Definition 4.4. Let $k \geq 1$ be an integer. Then

$$Q_k = \left\{ p; \, p(x) = \sum_{\substack{0 \leq i_j \leq k \\ 1 \leq j \leq n}} a_{i_1, \ldots, i_n} x_1^{i_1} \ldots x_n^{i_n} \right\}.$$

42

We have the inclusions $P_k \subset Q_k \subset P_{nk}$.

Example 4.4. The Rectangle of type (1).

Let K be the unit square in $\mathbb{R}^2$, i.e., $K = [0,1]^n$. Let $P_K = Q_1$. The set of degrees of freedom is given by $\sum_K = \{p(a_i), 1 \leq i \leq 4\}$; cf. Fig. 4.4 in the case $n = 2$.

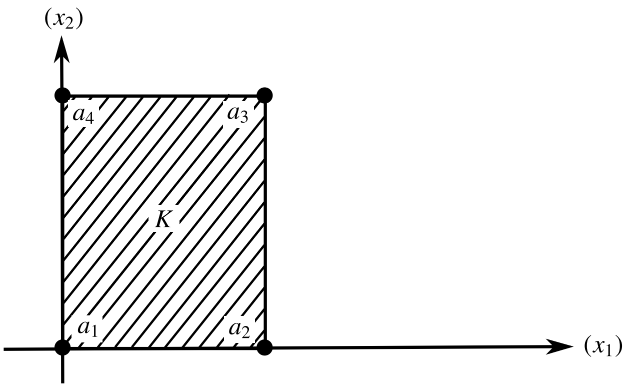

Figure 4.4:

To show that $\sum_K$ indeed determines $p \in Q_1$ uniquely we adopt a different method now. (There are essentially two methods to show that $\sum_K$ completely determines P_K; the first was used in the previous examples where we exhibited a basis for P_K such that the corresponding coefficients in the expansion of p in terms of this basis came from $\sum_K$; the second is illustrated now).

Observe first that $\dim P_K = \text{card} \sum_K = 2^n$. To determine a polynomial completely in terms of the elements of $\sum_K$ we must solve 2^n linear

equations in as many unknowns. That every polynomial is determined this way is deduced from the existence of a solution to this system. But for such a system the existence and uniqueness of the solution are equiv- 43
alent, and one establishes the latter. Thus we show that if $p \in P_K$ such that all its degrees of freedom are zero, the $p \equiv 0$.

Returning to our example, consider a polynomial $p \in Q_1$ such that $p(a_i) = 0$ for all $1 \leq i \leq 4$. On each side p is a polynomial of degree 1 either in x_1 alone or in x_2 alone. Since it vanishes at two points, the polynomial p vanishes on the sides of the square. Now consider various lines parallel to one of the axes. Here too p is a polynomial of degree 1 in one variable only. Since it vanishes at the points where the line meets the side, it also vanishes on this line. Varying the line we get $p \equiv 0$[1].

Example 4.5. The Rectangle of type (2).

Again consider the unit square (or hypercube in $\mathbb{R}^n$) to be the finite element K. Set $P_K = Q_2$, and $\sum_K = \{p(a_i), 1 \leq i \leq 9\}$ where the a_i are as in the figure below.

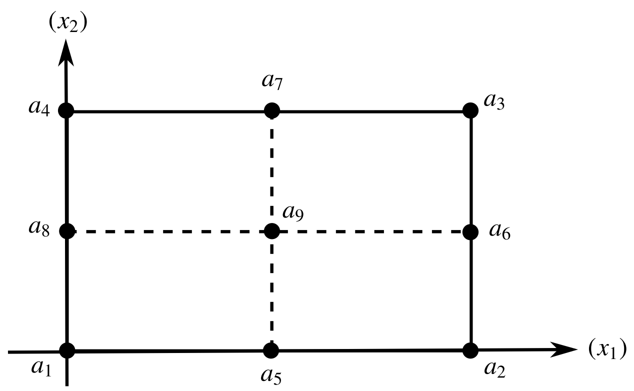

Figure 4.5:

Here again one can prove the unisolvency as above. Now let $p \in Q_2$ be given such that $p(a_i) = 0$ for all $1 \leq i \leq 9$; then $p = 0$ on the four

[1]It is not necessary to restrict ourselves to a square. Any rectangle with sides parallel to the coordinate axes would do.

sides and on the two central (dotted, in Fig. 4.5) lines. Now take lines
44 parallel to one of the axis and p vanishes on each of these. Thus $p \equiv 0$
on K and we get that $\sum_K$ uniquely determines $p \in Q_2$.

Exercise 4.3. Describe the rectangle of type (3) and generalize to hyper-rectangles of type (k).

Exercise 4.4. Prove that in all the preceding examples, we get $V_h \subset C^0(\Omega)$.

Exercise 4.5. The Rectangle of type (2′).

Let K be as in example 4.5. However omit the node a_9 (the centroid of K). Let $\sum_K = \{p(a_i), 1 \leq i \leq 8\}$ and show that this determines uniquely a function in the space

$$P_K = \left\{ p \in Q_2; 4p(a_9) + \sum_{i=1}^{4} p(a_i) - 2\sum_{i=5}^{8} p(a_i) = 0 \right\}.$$

and that $P_2 \subset P_K$.

We now turn to different types of finite elements. They differ from the preceding ones in the choice of degrees of freedom as will be seen presently.

Example 4.6. The Hermite Triangle of Type (3).

Let $K \subset \mathbb{R}^2$ be a triangle with vertices $\{a_1, a_2, a_3\}$. Let $\lambda_i, 1 \leq i \leq 3$, be the barycentric coordinate functions. Then one can check that any polynomial $p \in P_3 = P_K$ can be expanded as

$$\begin{aligned} p = {} & \sum_{i=1}^{3} (-2\lambda_i^3 + 3\lambda_i^2 - 7\lambda_1\lambda_2\lambda_3) p(a_i) + 27\lambda_1\lambda_2\lambda_3 p(a_{123}) \\ & + \sum_{i=1}^{3} \sum_{\substack{j=1 \\ j \neq i}} \lambda_i\lambda_j(2\lambda_i + \lambda_j - 1)\, \mathrm{Dp}(a_i)(a_j - a_i). \end{aligned}$$

Thus, $\sum_K = \{p(a_i), 1 \leq i \leq 3; \mathrm{Dp}(a_i)(a_j - a_i), 1 \leq i \neq j \leq 3; p(a_{123})\}$ is the corresponding set of degrees of freedom.

45 Note that $\mathrm{Dp}(a_i)$ is the Frechet derivative of p evaluated at a_i: If $(e_1, \ldots, e_n)$ is the standard basis for $\mathbb{R}^n$, then for $v : \mathbb{R}^n \to R$, we have, $(\mathrm{Dv})(x)(e_i) = \dfrac{\partial v}{\partial x_i}(x)$, the usual partial derivative.

Notice that we may replace $\sum_K$ by the set

$$\sum_K{}' = \left\{ p(a_i), 1 \leq i \leq 3; p(a_{123}); \frac{\partial p}{\partial x_1}(a_i), \frac{\partial p}{\partial x_2}(a_i), 1 \leq i \leq 3 \right\}.$$

Remark 4.2. The term "*Hermite*" means that we assume knowledge of derivatives at some of the nodes. If only the values of p at the nodes appear in the set of degrees of freedom, as was the case upto Example 4.5, we refer to the finite elements as of "*Lagrange*" type. These ideas will be made precise in Sec. 5. We usually indicate degrees of freedom involving derivatives by circling the nodes - one circle for first derivatives, two for first and second derivatives and so on. Thus the finite element of example 4.6 may be pictured as in Fig. 4.6.

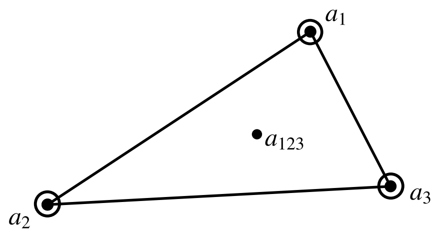

Figure 4.6:

Exercise 4.6. The Hermite Triangle of Type (3′).

This is also known as the *Zienkiewicz triangle* in Engineering literature. Set $\sum_K = \left\{ p(a_i), \dfrac{\partial p}{\partial x_1}(a_i), \dfrac{\partial p}{\partial x_2}(a_i), 1 \leq i \leq 3 \right\}$. Show that $\sum_K$ uniquely determines a function in the space

$$P_K = \left\{ p \in P_3;\ 6p(a_{123}) - 2\sum_{i=1}^{3} p(a_i) + \sum_{i=1}^{3} Dp(a_i)(a_i - a_{123}) = 0 \right\}.$$

All examples cited upto now yield the inclusion $V_h \subset C^0(\overline{\Omega})$ and consequently are useful to solve second-order problems. In order to solve problems of fourth order, we need the inclusion $V_h \subset C^1(\overline{\Omega})$. Our subsequent examples will be in this direction. 46

Remark 4.3. Consider a 1-simplex $K \subset \mathbb{R}^1$. A triangulation is merely a subdivision of Ω into subintervals. In any subinterval K not only $v|K$ but also $\dfrac{d(v|K)}{dx}$ must be continuous at both end points. Thus we get 4 conditions on $v|K$. Consequently P_K must contain all polynomials of degree 3 in it. The analogous result (which is non-trivial) is due to A. Ženišek [24] that is case of $\mathbb{R}^2$, and K a triangle of $\mathbb{R}^2$, at least polynomials of degree 5 must be contained in P_K.

Example 4.7. The Argyris triangle.

This is also known as the *21-degree-of-freedom-triangle.* We set $P_K = P_5$ and

$$\sum_K = \left\{ p(a_i), \frac{\partial p}{\partial x_1}(a_i), \ldots, \frac{\partial^2 p}{\partial x_2^2}(a_i),\ 1 \le i \le 3; \right.$$
$$\left. \frac{\partial p}{\partial \nu}(a_{ij}),\ 1 \le i \le j \le 3 \right\}.$$

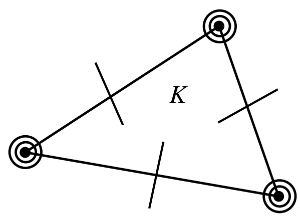

Figure 4.7:

47 The knowledge of the normal derivative $\dfrac{\partial p}{\partial \nu}$ is indicated by a line perpendicular to the side at the appropriate point; cf. Fig. 4.7.

We now show that any $p \in P_K$ is uniquely determined by $\sum_K$. Let $p \in P_K = P_5$ be given such that all its degrees of freedom are zero. If K'

is any side of K and t is an abscissa along K' then $p|K'$ is a polynomial $p_1(t)$ of degree 5. The vanishing of p, $\dfrac{dp}{dt}$, $\dfrac{d^2p}{dt^2}$ at the end points, say b, b', of K' imply that all the 6 coefficients of p_1 are 0 and hence $p_1 \equiv 0$. Thus $p = 0 = \dfrac{dp}{dt}$ on K'. The polynomial $r(t) = \dfrac{\partial p}{\partial \nu}(t)$ is of degree 4 on K' and we also have $r(b) = r(b') = \dfrac{dr}{dt}(b) = \dfrac{dr}{dt}(b') = r\left(\dfrac{b+b'}{2}\right) = 0$ which imply that $r \equiv 0$ on K'. Hence p, $\dfrac{\partial p}{\partial x_1}$, $\dfrac{\partial p}{\partial x_2}$ all vanish on the sides of the triangle K. The sides of K are defined by the equations $\lambda_i(x_1, x_2) = 0$, $(i = 1, 2, 3)$ where λ_i are the barycentric coordinate functions. We claim that λ_i^2 divides p for $i = 1, 2, 3$. To see this it is enough to prove that if p is a polynomial such that p, $\dfrac{\partial p}{\partial x_1}$, $\dfrac{\partial p}{\partial x_2}$ vanish on any straight line $L = \{(x_1, x_2); \lambda(x_1, x_2) = 0\}$ then λ^2 divides p. In the special case, when $\lambda(x_1, x_2) = x_1$ writing $p(x_1, x_2) = \sum_{j=0}^{5} a_j(x_2)x_1^j$ (with deg. $a_j \leq 5 - j$) it follows that $a_0(x_2) = a_1(x_2) = 0$ since $p = \dfrac{\partial p}{\partial x_1} = 0$ on L. Thus x_1^2 divides p. The general case reduces to this case by an affine transformation. In fact, by translating the origin to a point P, fixed arbitrarily on L and by rotation of the coordinate axes we can assume that $L = \{(X_1, X_2); X_1 = 0\}$ in the new coordinates. If p' is the image of p under this transformation then p' is also a polynomial (of degree 5) and p', $\dfrac{\partial p'}{\partial X_1}$, $\dfrac{\partial p'}{\partial X_2}$ vanish on L by chain rule for differentiation. Hence X_1^2 divides p'. This is the same thing as saying λ^2 divides p which proves the claim. Since λ_i are mutually coprime we may now write $p = q\lambda_1^2\lambda_2^2\lambda_3^2$. Then we necessarily have $q(x_1, x_2) \equiv 0$ for, otherwise deg. $p \geq 6$ which is impossible since $p \in P_5$. Hence $p \equiv 0$ on K which proves that $\sum_K$ determines $p \in P_5$.

To define the corresponding space V_h, we number all the vertices of the triangles by $\{b_j\}$ and all midpoints of the sides by $\{c_k\}$. The *the set of degrees of freedom of the space V_h is* **48**

$$\sum_h = \left\{ v(b_j), \frac{\partial v}{\partial x_1}(b_j), \frac{\partial v}{\partial x_2}(b_j), \frac{\partial^2 v}{\partial x_1^2}(b_j), \frac{\partial^2 v}{\partial x_1 \partial x_2}(b_j), \frac{\partial^2 v}{\partial x_2^2}(b_j), \frac{\partial v}{\partial \nu_k}(c_k) \right\}$$

where $\dfrac{\partial}{\partial \nu_k}$ is one of the two possible normal derivatives at the mid-point c_k.

We now show that $V_h \subset C^1(\overline{\Omega})$. Consider two adjacent Argyris triangles K_1 and K_2 with common boundary K' along which t is an abscissa (Fig. 4.8). Let $v \in V_h$. Now $v|K_1$ and $v|K_2$ are polynomials of degree 5 in t along K' and they agree together with their first and second derivatives at the end points. Thus $v|K_1 = v|K_2$ on K', proving continuity.

Now $\dfrac{\partial(v|K_1)}{\partial \nu}$ and $\dfrac{\partial(v|K_2)}{\partial \nu}$ along K' are polynomials of degree 4 agreeing in their values with first derivatives at end points and agree at the mid-point in their values. Thus $\dfrac{\partial(v|K_1)}{\partial \nu} = \dfrac{\partial(v|K_2)}{\partial \nu}$ on K'. Similarly, $\dfrac{\partial(v|K_1)}{\partial t} = \dfrac{\partial(v|K_2)}{\partial t}$ on K' and hence $v \in C^1(\overline{\Omega})$. Thus $V_h \subset C^1(\overline{\Omega})$.

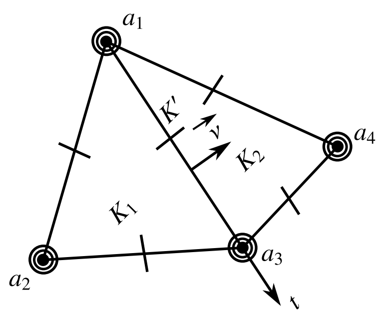

Figure 4.8:

Exercise 4.7. The 18-Degree-of-Freedom-Triangle

Let K be a triangle in $\mathbb{R}^2$. Let P_K consist of those polynomials of degree ≤ 5 for which, along each side of K, the normal derivative is a polynomial of degree ≤ 3, in one variable of course. Show that a polynomial in P_K is uniquely determined by the following set of degrees of freedom:

$$\sum_K = \left\{ p(a_i), \frac{\partial p}{\partial x_1}(a_i), \ldots, \frac{\partial^2 p}{\partial x_2^2}(a_i), 1 \leq i \leq 3 \right\}.$$

49

Note that $P_4 \subset P_K$ and $\dim P_K = 18$.

Exercise 4.8. The HCT-Triangle.

This element is due to Hsieh, Clough and Tocher. Let a be any interior point of the triangle K with vertices a_1, a_2, a_3. With a as common vertex subdivide the triangle into triangles K_1, K_2, K_3; cf. Fig. 4.9. Define

$$P_K = \left\{ p \in C^1(K);\ P|K_i \in P_3,\ 1 \le i \le 3 \right\}.$$

Obviously, $P_3 \subset P_K$. The degrees of freedom are given by

$$\sum_K = \left\{ p(a_i), \frac{\partial p}{\partial x_1}(a_i), \frac{\partial p}{\partial x_2}(a_i), 1 \le i \le 3;\ \frac{\partial p}{\partial \nu}(a_{ij}), 1 \le i < j \le 3 \right\}.$$

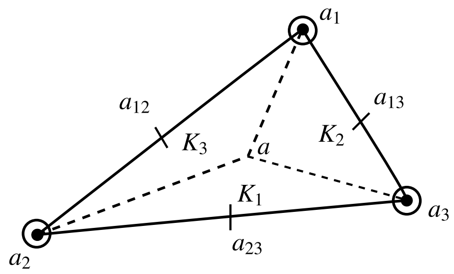

Figure 4.9:

Show that $\sum_K$ uniquely determines $p \in P_K$.

Note: Since we have to determine 3 polynomials $p_i = p|K_i$ each of degree ≤ 3, we need to determine 30 coefficients on the whole. For this we have the following conditions:

(i) The values at the vertices together with first derivatives and also the normal derivative at the mid points give 7 conditions for each $p_i = p|K_i$. Thus we have 21 conditions from these.

(ii) $p_1(a) = p_2(a) = p_3(a)$ gives 2 conditions.

(iii) $\dfrac{\partial p_1}{\partial x_i}(a) = \dfrac{\partial p_2}{\partial x_i}(a) = \dfrac{\partial p_3}{\partial x_1}(a)$ for $i = 1, 2$, gives 4 more conditions. **50**

(iv) $\dfrac{\partial p_1}{\partial \nu} = \dfrac{\partial p_2}{\partial \nu}$ along $a_1 a$ and two more similar conditions give 3 conditions.

Thus we have 30 conditions to determine the 30 coefficients. But, of course this is no proof, which is left as an exercise!

Exercise 4.9. The Bogner-Lox-Schmidt Rectangle; cf. Fig. 4.10.

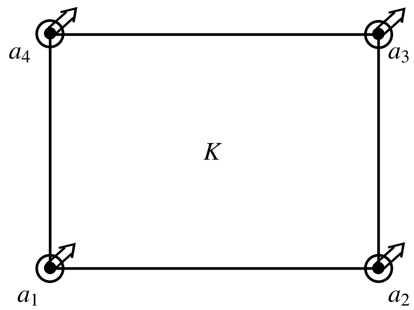

Figure 4.10:

Let $P_K = Q_3$, the degrees of freedom being given by

$$\sum_K = \left\{ p(a_i), \frac{\partial p}{\partial x_1}(a_i),\ \frac{\partial p}{\partial x_2}(a_i),\ \frac{\partial^2 p}{\partial x_1 \partial x_2}(a_i),\ 1 \leq i \leq 4 \right\}.$$

Show that $\sum_K$ determines uniquely a polynomial $p \in Q_3$ (a double dotted arrow indicates that the mixed second derivative is a degree of freedom). Show also that in this case $V_h \subset C^1(\overline{\Omega})$.

So far, we have verified requirements (i) and (ii) mentioned at the beginning of this Section. Let us now examine requirement (iii), which will be fulfilled by a "canonical" choice for the basis functions. Let $\sum_h$ be the set of degrees of freedom of the space V_h derived in an obvious way from the sets $\sum_K$, $K \in \mathfrak{t}_h$; Examples of such sets $\sum_h$ have been
51 given for n-simplices of type (k) and for Argyris triangles. Then if

$$\sum_h = \left\{ \varphi_{jh}, 1 \leq j \leq M \right\},$$

we let the basis functions w_j, $1 \leq j \leq M$, be those functions in the space V_h which satisfies

$$\varphi_i(w_j) = \delta_{ij},\ 1 \leq i \leq M.$$

Then it is easily seen that this choice will result in functions with "small" support: in Fig. 4.11, we have represented there types of supports encountered in this fashion, depending upon the position of the node associated

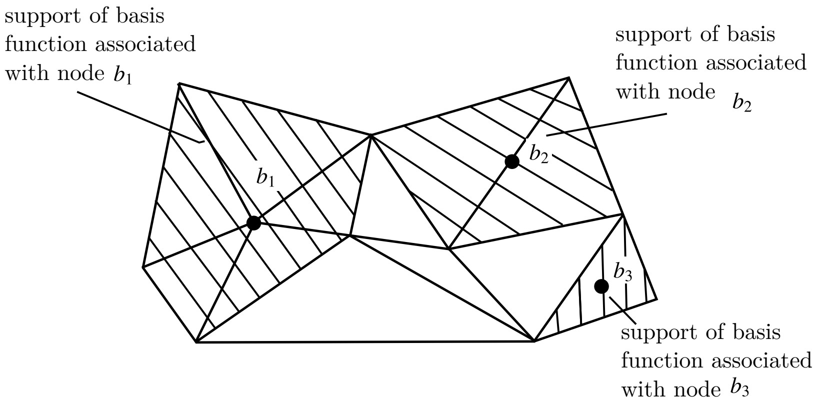

Figure 4.11:

with the degree of freedom.

Chapter 5

General Properties of Finite Elements

IT WOULD HAVE been observed that upto now we have not defined fi- **52**
nite elements in a precise manner. Various polygons like triangles, rectangles, etc. were loosely called finite elements. We rectify this omission and make precise the ideas expressed in the previous sections.

Definition 5.1. A finite element is a triple (K, Σ, P) such that

(i) $K \subset \mathbb{R}^n$ with a Lipschitz continuous boundary ∂K and $\operatorname{Int} K \neq \phi$.

(ii) Σ is a finite set of linear forms over $C^\infty(K)$. The set Σ is said to be the set of degrees of freedom of the finite element.

(iii) P is a finite dimensional space of real-valued functions over K such that Σ is P-unisolvent: i.e. if $\Sigma = \{\varphi_i\}_{i=1}^N$ and α_i, $1 \leq i \leq N$ are any scalars, then there exists a unique function $p \in P$ such that

$$\varphi_i(p) = \alpha_i,\ 1 \leq i \leq N. \tag{5.1}$$

Condition (iii) of definition (5.1) is equivalent to the conditions that $\dim P = N = \operatorname{card} \Sigma$ and that there exists a set of functions $\{p_j\}_{j=1}^N$ with $\varphi_i(p_j) = \delta_{ij} (1 \leq i, j \leq N)$, which forms a basis of P over $\mathbb{R}$. Given any

$p \in P$ we may write

$$p = \sum_{i=1}^{N} \varphi_i(p) p_i. \tag{5.2}$$

Instead of $(K, \sum, P)$ one writes at times $(K, \sum_K, P_K)$ for the finite element.

In the various examples we cited in Sec. 4 our set of degrees of freedom for a finite element K (which was an n-simplex or hyper-rectangle)
53 had elements of the following type:

Type 1: φ_i^0 given by $p \mapsto p(a_i^0)$. The points $\{a_i^0\}$ were the vertices, the mid-points of sides, etc...

Type 2: $\varphi_{i,k}^1$ given by $p \mapsto \mathrm{Dp}(a_i^1)(\xi_{i,k}^1)$. For instance, in the Hermite triangle of type (3) (cf. Example 4.6), we had $a_i^1 = a_i$, $\xi_{i,k}^1 = a_i - a_k$, where a_1, a_2, a_3 were the vertices.

Type 3: $\varphi_{i,kl}^2$ given by $p \mapsto D^2 p(a_i^2)(\xi_{i,k}^2, \xi_{i,l}^2)$. For example, in the 18-degree-of-freedom triangle, $a_i^2 = a_i$, $\xi_{i,k}^2 = e_1 = \xi_{i,1}^2$, the unit vector in the x_1-direction so that we have $D^2 p(a_i)(e_1, e_1) = \dfrac{\partial^2 p}{\partial x_1^2}(a_i)$ as a degree of freedom. (cf. Exercise 4.7).

In all these cases the points $\{a_i^s\}$ for $s = 0, 1$ and 2, are points of K and are called the *nodes of the finite element.*

Definition 5.2. A finite element is called a Lagrange finite element if its degrees of freedom are only of Type 1. Otherwise it is called a Hermite finite element. (cf. Remark 4.2)

Let (K, Σ, P) be a finite element and $v : K \to \mathbb{R}$ be a "smooth" function on K. Then by virtue of the P-unisolvency of Σ, there exists a unique element, say, $\pi v \in P$ such that $\varphi_i(\pi v) = \varphi_i(v)$ for all $1 \leq i \leq N$, where $\Sigma = \{\varphi_i\}_{i=1}^N$. The function πv is called the *P-interpolate function* of v and the operator $\pi : C^\infty(K) \to P$ is called the *P-interpolation* operator. If $\{p_j\}_{j=1}^N$ is a basis for P satisfying $\varphi_i(p_j) = \delta_{ij}$ for $1 \le i$, $j \le N$ then we have the explicit expression

$$\pi(\cdot) = \sum_{i=1}^{N} \varphi_i(\cdot) p_i. \tag{5.3}$$

Example 5.1. In the triangle of type (1) (see Example 4.1), $P = P_1$, $\Sigma = \{\varphi_i; \varphi_i(p) = p(a_i), 1 \le i \le 3\}$ and $p_i = \lambda_i$, the barycentric coordinate **54**
functions. Thus we also have

$$\pi v = \sum_{i=1}^{3} v(a_i)\lambda_i. \tag{5.4}$$

Exercise 5.1. Let K be a triangle with vertices a_1, a_2 and a_3. Let $a_{ij}(i < j)$ be the mid-point of the side joining a_i and a_j. Define $\Sigma_K = \{p \mapsto p(a_{ij}), 1 \le i \le j \le 3\}$. Show that Σ is P_1-unisolvent and that in general $V_h \not\subset C^0(\overline{\Omega})$ for a triangulation made up of such finite elements.

Exercise 5.2. Let K be a rectangle in $\mathbb{R}^2$ with vertices a_1, a_2, a_3, a_4. Let a_5, a_6, a_7, a_8 be the midpoints of the sides as in Fig. 4.5. If $\Sigma = \{p \mapsto p(a_i), 5 \le i \le 8\}$, show that Σ is not Q_1-unisolvent.

Let us now consider a *family* of finite elements of a given type. To be more specific, we will consider for instance a family of triangles of type (2) (see Example 4.2), but our subsequent descriptions extend to all types of finite elements in all dimensions.

Pick, in particular, a triangle $\hat{K}$ with vertices $\{\hat{a}_1, \hat{a}_2, \hat{a}_3\}$ from this family. Let the mid-points of the sides be $\{\hat{a}_{12}, \hat{a}_{23}, \hat{a}_{13}\}$. Set $\hat{P} = P_{\hat{K}} = P_2$ and define accordingly the associated set of degrees of freedom for $\hat{K}$ as

$$\hat{\sum} = \sum_{\hat{K}} = \{p \mapsto p(\hat{a}_i), 1 \le i \le 3;\ p \mapsto p(\hat{a}_{ij}), 1 \le i < j \le 3\}.$$

In as much as we consider the finite element $(\hat{K}, \hat{\Sigma}, \hat{P})$ as fixed in the sequel, it will be called the *reference finite element* of the family.

Given any finite element K with vertices a_1, a_2, a_3 in this family, there exists a unique invertible affine transformation of $\mathbb{R}^2$ i.e. of the

form $F_K(\hat{x}) = B_K\hat{x} + b_K$, where B_K is an invertible 2×2 matrix and
55 $b_K \in \mathbb{R}^2$, such that $F_K(\hat{K}) = K$ and $F_K(\hat{a}_i) = a_i$, $1 \le i \le 3$. It is easily verified that $F_K(\hat{a}_{ij}) = a_{ij}$ for $1 \le i < j \le 3$. Also, the space $\{p : K \to \mathbb{R}; p = \hat{p} \circ F_K^{-1}, \hat{p} \in \hat{P}\}$ is precisely the space $P_K = P_2$. Hence the family $\{(K, \Sigma, P)\}$ is equivalently defined by means of the following data:

(i) A reference finite element $(\hat{K}, \hat{\Sigma}, \hat{P})$,

(ii) A family of affine mappings $\{F_K\}$ such that $F_K(\hat{K}) = K$, $a_i = F_K(\hat{a}_i)$,

$$\begin{aligned} &1 \le i \le 3, a_{ij} = F_K(\hat{a}_{ij}), 1 \le i < j \le 3, \quad \text{and} \\ &\Sigma_K = \{p \mapsto p(F_K(\hat{a}_i)); p \mapsto p(F_K(\hat{a}_{ij}))\}, \\ &P_K = \{p : K \to \mathbb{R};\ p = \hat{p} \circ F_K^{-1},\ \hat{p} \in \hat{P}\}. \end{aligned}$$

This special case leads to the following general definition.

Definition 5.3. Two finite elements $(\hat{K}, \hat{\Sigma}, \hat{P})$ and (K, Σ, P) are affine equivalent if there exists an affine transformation F on $\mathbb{R}^n$ such that

(i) $F(\hat{x}) = B\hat{x} + b$, $b \in \mathbb{R}^n$, B an invertible $n \times n$ matrix,

(ii) $K = F(\hat{K})$,

(iii) $a_i^s = F(\hat{a}_i^s)$, $s = 0, 1, 2$,

(iv) $\xi_{i,k}^1 = B\hat{\xi}_{i,k}^1$, $\xi_{i,k}^2 = B\hat{\xi}_{i,k}^2$, $\xi_{i,l}^2 = B\hat{\xi}_{i,l'}^2$
and

(v) $P\{p : K \mapsto \mathbb{R};\ p = \hat{p} \circ F^{-1},\ \hat{p} \in \hat{P}\}$.

This leads to the next definition.

Definition 5.4. A family $\{(K, \Sigma_K, P_K)\}$ of finite elements is called an affine family if all the finite elements (K, Σ_K, P_K) are equivalent to a single reference finite element $(\hat{K}, \hat{\Sigma}, \hat{P})$.

56 Let us see why the relations given by (iv) must be precisely of that form. We have by (v), $p(x) = \hat{p}(\hat{x})$. This must be valid when we use the basis functions as well. We have:

$$\hat{p}(\hat{x}) = \sum_i \hat{p}(\hat{a}_i^0)\hat{p}_i^0(\hat{x}) + \sum_{i,k} D\hat{p}(\hat{a}_i^1)(\hat{\xi}_{i,k}^1)\hat{p}_{i,k}^1(\hat{x}) + \sum_{i,k,l} D^2\hat{p}(\hat{a}_i^2)(\hat{\xi}_{i,k}^2, \hat{\xi}_{i,l}^2)\hat{p}_{i,kl}^2(\hat{x}).$$

Now $D\hat{p}(\hat{a}_i^1)(\hat{\xi}_{i,k}^1) = Dp(a_i^1)B\hat{\xi}_{i,k}^1$ by a simple application of the chain rule and therefore

$$D\hat{p}(\hat{a}_i^1)(\hat{\xi}_{i,k}^1) = Dp(a_i^1)\xi_{i,k}^1, \quad \text{by (iv)}.$$

By a similar treatment of the second derivative term, we get

$$\hat{p}(\hat{x}) = \sum_i p(a_i^0)p_i^0(x) + \sum_{i,k} Dp(a_i^1)(\xi_i^1)(\xi_{i,k}^1)p_{i,k}(x) + \sum_{i,k,l} D^2 p(a_i^2)(\xi_{ik}^2, \xi_{il}^2)p_{ikl}(x) = p(x).$$

Thus the relations (iv) and (v) are compatible.

Theorem 5.1. *Let* (K, Σ, P) *and* $(\hat{K}, \hat{\Sigma}, \hat{P})$ *be affine equivalent with* F_K *as the* affine transformation. If $v : K \to \mathbb{R}$ induces $\hat{v} : \hat{K} \to \mathbb{R}$ by $\hat{v}(\hat{x}) = v(x)$ for $\hat{x} \in \hat{K}$, $(x = F_K(\hat{x}))$, then $\widehat{\pi v} = \hat{\pi}\hat{v}$.

Proof. Let $\hat{\Sigma} = \{\hat{\varphi}_i\}_{i=1}^N, \Sigma = \{\varphi_i\}_{i=1}^N$. By definition,

$$\hat{\varphi}_i(\widehat{\pi v}) = \varphi_i(\pi v) = \varphi_i(v),\ 1 \le i \le N. \quad \hat{\varphi}_i(\hat{\pi}\hat{v}) = \hat{\varphi}_i(\hat{v}) = \varphi_i(v),\ 1 \le i \le N.$$

Thus, $\hat{\varphi}_i(\hat{\pi}\hat{v}) = \hat{\varphi}_i(\widehat{\pi v})$ for $1 \le i \le N$. Hence $\hat{\pi}\hat{v} = \widehat{\pi v}$ by uniqueness of the function $\hat{\pi}\hat{v}$. □ 57

Let us consider a polygonal domain Ω with a triangulation $\mathfrak{t}_h$. Suppose to each $K \in \mathfrak{t}_h$ is associated a finite element, (K, Σ_K, P_K), Σ_K being the set of degrees of freedom, and P_K the finite dimensional space such that Σ_K is P_K-unisolvent. Then we have defined the interpolation operator π_K. All these make sense *locally* i.e. at a particular finite element K. We now define the *global* counterparts of these terms. The comparison is given in the following table.

Table 5.1.

Local definition	Global definition
1. Finite element K.	1. The set $\overline{\Omega} = \bigcup_{K=\mathrm{t}_n} \overline{K}$
2. The boundary of K, ∂K.	2. The boundary of Ω, $\partial\Omega = \Gamma$.
3. The space P_K of functions $K \to \mathbb{R}$, which is finite-dimensional.	3. The space V_h of functions $\Omega \to \mathbb{R}$, which is also finite-dimensional.
4. The set $\sum_K = \{\varphi_{i,K}\}_{i=1}^N$ of degrees of freedom of K.	4. The set of degrees of freedom $\sum_h = \{\varphi_i\}_{i=1}^N$, where $\varphi_i(p\vert K) = \varphi_{i,K}(p\vert K)$.
5. Basis functions of P_K are $\{p_{i,K}\}_{i=1}^N$.	5. Basis function of V_h are $\{w_j\}$.
6. The nodes of K are $\{a_i^0, a_i^1, a_i^2, \ldots\}$.	6. The nodes of t_h are by definition, $\bigcup_{K\in\mathrm{t}_h} \{\text{Nodes of } K\} = \cup\{b_j\}$ say.
7. π_K is the P_K-interpolation operator, defined by $\varphi_{i,K}(\pi_K v) = \varphi_{i,K}(v)$, for all $\varphi_{i,K} \in \sum_K$.	7. The V_h interpolation operator π_h is defined by $\pi_h v \in V_h$ such that, $\varphi_{i,K}(\pi_h v\vert K) = \varphi_{i,K}(v\vert K)$ for all $\varphi_{i,K} \in \sum_K$.

58 Notice that, by definition,

$$(\pi_h v)|K = \pi_K(v|K) \quad \text{for all} \quad K \in \mathrm{t}_h. \tag{5.5}$$

It is this property and the conclusion of theorem 5.1 that will be essential in our future error analysis.

Definition 5.5. We say that *a finite element of a given type is of class* C^0, resp. *of class* C^1, if, whenever it is the generic finite element of a triangulation, the associated space V_h satisfies the inclusion $V_h \subset C^0(\overline{\Omega})$, resp. $V_h \subset C^1(\overline{\Omega})$. By extension, a *triangulation is of class* C^0, resp. *of class* C^1 if it is made up of finite elements of class C^0, resp. of class C^1.

Reference: A forthcoming book of Ciarlet and Raviart [5].

Chapter 6

Interpolation Theory in Sobolev Spaces

WE OUTLINED THE internal approximation method in Sec. 3. We are **59**
naturally interested in the convergence of the solutions $u_h \in V_h$ to the global solution $u \in V$. As a key step in this analysis we obtained the error estimate (cf. Theorem 3.1):

$$\|u - u_h\| \leq C \inf_{v_h \in V_h} \|u - v_h\|. \tag{6.1}$$

To be more specific let us consider an example. Given $\Omega \subset \mathbb{R}^2$ a polygon, consider the solution of the following problem, which is therefore posed in the space $V = H_0^1(\Omega)$:

$$\begin{cases} -\Delta u + au = f \text{ in } \Omega, \\ u = 0 \text{ on } \Gamma. \end{cases} \tag{6.2}$$

Let t_h be a triangulation of Ω by triangles of type (1), (2) or (3). Then $u_h \in V_h \subset H_0^1(\Omega)$ and (6.1) reads as

$$\|u - u_h\|_{1,\Omega} \leq C \inf_{v_h \in V_h} \|u - v_h\|_{1,\Omega}. \tag{6.3}$$

We know 'a priori' that $u \in H_0^1(\Omega)$. Let us assume for the moment that $u \in C^0(\overline{\Omega})$. (Such assumptions are made possible by the various regularity theorems. For instance, $u \in H^2(\Omega) \subset C^0(\overline{\Omega})$ if $f \in L^2(\Omega)$ and Ω is a convex polygon). If $u \in C^0(\overline{\Omega})$, then we may define the V_h-interpolate of u, i.e., $\pi_h u$ by $\pi_h u(b_j) = u(b_j)$ for the nodes b_j of the triangulation. Note also that $\pi_h u|_K = \pi_K u$ (cf. (5.5)). Now from (6.3) we get,

$$\begin{aligned} \|u - u_h\|_{1,\Omega} &\leq C\|u - \pi_h u\|_{1,\Omega} \\ &= C\left[\sum_{K \in t_h} \|u - \pi_h u\|_{1,K}^2\right]^{\frac{1}{2}} \\ &= C\left[\sum_{K \in t_h} \|u - \pi_K u\|_{1,K}^2\right]^{\frac{1}{2}} \end{aligned}$$

60 *Thus the problem of estimating* $\|u - u_h\|_{1,\Omega}$ *is reduced to the problem of estimating* $\|u - \pi_K u\|_{1,K}$. This is one central problem in the finite element method and motivates the study of interpolation theory in Sobolev spaces.

We consider more general types of Sobolev spaces for they are no more complicated for this purpose than those defined in Sec. 2.

Definition 6.1. Let $m \geq 0$ be an integer, and $1 \leq p \leq +\infty$. Then the Sobolev space $W^{m,p}(\Omega)$ for $\Omega \subset \mathbb{R}^n$, open, is defined by

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

Remark 6.1. $H^m(\Omega) = W^{m,2}(\Omega)$.

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

$$\|v\|_{m,p,\Omega} = \left(\int_\Omega \sum_{|\alpha| \leq m} |\partial^\alpha v|^p dx\right)^{1/p} \tag{6.4}$$

and the semi-norm $|\cdot|_{m,p,\Omega}$ by

$$|v|_{m,p,\Omega} = \left(\int_\Omega \sum_{|\alpha| = m|} |\partial^\alpha v|^p dx\right)^{1/p} \tag{6.5}$$

If $k \geq 1$ is an integer, consider the space $W^{k+1,p}(\Omega)/P_k$. If $\dot{v}$ stands for the equivalence class of $v \in W^{k+1,p}(\Omega)$ we may define the analogues of (6.4) and (6.5) respectively by

$$\|\dot{v}\|_{k+1,p,\Omega} = \inf_{p \in P_k} \|v + p\|_{k+1,p,\Omega} \tag{6.6}$$

and 61

$$|\dot{v}|_{k+1,p,\Omega} = |v|_{k+1,p,\Omega}. \tag{6.7}$$

These are obviously well-defined and $\|\cdot\|_{k+1,p,\Omega}$ defines the quotient norm on the quotient space above. We then have the following key result, whose proof may be found in Nečas [20] for instance.

Theorem 6.1. *In $W^{k+1,p}(\Omega)/P_k$, the semi-norm $|\dot{v}|_{k+1,p,\Omega}$ is a norm equivalent to the quotient norm $\|\dot{v}\|_{k+1,p,\Omega}$, i.e., there exists a constant $C = C(\Omega)$ such that for all $\dot{v} \in W^{k+1,p}(\Omega)/P_k$*

$$|\dot{v}|_{k+1,p,\Omega} \leq \|\dot{v}\|_{k+1,p,\Omega} \leq C|\dot{v}|_{k+1,p,\Omega}. \tag{6.8}$$

Equivalently, we may state

Theorem 6.2. *There exists a constant $C = C(\Omega)$ such that for each $v \in W^{k+1,p}(\Omega)$*

$$\mathop{inf.}_{p \in P_k} \|v + p\|_{k+1,p,\Omega} \leq C|v|_{k+1,p,\Omega}. \tag{6.9}$$

(Note: This result holds if Ω has a continuous boundary and if it is bounded so that $P_k \subset W^{k+1,p}(\Omega)$.)

We now prove the following

Theorem 6.3. *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 $\pi \in \mathscr{L}(W^{k+1}, \rho(\Omega), W^{m,q}(\Omega))$, i.e. a continuous linear map, such that for each $p \in P_k$, $\pi p = p$. Then there exists $C = C(\Omega)$ such that for each $v \in W^{k+1,p}(\Omega)$*

$$|v - \pi v|_{m,q,\Omega} \leq C\|I - \pi\|_{\mathscr{L}(W^{k+1}, P(\Omega), W^{m,q}(\Omega))} |v|_{k+1,p,\Omega}$$

62

Proof. For each $v \in W^{k+1,p}(\Omega)$ and for each $p \in P_k$, we can write

$$v - \pi v = (I - \pi)(v + p).$$

Thus,

$$\begin{aligned} |v - \pi v|_{m,q,\Omega} &\leq \|v - \pi v\|_{m,q,\Omega} \\ &= \|I - \pi\|_{\mathscr{L}(W^{k+1,p}(\Omega), W^{m,q}(\Omega))} \|v + p\|_{k+1,p,\Omega}, \end{aligned}$$

for all $p \in P_K$. Hence,

$$\begin{aligned} |v - \pi v|_{m,q,\Omega} &\leq \|I - \pi\|_{\mathscr{L}(W^{k+1,p}(\Omega), W^{m,q}(\Omega))} \inf_{p \in P_k} \|v + p\|_{k+1,p,\Omega} \\ &\leq |C| \|I - \pi\|_{\mathscr{L}(W^{k+1,p}(\Omega), W^{m,q}(\Omega))} |v|_{k+1,p,\Omega} \end{aligned}$$

By theorem 6.2, this completes the proof. □

Definition 6.2. Two open subsets Ω, $\hat{\Omega}$ of $\mathbb{R}^n$ are said to be affine equivalent if there exists an invertible affine map F mapping $\hat{x}$ to $B\hat{x} + b$, B an invertible $(n \times n)$ matrix and $b \in \mathbb{R}^n$, such that $F(\hat{\Omega}) = \Omega$.

If Ω, $\hat{\Omega}$ are affine equivalent, then we have a bijection between their points given by $\hat{x} \leftrightarrow x = F(\hat{x})$. Also we have bijections between smooth functions on Ω and $\hat{\Omega}$ defined by $(v : \Omega \to \mathbb{R}) \leftrightarrow (\hat{v} : \hat{\Omega} \to \mathbb{R})$ where $v(x) = \hat{v}(\hat{x})$.

The following theorem gives estimates of $|v|_{m,p,\Omega}$ and $|\hat{v}|_{m,p,\hat{\Omega}}$ each in terms of the other.

63 **Theorem 6.4.** *Let Ω, $\hat{\Omega} \subset \mathbb{R}^n$ be affine equivalent. Then there exist constants C, $\hat{C}$ such that for all $v \in W^{m,p}(\Omega)$*

$$|\hat{v}|_{m,p,\hat{\Omega}} \leq C\|B\|^m \|\det B|^{-1/p} |v|_{m,p,\Omega} \tag{6.11}$$

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

$$|v|_{m,p,\Omega} \leq C\|B^{-1}\|^m |\det B|^{1/p} |\hat{v}|_{m,p,\hat{\Omega}}. \tag{6.12}$$

Note:

(i) It suffices to prove either (6.11) or (6.12). We get the other by merely interchanging the roles of Ω and $\hat{\Omega}$. We will prove the former.

(ii) $\|B\|$ is the usual norm of the linear transformation defined by B, viz. $\|B\| = \sup\limits_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \dfrac{\|Bx\|}{\|x\|}$. (Recall that $F(\hat{\Omega}) = \Omega$, $F(\hat{x}) = B\hat{x} + b$).

Proof. Let $\{e_1, \ldots, e_n\}$ be the standard basis for $\mathbb{R}^n$. Let α be a multi-index with $|\alpha| = m$. By choosing a suitable collection $\{e_{1\alpha}, \ldots, e_{m\alpha}\}$ with appropriate number of repetitions from the basis, we may write,

$$(\partial^\alpha \hat{v})(\hat{x}) = (D^m \hat{v})(\hat{x})(e_{1\alpha}, \ldots, e_{m\alpha}),$$

where $D^m\hat{v}$ is the m^{th} order Fréchet derivative of $\hat{v}$ and $D^m\hat{v}(\hat{x})$ is consequently an m-linear form on $\mathbb{R}^n$. Thus,

$$|\partial^\alpha \hat{v}(\hat{x})| \leq \|D^m\hat{v}(\hat{x})\| = \sup_{\substack{\|\xi_i\|=1 \\ 1\leq i\leq m}} |D^m\hat{v}(\hat{x})(\xi_1, \ldots, \xi_m)|.$$

Since this is true for all $|\alpha| = m$, we get

$$|\hat{v}|_{m,p,\hat{\Omega}} \leq C_1 \left(\int_{\hat{\Omega}} \|D^m\hat{v}(\hat{x})\|^p d\hat{x}\right)^{1/p} \leq C_2 |\hat{v}|_{m,p,\hat{\Omega}}. \tag{6.13}$$

The first inequality is a consequence of our preceding argument. The second follows by a straightforward argument. By composition of functions in differentiation:

$$D^m\hat{v}(\hat{x})(\xi_1, \ldots, \xi_m) = D^m v(x)(B\xi_1, \ldots, B\xi_m); \tag{6.14}$$

This gives 64

$$\|D^m\hat{v}(\hat{x})\| \leq \|D^m v(x)\| \, \|B\|^m. \tag{6.15}$$

Hence the first inequality in (6.13) may be rewritten as

$$|\hat{v}|^p_{m,p,\hat{\Omega}} \leq C_1^p \|B\|^{mp} \int_{\hat{\Omega}} \|D^m v(F(\hat{x}))\|^p d\hat{x}$$

$$
\begin{aligned}
&= C_1^p \|B\|^{mp} |\det B|^{-1} \int_{\Omega} \|D^m v(x)\|^p dx \\
&\leq C^p \|B\|^{mp} |\det B|^{-1} |v|_{m,p,\Omega}.
\end{aligned}
$$

by an inequality similar to the second inequality of (6.13). Raising to power $1/p$ on either side we get (6.11). This completes the proof. □

We now estimate the norms $\|B\|$ and $\|B^{-1}\|$ in terms of the 'sizes' of Ω and $\hat{\Omega}$. More precisely, if h, (resp. $\hat{h}$) the supremum of the diameters of all balls that can be inscribed in Ω, (resp. $\hat{\Omega}$), we have the following:

Theorem 6.5. $\|B\| \leq h/\hat{\rho}$, *and* $\|B^{-1}\| \leq \hat{h}/\rho$.

Proof. Again it suffices to establish one of these. Now,

$$
\|B\| = \sup_{\|\xi\|=\hat{\rho}} \left(\frac{1}{\hat{\rho}} \|B\xi\| \right).
$$

Let $\xi \in \mathbb{R}^n$ with $\|\xi\| = \hat{\rho}$. Choose $\hat{y}, \hat{z} \in \overline{\hat{\Omega}}$ such that $\xi = \hat{y} - \hat{z}$. Then $B\xi = B\hat{y} - B\hat{z} = y - z$, where $F(\hat{y}) = y$, $F(\hat{z}) = z$. But $y, z \in \overline{\Omega}$ and hence $\|y - z\| \leq h$. Thus $\|B\xi\| \leq h$. Hence $\|B\| \leq h/\hat{\rho}$, which completes the proof. □

We conclude this section with an *important, often used, result.*

Theorem 6.6. *Let* $(\hat{K}, \hat{\Sigma}, \hat{P})$ *be a finite element. Let* $s(= 0, 1$ *or* $2)$ *be the maximal order of derivatives occurring in* $\hat{\Sigma}$*. Assume that:*

65 *(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*

$$
\text{(6.16)} \qquad |v - \pi_K v|_{m,q,K} \leq C(meas\ K)^{\frac{1}{q} - \frac{1}{p}} \frac{h_K^{k+1}}{\rho_K^m} |v|_{k+1,p,K}
$$

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

Proof. Since $P_k \subset \hat{P}$, for any polynomial $p \in P_k$ we have $\hat{\pi} p = p$. We may write

$$
(6.17)\qquad \begin{aligned}
\hat{\pi}\hat{v} &= \sum_i \hat{v}(\hat{a}_i^0)\hat{p}_i^0 + \sum_{i,k}(D\hat{v}(\hat{a}_i^1)(\hat{\xi}_{i,k}^1))\hat{p}_{i,k}^1 \\
&= + \sum_{i,k,l}(D^2\hat{v}(\hat{a}_i^2)(\hat{\xi}_{i,k}^2, \hat{\xi}_{i,l}^2))\hat{p}_{i,k,l}^2,
\end{aligned} .
$$

all these sums being finite (the second and third may or may not be present). We claim that $\hat{\pi} \in \mathscr{L}(W^{k+1}, p(\hat{K}), W^{m,q}(\hat{K}))$. Since $\hat{P} \subset W^{m,q}(\hat{K})$ all the basis functions in (6.17) are in $W^{m,q}(\hat{K})$. Thus,

$$
(6.18)\qquad \begin{aligned}
\|\hat{\pi}\hat{v}\|_{m,q,\hat{K}} \leq & \sum_i |\hat{v}(\hat{a}_i^0)| \, \|\hat{p}_i^0\|_{m,q,\hat{K}} \\
& + \sum_{i,k} |D\hat{v}(\hat{a}_i^1)(\xi_{i,k}^1)| \, \|\hat{p}_{ik}^1\|_{m,q,\hat{K}} \\
& + \sum_{i,k,l} |D^2\hat{v}(\hat{a}_i^2)(\hat{\xi}_{i,k}^2, \hat{\xi}_{i,l}^2)| \, \|\hat{p}_{ikl}^2\|_{m,q,\hat{K}}
\end{aligned}
$$

Since $W^{k+l,p}(\hat{K}) \hookrightarrow C^s(\hat{K})$ and all the numbers $\hat{v}(\hat{a}_i^0)$, etc..., are bounded by their essential supremum over $\hat{K}$,

$$
\|\hat{\pi}\hat{v}\|_{m,q,\hat{K}} \leq C\|\hat{v}\|_{k+1,p,\hat{K}}.
$$

66

Hence the claim is valid. Now by virtue of (ii) and also our observation on preservation of polynomials, we may apply theorem 6.3 to $\hat{\pi}$. Hence there exists $C = C(\hat{K}, \hat{\Sigma}, \hat{P})$ such that

$$
|\hat{v} - \hat{\pi}\hat{v}|_{m,q,\hat{K}} \leq C|\hat{v}|_{k+1,p,\hat{K}} \quad \text{for} \quad \hat{v} \in W^{k+1,p}(\hat{K}).
$$

Notice that $\hat{\pi}\hat{v} = \widehat{\pi_K v}$ by Theorem 5.1. Thus $\hat{v} - \hat{\pi}\hat{v} = \widehat{v - \pi_K v}$. Thus if $F_K(\hat{K}) = K$ where $F_K(\hat{x}) = B_K\hat{x} + b_K$, we get

$$
(6.19)\qquad |v - \pi_K v|_{m,q,K} \leq C_1\|B_K^{-1}\|^m |\det B_K|^{1/q} |\hat{v} - \hat{\pi}\hat{v}|_{m,q,\hat{K}},
$$

by Theorem 6.4. Also by the same theorem

$$
(6.20)\qquad |\hat{v}|_{k+1,p,\hat{K}} \leq C_2\|B_K\|^{k+1} |\det B_K|^{-1/p} |v|_{k+l,p,K}.
$$

Further $|\det B_K|$ being the Jacobian of the transformation, we have $|\det B_K| = \dfrac{\text{meas } K}{\text{meas } \hat{K}}$ and $\|B_K\| \leq \dfrac{h_K}{\hat{\rho}}$, $\|B_K^{-1}\| \leq \hat{h}/\rho_K$ by theorem 6.5. Since $\hat{h}$, $\hat{\rho}$, meas $\hat{K}$ are constants, combining (6.19), (6.20) and the preceding observation we complete the proof of the theorem. ◻

References: See Bramble and Hilbert [28], Bramble and Zlǫ́mal [2], Ciarlet and Raviart [7], Ciarlet and Wagschal [8], Strang [21], Ženišek [23, 24] and Zlámal [25, 32].

Chapter 7

Applications to Second-Order Problems Over Polygonal Domains

WE APPLY THE results of the preceding section in studying the convergence of the finite element method, i.e. the convergence of the solutions u_h of (P_h) to the solution u of a problem (P) which corresponds to the choice $V = H^1(\Omega)$ or $H_0^1(\Omega)$, which we saw in Sec. 2 led to second-order problems. **67**

Let Ω be a polygonal domain throughout.

Definition 7.1. A family (t_h) of triangulations of Ω is regular is

(i) for all t_h and for each $K \in \mathrm{t}_h$, the finite elements (K, Σ, P) are all affine equivalent to a single finite element, $(\hat{K}, \hat{\Sigma}, \hat{P})$ called the reference finite element of the family;

(ii) there exists a constant σ such that for all t_h and for each $K \in \mathrm{t}_h$ we have

$$\frac{h_K}{\rho_K} \leq \sigma \tag{7.1}$$

where h_K, ρ_K are as in Theorem 6.6;

(iii) for a given triangulation t_h if

$$h = \max_{K \in \mathrm{t}_h} h_K, \tag{7.2}$$

then $h \to 0$.

Remark 7.1. The condition (ii) in definition 7.1 assures us that as $h \to 0$ the triangles do not become "flat"; cf. Exercise 7.1.

Exercise 7.1. If $n = 2$ and the sets K are triangles, show that condition (ii) of definition 7.1 is valid if and only if there exists $\theta_0 > 0$ such that
68 for all t_h and $K \in \mathrm{t}_h$, $\theta_K \geq \theta_0 > 0$, θ_K being the smallest angle in K.

Exercise 7.2. Consider the space V_h associate with t_h. Since V_h is finite dimensional all norms are equivalent and hence

$$|v_h|_{0,\infty,\Omega} \leq C_h |v_h|_{0,\Omega} \quad \text{for all} \quad v_h \in V_h,$$

for some constant C_h, a priori dependent upon h, which we may evaluate as follows: If (t_h) is a *regular family* of triangulations, show that there exists a constant C, independent of h, such that

$$|v_h|_{0,\infty,\Omega} \leq \frac{C}{h^{n/2}} |v_h|_{0,\Omega} \quad \text{for} \quad v_h \in V_h. \tag{7.3}$$

Also show that there exists a constant C such that

$$|v_h|_{1,\Omega} \leq \frac{C}{h} |v_h|_{0,\Omega} \quad \text{for all} \quad v_h \in V_h. \tag{7.4}$$

We now obtain an estimate for the error $||u - u_h||_{1,\Omega}$ when the family of triangulations is regular, which also gives convergence.

Theorem 7.1. *Let (t_h) be a regular family of triangulations on Ω of class C^0 (i.e. $V_h \subset C^0(\overline{\Omega})$) with reference finite element $(\hat{K}, \hat{\Sigma}, \hat{P})$. We assume that there exists an integer $k \geq 1$ such that*

(i) $P_K \subset \hat{P} \subset H^1(\hat{K})$

(ii) $H^{k+1}(\hat{K}) \hookrightarrow C^s(\hat{K})$ *where* $s(= 0, 1,$ *or* $2)$ *is the maximal order of derivatives in* $\hat{\sum}$.

(iii) $u \in H^{k+1}(\Omega)$ *(Regularity assumption).*

Then there exists a constant C (independent of V_h) such that

$$\|u - u_h\|_{1,\Omega} \leq C h^k |u|_{k+1,\Omega}. \tag{7.5}$$

Proof. Since $V_h \subset C^0(\overline{\Omega})$, $P_K \subset H^1(K)$, we have $V_h \subset V$. By (ii) and 69
(iii) of the hypothesis we have that the V_h-interpolate of u, viz. $\pi_h u$ is well-defined. Since $\pi_h u \in V_h$, by our fundamental result (see Theorem 3.1 or relation (6.1)), it suffices to estimate $\|u - \pi_h u\|_{1,\Omega}$. Now,

$$\begin{aligned} &H^{k+1}(\hat{K}) \hookrightarrow C^s(\hat{K}),\\ &H^{k+1}(\hat{K}) \hookrightarrow H^1(\hat{K}) \quad (k \geq 1),\\ &P_k \subset P \subset H^1(\hat{K}), \end{aligned}$$

and we may apply Theorem 6.6 with $p = q = 2$, $m = 1$ to get

$$\begin{aligned} |u - \pi_K u|_{1,K} &\leq C|u|_{k+1,K} \frac{h_K^{k+1}}{\rho_K} \\ &\leq C|u|_{k+1,K} h_K^k \quad \left(\text{Since } \frac{h_K}{\rho_K} \leq \sigma\right). \end{aligned} \tag{7.6}$$

Similarly with $m = 0$ we get

$$|u - \pi_K u|_{0,K} \leq C|u|_{k+1,K} h_K^{k+1}. \tag{7.7}$$

These together give

$$\|u - \pi_K u\|_{1,K} \leq C h_K^k |u|_{k+1,K}. \tag{7.8}$$

Now since $h_K \leq h$,

$$\begin{aligned} \|u - \pi_h u\|_{1,\Omega} &= \left(\sum_{K \in \mathfrak{t}_h} \|u - \pi_K u\|_{1,K}^2\right)^{\frac{1}{2}} \\ &\leq C\, h^k \left(\sum_{K \in \mathfrak{t}_h} |u|_{k+1,K}^2\right)^{\frac{1}{2}} \\ &= C\, h^K |u|_{k+1,\Omega}. \end{aligned}$$

This completes the proof. □

Example 7.1. Consider triangulations by triangles of type (1). Then $k = 1$, $\hat{P} = P_1$ and if $n = 2$ or 3, $H^2(\hat{K}) \hookrightarrow C^0(\hat{K})$. If $u \in H^2(\Omega)$, Theorem 7.1 says that 70

$$\|u - u_h\|_{1,\Omega} \leq C\, h|u|_{2,\Omega}.$$

We conclude the analysis of convergence in the norm $\|\cdot\|_{1,\Omega}$ with the following result.

Theorem 7.2. *Let* $(\mathfrak{t}_h)$ *be a regular family of triangulations of* Ω, *of class* C^0. *Let* $s = 0$ *or* 1 *and let* $P_1 \subset \hat{P} \subset H^1(\hat{K})$. *Then (with the assumption that* $u \in V = H^1(\Omega)$ *or* $H_0^1(\Omega)$*) we have*

$$\lim_{h\to 0} \|u - u_h\|_{1,\Omega} = 0 \tag{7.9}$$

Proof. Let $\mathscr{V} = V \cap W^{2,\infty}(\Omega)$. Since $s \leq 1$, $W^{2,\infty}(\cdot) \hookrightarrow C^s(\cdot)$ and $W^{2,\infty}(\cdot) \hookrightarrow H^1(\cdot)$. The second inclusion follows 'a fortiori' from the first with $s = 1$. Also, $P_1 \subset \hat{P} \subset H^1(\hat{K})$. Thus we may apply Theorem 6.6 with $k = 1$, $p = \infty$, $m = 1$, $q = 2$. Then for all $v \in \mathscr{V}$,

$$\begin{aligned}\|v - \pi_K v\|_{1,K} &\leq C(\text{meas } K)^{\frac{1}{2}} h|v|_{2,\infty,K}\\ &\leq C(\text{meas } K)^{\frac{1}{2}} h|v|_{2,\infty,\Omega}.\end{aligned}$$

Summing over K, we get

$$\begin{aligned}\|v - \pi_h v\|_{1,\Omega} &\leq C\, h|v|_{2,\infty,\Omega}\left(\sum_{K\in\mathfrak{t}_h} \text{meas } K\right)^{\frac{1}{2}}\\ &= C\, h|v|_{2,\infty,\Omega},\end{aligned}$$

since $\sum_{K\in\mathfrak{t}_h}$ meas K = meas Ω, a constant. Thus, for all $v \in \mathscr{V}$,

$$\lim_{h\to 0} \|v - \pi_h v\|_{1,\Omega} = 0 \tag{7.10}$$

71

Notice that $\overline{\mathscr{V}} = V$. Hence choose $v_0 \in \mathscr{V}$ such that $\|u - v_0\|_{1,\Omega} \leq \epsilon/2$ where $\epsilon > 0$ is any preassigned quantity. Then once v_0 is chosen, by

(7.10) choose h_0 such that for all $h \leq h_0$, $\|v_0 - \pi_h v_0\|_{1,\Omega} \leq \epsilon/2$. Now, by (6.1)

$$\begin{aligned}\|u - u_h\|_{1,\Omega} &\leq C\|u - \pi_h v_0\|_{1,\Omega}\\ &\leq C\left(\|u - v_0\|_{1,\Omega} + \|v_0 - \pi_h v_0\|_{1,\Omega}\right)\\ &\leq C\epsilon, \quad \text{for} \quad h \leq h_0.\end{aligned}$$

This gives (7.9) and completes the proof. ◻

We now have, by Theorem 7.1, $|u - u_h|_{0,\Omega} \leq \|u - u_h\|_{1,\Omega} = 0(h^k)$. We now show, by another argument that $|u - u_h|_{0,\Omega} = 0(h^{k+1})$, (at least in some cases) there by giving a more rapid convergence than expected. This is done by the *Aubin-Nitsche argument* (also known as the *duality argument*). We describe this in an abstract setting.

Let V be a normed space with norm denoted by $\|\cdot\|$. Let H be a Hilbert space with norm $|\cdot|$ and inner product $(\cdot,\cdot)$ such that

$$(7.11) \qquad \begin{cases}\text{(i) } V \hookrightarrow H, \quad \text{and}\\ \text{(ii) } \overline{V} = H.\end{cases}$$

For second-order problems: $V = H^1(\Omega)$ or $H^1_0(\Omega)$ and $H = L^2(\Omega)$.

Since H is a Hilbert space, we may identify it with its dual. Further since V is dense in H, we have that H may be identified with a subspace of V', the dual of V. For, if $g \in H$, define $\tilde{g} \in V'$ by $\tilde{g}(v) = (g, v) \cdot \tilde{g} \in V'$ since $|\tilde{g}(v)| \leq C|g|\,\|v\|$. If $\tilde{g}(v) = 0$ for all $v \in V$, then $(g, v) = 0$ for all $v \in H$ as well since $\overline{V} = H$. Thus $g = 0$. This proves the identification. In the sequel we will set $g = \tilde{g}$. 72

Recall that u and u_h are the solutions of the problems:

$$(P) \qquad a(u, v) = f(v) \quad \text{for all} \quad v \in V,$$

$$(P_h) \qquad a(u_h, v_h) = f(v_h) \quad \text{for all} \quad v_h \in V_h \subset V,$$

and that the assumptions on (P) are as in the Lax-Milgram lemma. Then we have the following theorem.

Theorem 7.3. *Let the spaces H and V satisfy* (7.11). *Then with our above mentioned notations,*

$$|u-u_h| \leq M\|u-u_h\| \left[\sup_{g\in H}\left\{\frac{1}{|g|}\inf_{\varphi_h\in V_h}\|\varphi-\varphi_h\|\right\}\right], \tag{7.12}$$

where for each $g \in H$, $\varphi \in V$ is the corresponding unique solution of the problem

$$a(v,\varphi) = (g,v) \quad \textit{for all} \quad v \in V, \tag{P^*}$$

and M the constant occurring in the inequality giving continuity of $a(\cdot,\cdot)$.

Remark 7.2. Note that unlike in (P), we solve for the *second* argument of $a(\cdot,\cdot)$ in (P^*). This is called the *adjoint problem* of (P). The existence and uniqueness of the solution of (P^*) are proved in an identical manner. Note that if $a(\cdot,\cdot)$ is symmetric, then (P) is *self-adjoint* in the sense that $(P) = (P^*)$.

Proof. From the elementary theory of Hilbert spaces, we have

$$|u-u_h| = \sup_{\substack{g\in H\\ g\neq 0}} \frac{|(g,u-u_h)|}{|g|}. \tag{7.13}$$

For a given $g \in H$,

$$(g,u-u_{\underline{h}}) = a(u-u_{\underline{h}},\varphi) \tag{7.14}$$

73 Also if $\varphi_h \in V_h$ we have,

$$a(u-u_h,\varphi_h) = 0. \tag{7.15}$$

Thus (7.14) and (7.15) give

$$(g,u-u_h) = a(u-u_h,\varphi-\varphi_h), \tag{7.16}$$

which gives us

$$|(g,u-u_h)| \leq M\|u-u_h\|\,\|\varphi-\varphi_h\|,$$

and hence

$$|u - u_h| \leq M||u - u_h| \sup_{\substack{g \in H \\ g \neq 0}} \left(\frac{||\varphi - \varphi_h||}{|g|} \right).$$

by (7.13). Since this is true for any $\varphi_h \in V_h$ we may take infimum over V_h to get (7.12), which completes the proof. □

For dimensions ≤ 3 and Lagrange finite elements we now show that $|u - u_h|_{0,\Omega} = 0(h^{k+1})$. For this we need one more definition.

Definition 7.2. Let $V = H^1(\Omega)$ or $H_0^1(\Omega)$, $H = L^2(\Omega)$. The adjoint problem is said to be regular if the following hold:

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

(ii) there exists a constant C such that for all $g \in L^2(\Omega)$

$$||\varphi||_{2,\Omega} \leq C|g|_{0,\Omega}, \tag{7.17}$$

where φ is the solution of the adjoint problem for g.

Theorem 7.4. *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* $u \in H^{k+1}(\Omega)$, $P_k \subset \hat{P} \subset H^1(\hat{K})$. *Assume* **74**
further that the adjoint problem is regular in the sense of Definition 7.2. Then there exists a constant C *independent of* h *such that*

$$|u - u_h|_{0,\Omega} \leq C\, h^{k+1}|u|_{k+1,\Omega}. \tag{7.18}$$

Proof. Since $n \leq 3$, $H^2(\cdot) \hookrightarrow C^0(\cdot)$. Also, $H^2(\cdot) \hookrightarrow H^1(\cdot)$ and $P_1 \subset \hat{P} \subset H^1(\hat{H})$. Thus for $\varphi \in H^2(\Omega)$, by Theorem 7.1,

$$||\varphi - \pi_h\varphi||_{1,\Omega} \leq C\, h|\varphi|_{2,\Omega}.$$

Hence

$$\inf_{\varphi_h \in V_h} ||\varphi - \varphi_h||_{1,\Omega} \leq C\, h|\varphi|_{2,\Omega}. \tag{7.19}$$

By (7.12) and (7.19).

$$|u-u_h|_{0,\Omega} \leq M\|u-u_h\|_{1,\Omega} \sup_{g\in L^2(\Omega)} \left(\frac{1}{|g|_{0,\Omega}} Ch|\varphi|_{2,\Omega}\right).$$

By the regularity of (P^*),

$$\frac{|\varphi|_{2,\Omega}}{|g|_{0,\Omega}} \leq \frac{\|\varphi\|_{2,\Omega}}{|g|_{0,\Omega}} \leq \text{ constant.} \tag{7.20}$$

Thus, $|u-u_h|_{0,\Omega} \leq C\, h\|u-u_h\|_{1,\Omega}$

$$\leq C\, h(h^k|u|_{k+1,\Omega}) \quad \text{(by theorem 7.1).}$$

This gives (7.18) and completes the proof. □

We finally give an estimate for the error in the L^∞-norm.

Theorem 7.5. *Let* $(\mathfrak{t}_h)$ *be a regular family of triangulations on* $\Omega \subset \mathbb{R}^n$, *where* $n \leq 3$. *Assume further that for all* $\mathfrak{t}_h$ *and* $K \in \mathfrak{t}_h$.

$$0 < \tau \leq \frac{h_K}{h} \leq, frm[o]--, \quad \tau \text{ being a constant.} \tag{7.21}$$

75 *Let* $u \in H^2(\Omega)$ *and* $P_1 \subset \hat{P} \subset H^1(\hat{K}) \cap L^\infty(\hat{K})$. *If* (P^*) *is regular, then there exists a constant* C *independent of* h *such that*

$$\begin{cases} |u-u_h|_{0,\infty,\Omega} \leq C\, h|u|_{2,\Omega}; & \text{if } n=2 \\ |u-u_h|_{0,\infty,\Omega} \leq C\, \sqrt{h}|u|_{2,\Omega} & \text{if } n=3. \end{cases} \tag{7.22}$$

Proof. Assume $n = 2$. Now

$$|u-u_h|_{0,\infty,\Omega} \leq |u-\pi_h u|_{0,\infty,\Omega} + |\pi_h u - u_h|_{0,\infty,\Omega}. \tag{7.23}$$

Note that since $(u_h - \pi_h u) \in V_h$, we may apply Exercise 7.1 to get

$$|u_h-\pi_h u|_{0,\infty,\Omega} \leq \frac{C}{h}|u_h-\pi_h u|_{0,\Omega}. \tag{7.24}$$

Thus,

$$|u_h-\pi_h u|_{0,\infty,\Omega} \leq \frac{C}{h}\left[|u_h-u|_{0,\Omega} + |u-\pi_h u|_{0,\Omega}\right]$$

$$\leq \frac{C}{h}\left[C_1 h^2 |u|_{2,\Omega} + C_2 h^2 |u|_{2,\Omega}\right]$$
$$\leq C\, h|u|_{2,\Omega} \quad \text{(by Theorem 7.4 and Theorem 6.6).}$$

Also $H^2(\cdot) \hookrightarrow C^0(\cdot)$; $H^2(\cdot) \hookrightarrow L^\infty(\cdot)$ and $P_1 \subset \hat{P} \subset L^\infty(\hat{K})$. Thus, by Theorem 6.6 with $k = 1$, $p = 2$, $m = 0$, $q = \infty$,

$$|u - \pi_K u|_{0,\infty,K} \leq C(\text{meas}\, K)^{-\frac{1}{2}} h^2 |u|_{2,K}.$$

Since $n = 2$,

$$\text{meas}\, K \geq C\rho_K^2 \geq \frac{C}{\sigma^2} h_K^2 \geq \frac{C\tau^2}{\sigma^2} h^2$$

by (7.1), so that $(\text{meas}\, K)^{-\frac{1}{2}} \leq C\, h^{-1}$ and therefore,

$$|u - \pi_K u|_{0,\infty,K} \leq C\, h|u|_{2,K}.$$

Hence we obtain (7.22) for $n = 2$ since

$$|u - \pi_h u|_{0,\infty,\Omega} = \max_{K \in t_h} |u - \pi_K u|_{0,\infty,K} \leq C\, h|u|_{2,\Omega}.$$

76

For $n = 3$, the only variation in the proof occurs in the fact that

$$|u_h - \pi_h u|_{0,\infty,\Omega} \leq \frac{C}{h^{3/2}} |u_h - \pi_h u|_{0,\Omega}$$

as in Exercise 7.1 and that now

$$\text{meas}\, K \geq C\rho_K^3 \geq \frac{C}{\sigma^3} \cdot h_K^3 \geq \frac{C\tau^3}{\sigma^3} \cdot h^3.$$

This completes the proof. □

References: One may refer to Ciarlet and Raviart [6] for $0(h)$ convergence in the norm $|\cdot|_{0,\infty,\Omega}$ for any n. See also Bramble and Thomée [1].

Chapter 8

Numerical Integration

LET US START with a specific problem. Let Ω be a polygonal domain in $\mathbb{R}^n$. Consider the problem 77

$$
\begin{cases} -\sum\limits_{i,j=1}^{n} \frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial u}{\partial x_j}\right) = f \text{ in } \Omega, \\ u = 0 \text{ on } \Gamma = \partial\Omega. \end{cases} \tag{8.1}
$$

where the (a_{ij}) and f are functions over $\overline{\Omega}$ which are smooth enough. Let us further assume that there exists $\alpha > 0$ such that, for all $\xi \in \mathbb{R}^n$,

$$
\sum_{i,j=1}^{n} a_{ij}\xi_i\xi_j \geq \alpha \sum_{i=1}^{n} \xi_i^2. \tag{8.2}
$$

It has been seen earlier (cf. Remark 2.2) that the above problem (8.1) is obtained from a problem (P) with $a(\cdot,\cdot)$ and $f(\cdot)$ being defined by

$$
\begin{cases} a(u,v) = \int_\Omega \sum_{i,j=1}^{n} a_{ij}\frac{\partial u}{\partial x_j}\frac{\partial v}{\partial x_i} dx \\ f(v) = \int_\Omega fv\, dx \end{cases} \tag{8.3}
$$

for $u, v \in V = H_0^1(\Omega)$.

Approximating the solution by the finite element method, i.e. by constructing a regular family of triangulations (t_h) with reference finite element $(\hat{K}, \hat{\Sigma}, \hat{P})$, we get the problems (P_h), i.e., to find $u_h \in V_h$ such that

$$a(u_h, v_h) = f(v_h), \quad \text{for all} \quad v_h \in V_h. \tag{8.4}$$

If we choose a basis $\{w_k\}_{k=1}^{M}$ for V_h, then we may write

$$u_h = \sum_{k=1}^{M} u_k w_k. \tag{8.5}$$

Thus to solve (P_h) we have to solve the linear system

$$\sum_{k=1}^{M} a(w_k, w_m) u_k = f(w_m),\ 1 \le m \le M. \tag{8.6}$$

78

Notice that

$$\begin{aligned} a(w_k, w_m) &= \sum_{K \in \mathfrak{t}_h} \int_K \sum_{i,j=1}^{n} a_{ij} \frac{\partial w_k}{\partial x_j} \frac{\partial w_m}{\partial x_i} dx \\ f(w_m) &= \sum_{K \in \mathfrak{t}_h} \int_K f w_m dx. \end{aligned} \tag{8.7}$$

Thus we have ended up with the computations of integrals over $K \in \mathfrak{t}_h$. These are, in general, difficult or impossible to evaluate exactly and one thus has to resort to numerical methods. We now study briefly how this may be done.

Let us assume $F_K(\hat{K}) = K$, where, $F_K(\hat{x}) = B_K \hat{x} + b_K$, with $\det B_K > 0$. There is no loss in generality in the last assumption. Then if φ is a function over K, we have

$$\int_K \varphi(x) dx = (\det B_K) \int_{\hat{K}} \hat{\varphi}(\hat{x}) d\hat{x} \tag{8.8}$$

the functions φ and $\hat{\varphi}$ being in the usual correspondence. We then replace the expression in the right-hand side by the following:

$$\int_{\hat{K}} \hat{\varphi}(\hat{x}) \widehat{dx} \sim \sum_{1=1}^{L} \hat{\omega}_1 \hat{\varphi}(\hat{b}_1). \tag{8.9}$$

In this section $\sim$ will denote the right-hand side replacing the expression in the left hand side in similar relations). In (8.9) the quantities $\hat{\omega}_1$ are called the *weights* and the points $\hat{b}_1$ are called the *nodes* of the *quadrature scheme*. While in general we may assume $\hat{\omega}_1 \in \mathbb{R}$ and $\hat{b}_1 \in \mathbb{R}^n$, *we will restrict ourselves to the most common case where* $\hat{\omega}_1 > 0$ and $\hat{b}_1 \in \hat{K}$, $1 \leq l \leq L$.

We may now define the *error functional* $\hat{\mathscr{E}}$ by

$$\hat{\mathscr{E}}(\hat{\varphi}) = \int_{\hat{K}} \hat{\varphi}(\hat{x})d\hat{x} - \sum_{1=1}^{L} \hat{\omega}_1 \hat{\varphi}(\hat{b}_1). \tag{8.10}$$

79

We will be interested in finding spaces of *polynomials* for which $\hat{\mathscr{E}}(\hat{\varphi}) = 0$, i.e., again we need "polynomial invariance", an idea already found in interpolation theory. *The above quadrature scheme for* $\hat{K}$ *induces one on* K *as well since if we set*

$$\begin{cases} \omega_{1,K} = (\det B_K)\hat{\omega}_1, \\ b_{1,k} = F_K(\hat{b}_1), \end{cases} \tag{8.11}$$

we then deduce the numerical quadrature scheme

$$\int_K \varphi(x)dx \sim \sum_{1=1}^{L} \omega_{1,K}\varphi(b_{1,K}). \tag{8.12}$$

We shall therefore define the error functional

$$\mathscr{E}_K(\varphi) = \int_K \varphi(x)dx - \sum_{l=1}^{L} \omega_{1,K}\varphi(b_{1,K}). \tag{8.13}$$

Notice that the following relation holds:

$$\mathscr{E}_K(\varphi) = (\det B_K)\hat{\mathscr{E}}(\hat{\varphi}). \tag{8.14}$$

Example 8.1. Consider $\hat{K}$ to be an n-simplex in $\mathbb{R}^n$. Let $\hat{a}$ be its centroid. Let

$$\int_{\hat{K}} \hat{\varphi}(\hat{x})d\hat{x} \sim (\operatorname{meas} \hat{K})\hat{\varphi}(\hat{a}).$$

Exercise 8.1. Show that $\mathscr{E}(\hat{\varphi}) = 0$ for $\hat{\varphi} \in P_1$ in Example 8.1.

Example 8.2. Let $\hat{K}$ be a triangle in $\mathbb{R}^2$. With the usual notations, set

$$\int_{\hat{K}} \hat{\varphi}(\hat{x})d\hat{x} \sim \frac{1}{3}(\text{meas}\,\hat{K}) \sum_{1\le i<j\le 3} \hat{\varphi}(\hat{a}_{ij}).$$

80 **Exercise 8.2.** Show that $\hat{\mathscr{E}}(\hat{\varphi}) = 0$ for $\hat{\varphi} \in P_2$ in Example 8.2.

Example 8.3. Let $\hat{K}$ be as in Example 8.2. Let (cf. Fig. 8.1):

$$\int_{\hat{K}} \hat{\varphi}(\hat{x})d\hat{x} \sim \frac{1}{60}(\text{meas}\,\hat{K}) \left[3\sum_{i=1}^{3} \hat{\varphi}(\hat{a}_i) + 8 \sum_{1\le i<j\le 3} \hat{\varphi}(\hat{a}_{ij}) + 27\hat{\varphi}(\hat{a}) \right].$$

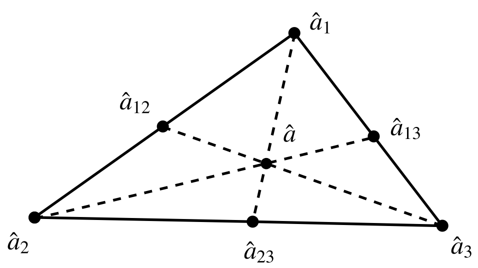

Figure 8.1:

Exercise 8.3. Show that $\hat{\mathscr{E}}(\hat{\varphi}) = 0$ for $\hat{\varphi} \in P_3$, in Example 8.3.

Let us now review the whole situation. We had the “original” approximation problem (P_h): To find $u_h \in V_h$ such that $a(u_h, v_h) = f(v_h)$ for all $v_h \in V_h$.

This led to the solution of the linear system (8.6). By virtue of the quadrature scheme we arrive at a solution of a “modified” approximation problem (P_h^*): To solve the linear system

$$\sum_{k=1}^{M} a_h(w_k, w_m)u_k^* = f_h(w_m), \; 1 \le m \le M, \tag{8.15}$$

where

$$\text{(8.16)} \qquad \begin{cases} a_h(w_k, w_m) = \sum\limits_{K\in \mathfrak{t}_h} \sum_{l=1}^{L} \omega_{1,K} \left(\sum\limits_{i,j=1}^{n} a_{ij} \frac{\partial w_k}{\partial x_j} \frac{\partial w_m}{\partial x_i} (b_{1,K}) \right) \\ f_h(w_m) = \sum\limits_{K\in \mathfrak{t}_h} \left(\sum\limits_{l=1}^{L} \omega_{1,K} (f w_m)(b_{1,K}) \right). \end{cases}$$

While u_h was given by (8.5) we now obtain

$$\text{(8.17)} \qquad u_h^* = \sum_{k=1}^{M} u_k^* w_k.$$

Thus the problem (P_h^*) (not to be confused with any adjoint problem!) *consists in finding* $u_h^* \in V_h$ *such that, for all* $w_h \in V_h$, **81**

$$\text{(8.18)} \qquad a_h(u_h^*, w_h) = f_h(w_h)$$

where

$$\text{(8.19)} \qquad \begin{cases} a_h(v_h, w_h) = \sum\limits_{K\in \mathfrak{t}_h} \sum\limits_{l=1}^{L} \omega_{1,K} \left(\sum\limits_{i,j=1}^{n} a_{ij} \frac{\partial v_h}{\partial x_j} \frac{\partial w_h}{\partial x_i} (b_{1,K}) \right) \\ f_h(v_h) = \sum\limits_{K\in \mathfrak{t}_h} \sum\limits_{l=1}^{L} \omega_{1,K} (f v_h)(b_{1,K}), \end{cases}$$

for $v_h, w_h \in V_h$.

Remark 8.1. The bilinear form $a_h(\cdot,\cdot) : V_h \times V_h \to \mathbb{R}$ and the linear form $f_h : V_h \to \mathbb{R}$ are *not* defined over V in general. For instance if $V = H_0^1(\Omega)(n = 2)$ in one of the examples, then as they require point values of the nodes, we see that they are not in general defined over V.

Having obtained the approximate solution u_h^* by numerical integration, we are naturally interested in its efficacy. Thus we require to know the error $||u - u_h^*||$. We now carry out the error analysis, first in an abstract setting.

Let us maintain our assumptions as in the Lax-Milgram lemma and consider the problem (P). Then we have problems (P_h^*) to find $u_h^* \in V_h \subset V$ such that for all $v_h \in V_h$, $a_h(u_h^*, v_h) = f_h(v_h)$ where $f_h \in V_h'$ and $a_h(\cdot,\cdot)$ is a bilinear form on V_h. Then we would like to answer the following questions:

(i) What are sufficient conditions such that (P_h^*) have unique solutions?

(ii) Can we find an abstract error estimate for $\|u - u_h^*\|$?

(iii) If $\|u - u_h\| = 0(h^k)$, i.e., without numerical quadrature, under what conditions is this order of convergence preserved, i.e. when can we say $\|u - u_h^*\| = 0(h^k)$?

82 The assumption of V_h-ellipticity of the bilinear forms $a_h(\cdot, \cdot)$ answers the first question (by the Lax-Milgram lemma) and we will see in Theorem 8.2 under which assumptions it is valid. The following theorem answers the second question.

Theorem 8.1. *Let the bilinear forms $a_h(\cdot, \cdot)$ be V_h-elliptic uniformly with respect to h, i.e., there exists a constant $\widetilde{\alpha} > 0$, independent of h, such that for all h and for all $v_h \in V_h$,*

$$a_h(v_h, v_h) \geq \widetilde{\alpha}\|v_h\|^2. \tag{8.20}$$

Then the approximate problems (P_h^*) all have unique solutions u_h^*, and further we have the estimate:

$$\|u - u_h^*\| \leq C\left(\inf_{v_h \in V_h}\left\{\|u - v_h\| + \sup_{w_h \in V_h}\frac{|a(v_h, w_h) - a_h(v_h, w_h)|}{\|w_h\|}\right\} + \sup_{w_h \in V_h}\frac{|f(w_h) - f_h(w_h)|}{\|w_h\|}\right). \tag{8.21}$$

Remark 8.2. If $a = a_h$, $f = f_h$ then we get our original estimate (3.1). Thus (8.21) generalizes our previous result.

Remark 8.3. The terms involving a, a_h and f, f_h merely mean that if u_h^* is to converge to u, then a_h and f_h must be "close to" a and f respectively. Their convergence to 0 with h may be viewed as "*consistency conditions*" which are so often found in Numerical Analysis.

Proof. The existence and uniqueness of the u_h^* are obvious by the Lax-Milgram lemma applied to the V_h. Since, for all $v_h \in V_h$, we have

$$a_h(u_h^*, u_h^* - v_h) = f_h(u_h^* - v_h),$$

$$a(u, u_h^* - v_h) = f(u_h^* - v_h),$$

83 we have the identity

$$\begin{aligned} a_h(u_h^* - v_h, u_h^* - v_h) &= a(u - v_h, u_h^* - v_h) + \{a(v_h, u_h^* - v_h) \\ &\quad - a_h(v_h, u_h^* - v_h)\} + \{f_h(u_h^* - v_h) - f(u_h^* - v_h)\}. \end{aligned} \tag{8.22}$$

Hence by (8.20) we get

$$\begin{aligned} \widetilde{\alpha}\|u_h^* - v_h\|^2 \leq{} & M\|u - v_h\| \, \|u_h^* - v_h\| \\ & + |a(v_h, u_h^* - v_h) - a_h(v_h, u_h^* - v_h)| \\ & + |f_h(u_h^* - v_h) - f(u_h^* - v_h)|. \end{aligned}$$

Thus,

$$\begin{aligned} \widetilde{\alpha}\|u_h^* - v_h\| &\leq M\|u - v_h\| + \frac{|a(v_h, u_h^* - v_h) - a_h(v_h, u_h^* - v_h)|}{\|u_h^* - v_h\|} \\ &\quad + \frac{|f_h(u_h^* - v_h) - f(u_h^* - v_h)|}{\|u_h^* - v_h\|} \\ &\leq M\|u - v_h\| + \sup_{w_h \in V_h} \frac{|a(v_h, w_h) - a_h(v_h, w_h)|}{\|w_h\|} \\ &\quad + \sup_{w_h \in V_h} \frac{|f(w_h) - f_h(w_h)|}{\|w_h\|} \end{aligned}$$

since $(u_h^* - v_h) \in V_h$. Hence,

$$\begin{aligned} \|u - u_h^*\| &\leq \|u - v_h\| + \|u_h^* - v_h\| \\ &\leq \left(1 + \frac{M}{\widetilde{\alpha}}\right)\|u - v_h\| + \frac{1}{\widetilde{\alpha}} \sup_{w_h \in V_h} \frac{|a(v_h, w_h) - a_h(v_h, w_h)|}{\|w_h\|} \\ &\quad + \frac{1}{\widetilde{\alpha}} \sup_{w_h \in V_h} \frac{|f(w_h) - f_h(w_h)|}{\|w_h\|}. \end{aligned}$$

Varying v_h over V_h and taking the infimum, and replacing $(1 + M/\widetilde{\alpha})$, 84
$(1/\widetilde{\alpha})$ by a larger constant C, we get (8.21), which completes the proof.

□

The following theorem tells us when the uniform V_h-ellipticity assumption of Theorem 8.1 is satisfied in the example we started with.

Theorem 8.2. *Let $a_h(\cdot,\cdot)$ and $f_h(\cdot)$ be as in* (8.19). *Assume further that* (i) $\hat{\omega}_1 > 0$, $1 \le l \le L$, (ii) $\hat{P} \subset P_k$, *and* (iii) $\bigcup_{l=1}^{L}\{\hat{b}_l\}$ *contains a $P_{k'-1}$ unisolvent subset. Then the $a_h(\cdot,\cdot)$ are V_h-elliptic uniformly with respect to h.*

Proof. We must produce an $\tilde{\alpha} > 0$, free of h, such that $a_h(v_h, v_h) \ge \tilde{\alpha}|v_h|^2_{1,\Omega}$ for all $v_h \in V_h$. We have

$$\begin{aligned} a_h(v_h,v_h) &= \sum_{K\in t_h}\sum_{l=1}^{L}\omega_{1,K}\left(\sum_{i,j=1}^{n} a_{ij}\frac{\partial v_h}{\partial x_j}\frac{\partial v_h}{\partial x_i}\right)(b_{l,K}) \\ &\ge \alpha\sum_{K\in t_h}\sum_{l=1}^{L}\omega_{l,K}\left(\sum_{i=1}^{n}\left(\frac{\partial p_K}{\partial x_i}(b_{l.K})\right)^2\right) \end{aligned} \tag{8.23}$$

where $p_K = v_h|_K$. The inequality (8.23) is a result of the ellipticity condition (8.2) on the matrix (a_{ij}) and the fact that $\omega_{l,K} > 0$, since $\widetilde{\omega}_1 > 0$ and we assumed without loss in generality that $\det B_K > 0$. Now let $\hat{p}_K(\widehat{x}) = p_K(x)$, where $x = B_K\hat{x} + b_K$. Let $B_K = (b_{ij})$, so that

$$x_j = \sum_{l=1}^{n} b_{jl}\hat{x}_l + b_{K,j}.$$

Then

$$\frac{\partial \hat{p}_K}{\partial \hat{x}_i} = \sum_{j=1}^{n}\frac{\partial p_K(x)}{\partial x_j}\frac{\partial x_j}{\partial \hat{x}_i} = \sum_{j=1}^{n}\frac{\partial p_K(x)}{\partial x_j}b_{ji}.$$

Thus is

$$\hat{D} = \left(\frac{\partial \hat{p}_K(\hat{x})}{\partial \hat{x}_1},\dots,\frac{\partial \hat{p}_K(\hat{x})}{\partial \hat{x}_n}\right) \quad \text{and} \quad D = \left(\frac{\partial p_K(x)}{\partial x_1},\dots,\frac{\partial p_K(x)}{\partial x_n}\right),$$

we have $\hat{D} = DB_K$. Hence $\|\hat{D}\|^2 \le \|D\|^2\|B_K\|^2$. Thus,

$$\sum_{i=1}^{n}\left(\frac{\partial p_K}{\partial x_i}(b_{l,K})\right)^2 \ge \|B_K\|^{-2}\sum_{i=1}^{n}\left(\frac{\partial \hat{p}_K}{\partial \hat{x}_i}(\hat{b}_1)\right)^2. \tag{8.24}$$

85 Now suppose $\hat{p} \in \hat{P}$ is such that

$$\sum_{l=1}^{L} \hat{\omega}_1 \sum_{i=1}^{n} \left(\frac{\partial \hat{p}}{\partial \hat{x}_i} (\hat{b}_1) \right)^2 = 0. \tag{8.25}$$

Then since $\hat{\omega}_1 > 0$, we have $\dfrac{\partial \hat{p}}{\partial \hat{x}_i}(\hat{b}_l) = 0$ for all $1 \leq l \leq L$ and $1 \leq i \leq n$. Since $\hat{P} \subset P_{k'}$, we have $\dfrac{\partial \hat{p}}{\partial \hat{x}_i} \in P_{k'-1}$ and hence $\dfrac{\partial \hat{p}}{\partial \hat{x}_i} = 0$ by the $P_{k'-1}$-unisolvency. Thus $\hat{p} \in P_0$, and on the finite dimensional space $\hat{P}/P_0$ (in practice we always have $P_0 \subset \hat{P}$) we have a norm defined by the square-root of the left hand side of (8.25). By the finite dimensionality this is equivalent to the norm defined by the $|\cdot|_{1,\hat{K}}$ norm on P. Hence we have a constant $\hat{\beta} > 0$ such that

$$\sum_{l=1}^{L} \hat{\omega}_l \sum_{i=1}^{n} \left(\frac{\partial \hat{p}}{\partial \hat{x}_i} (\hat{b}_l) \right)^2 \geq \hat{\beta} |\hat{p}|^2_{l,\hat{K}}. \tag{8.26}$$

We will apply this to $\hat{p}_K$. We also have

$$|\hat{p}_K|^2_{l,K} \geq C \|B_K^{-1}\|^{-2} (\det B_K)^{-1} |p_K|^2_{l,K}, \tag{8.27}$$

by Theorem 6.4. Combining the inequalities (8.23), (8.24), (8.26) and (8.27), we get

$$\begin{aligned}
a_h(v_h, v_h) &\geq \alpha \sum_{K \in t_h} (\det B_K) \|B_K\|^{-2} \hat{\beta} C \|B_K^{-1}\|^{-2} (\det B_K)^{-1} |p_K|^2_{l,K} \\
&= \alpha \hat{\beta} C \sum_{K \in t_h} (\|B_K\| \, \|B_K^{-1}\|)^{-2} |p_K|^2_{l,K} \\
&\geq \alpha \hat{\beta} C \gamma \sum_{K \in t_h} |p_K|^2_{l,K} \\
&= \alpha \hat{\beta} C \gamma |v_h|^2_{1,\Omega} = \tilde{\alpha} |v_h|^2_{l,\Omega}
\end{aligned}$$

since $(\|B_K\| \, \|B_K^{-1}\|)^{-2} \geq \gamma$ by Theorem 6.5. This proves the theorem. □

Let us now review our Examples 8.1 through 8.3 to see if the conditions of Theorem 8.2 are satisfied. 86

Let $n = 2$ and consider Example 8.1. Clearly $\hat{\omega} = \text{meas}\,\hat{K} > 0$. Also $\hat{P} = P_1$ for triangles of type (1). Since $\sum_K = \{p(\hat{a})\}$ is P_0-unisolvent, we have that for triangles of type (1) and the quadrature scheme of Example 8.1 the corresponding $a_h(\cdot,\cdot)$ are V_h-elliptic uniformly with respect to h.

For triangles of type (2), $\hat{P} = P_2$. The weights $\hat{\omega}_1$ are all > 0 in Example 8.2. Further we saw in Exercise 5.1 that $\{\hat{a}_{ij}\}_{1\leq i<j\leq 3}$ is P_1-unisolvent. Hence Theorem 8.2 is valid for this quadrature scheme as well.

For triangles of type (3), consider the quadrature scheme of Example 8.3. We have $\hat{\omega}_1 > 0$ and $\hat{P} = P_3$. It was seen in Example 4.2 that the set $\{a_i; 1 \leq i \leq 3\} \cup \{a_{ij}; 1 \leq i < j \leq 3\}$ is P_2-unisolvent. Hence the corresponding bilinear forms $a_h(\cdot,\cdot)$ are V_h-elliptic uniformly with respect to h.

Exercise 8.4. Let $(H, |\cdot|)$ be a Hilbert space and V a subspace with norm $\|\cdot\|$ such that $V \hookrightarrow H$ and $\overline{V} = H$ cf. Sec. 7. Then with the usual notations show that

$$|u - u_h^*| \leq \sup_{g\in H}\Big\{\frac{1}{|g|}\inf_{\varphi_h\in V_h}(M\|u - u_h^*\|\,\|\varphi - \varphi_h\| + |a(u_h^*, \varphi_h) - a_h(u_h^*, \varphi_h)| + |f(\varphi_h) - f_h(\varphi_h)|)\Big\}$$

where φ is the solution of adjoint problem for g.

We now turn our attention to the evaluation of the bound for $\|u - u_h^*\|$ given by (8.21). For second-order problems, for which the norm is $\|\cdot\|_{1,\Omega}$, we will take as usual for $v_h \in V_h$ the element $\pi_h u \in V_h$ so that we now get the bound

$$\|u - u_h^*\|_{1,\Omega} \leq C\Bigg[\|u - \pi_h u\|_{1,\Omega} + \sup_{w_h\in V_h}\frac{|a(\pi_h u, w_h) - a_h(\pi_h u, w_h)|}{\|w_h\|_{1,\Omega}} + \sup_{w_h\in V_h}\frac{|f(w_h) - f_h(w_h)|}{\|w_h\|_{1,\Omega}}\Bigg]. \tag{8.28}$$

87 Let us assume that we may apply Theorem 7.1, so that

$$\|u - \pi_h u\|_{1,\Omega} \leq C\, h^k |u|_{k+1,\Omega}. \tag{8.29}$$

In order to keep the same accuracy, we will therefore try to obtain estimates of the following form:

$$\begin{cases} \sup\limits_{w_h \in V_h} \dfrac{|a(\pi_h u, w_h) - a_h(\pi_h u, w_h)|}{\|w_h\|_{1,\Omega}} \leq C(u) h^k, \\ \sup\limits_{w_h \in V_h} \dfrac{|f(w_h) - f_h(w_h)|}{\|w_h\|_{1,\Omega}} \leq C(f) h^k, \end{cases} \tag{8.30}$$

and these will in turn be obtained from "local" estimates. (cf. Theorem 8.4 and Exercise 8.5).

As a preliminary step, we need two results which we prove now.

The first of these is a historically important result in the interpolation theory in Sobolev spaces.

Theorem 8.3 (BRAMBLE-HILBERT LEMMA; cf. Bramble and Hilbert [27]). *Let $\Omega \subset \mathbb{R}^n$ be open with Lipschitz continuous boundary Γ. Let $f \in W^{k+1,p}(\Omega))$ which vanishes over P_k. Then there exists a constant $C = C(\Omega)$ such that, for all $v \in W^{k+1,p}(\Omega)$,*

$$|f(v)| \leq C \|f\|^*_{k+1,p,\Omega} |v|_{k+1,p,\Omega}. \tag{8.31}$$

Proof. For $v \in W^{k+1,p}(\Omega)$ and all $p \in P_k$, we have

$$f(v) = f(v + p),$$

so that

$$|f(v)| = |f(v + p)| \leq \|f\|^*_{k+1,p,\Omega} \|v + p\|_{k+1,p,\Omega},$$

and thus, **88**

$$\begin{aligned} |f(v)| &\leq \|f\|^*_{k+1,p,\Omega} \inf_{p \in P_k} \|v + p\|_{k+1,p,\Omega} \\ &\leq C \|f\|^*_{k+1,p,\Omega} |v|_{k+1,p,\Omega}, \end{aligned}$$

by Theorem 6.2, which completes the proof. □

Lemma 8.1. *Let $\varphi \in W^{m,q}(\Omega)$, $w \in W^{m,\infty}(\Omega)$. Then $\varphi w \in W^{m,q}(\Omega)$, and there exists a numerical constant C, independent of φ and w such that*

$$|\varphi w|_{m,q,\Omega} \leq C \sum_{j=0}^{M} |\varphi|_{m-j,q,\Omega} |w|_{j,\infty,\Omega}. \tag{8.32}$$

Proof. The result is an immediate consequence of the Leibniz formula: For any $|\alpha| = m$,

$$\partial^{\alpha}(\varphi w) = \sum_{j=0}^{m} \sum_{|\beta|=j} C_{\alpha,\beta} \partial^{\alpha-\beta}(\varphi) \partial^{\beta}(w)$$

which yields (8.32). □

We may now apply Lemma 8.1 and Theorem 8.3 to get the estimates (8.30). We do this in two stages (Theorem 8.4 and Exercise 8.5) in which, for the sake of simplicity, we present our results for the special case $P_K = P_2$.

Theorem 8.4. *Let $P_K = P_2$ and consider a quadrature scheme such that for all $\hat{\varphi} \in P_2$, $\hat{\xi}(\hat{\varphi}) = 0$. Then there exists a constant C, independent of K, such that for all $a_{ij} \in W^{2,\infty}(K)$ and for all p, $p' \in P_K$ we have*

$$\left|\mathscr{E}_K\left[(a_{ij})\frac{\partial p}{\partial x_j}\frac{\partial p'}{\partial x_i}\right]\right| \leq C\, h_K^2 \|a_{ij}\|_{2,\infty,K} \left\|\frac{\partial p}{\partial x_j}\right\|_{1,K} \left|\frac{\partial p'}{\partial x_i}\right|_{0,K}. \tag{8.33}$$

Proof. Since we have $\dfrac{\partial p}{\partial x_j}, \dfrac{\partial p'}{\partial x_i} \in P_1$, it suffices to find a suitable estimate for $\mathscr{E}_K(avw)$, for $a \in W^{2,\infty}(K)$, v, $w \in P_1$. Further, since

$$\mathscr{E}_K(avw) = (\det B_K)\hat{\mathscr{E}}(\hat{a}\hat{v}\hat{w}), \tag{8.34}$$

89 we will first find an estimate for $\hat{\mathscr{E}}(\hat{a}\hat{v}\hat{w})$, with $\hat{a} \in W^{2,\infty}(\hat{K})$ and $\hat{v}$, $\hat{w} \in P_1$. Let $\hat{\pi}_0\hat{w}$ be the orthogonal projection of $\hat{w}$ onto the subspace P_0 in the sense of $L^2(\hat{K})$. Then we may write

$$\hat{\mathscr{E}}(\hat{a}\hat{v}\hat{w}) = \hat{\mathscr{E}}(\hat{a}\hat{v}\hat{\pi}_0\hat{w}) + \hat{\mathscr{E}}(\hat{a}\hat{v}(\hat{w} - \hat{\pi}_0\hat{w})). \tag{8.35}$$

(i) *Estimate for* $\hat{\mathscr{E}}(\hat{a}\hat{v}\hat{\pi}_0\hat{w})$.

Consider the functional $\hat{\mathscr{E}} : W^{2,\infty}(\hat{K}) \to \mathbb{R}$ defined by

$$\hat{\psi} \mapsto \hat{\mathscr{E}}(\hat{\psi}) = \int_{\hat{K}} \hat{\psi}(\hat{x})d\hat{x} = \sum_{l=1}^{L} \hat{\omega}_1\hat{\psi}(\hat{b}_1).$$

$|\hat{\mathscr{E}}(\hat{\psi})| \leq \hat{C}|\hat{\psi}|_{0,\infty,\hat{K}} \leq \hat{C}||\hat{\psi}||_{2,\infty,\hat{K}}$. Thus $\hat{\mathscr{E}}$ is a continuous linear functional on $W^{2,\infty}(\hat{K})$. Hence by Theorem 8.3, since $\hat{\mathscr{E}}$ vanishes on $P_1(\subset P_2)$[1], we have a constant $\hat{C}$ such that

$$|\hat{\mathscr{E}}(\hat{\psi})| \leq \hat{C}|\hat{\psi}|_{2,\infty,\hat{K}}. \tag{8.36}$$

Thus

$$\begin{aligned}|\hat{\mathscr{E}}(\hat{a}\hat{v}\hat{\pi}_0\hat{w})| &\leq \hat{C}|\hat{a}\hat{v}\hat{\pi}_0\hat{w}|_{2,\infty,\hat{K}}\\ &\leq \hat{C}|\hat{a}\hat{v}|_{2,\infty,\hat{K}}|\hat{\pi}_0\hat{w}|_{0,\infty,\hat{K}}\end{aligned}$$

since $\hat{\pi}_0\hat{w} \in P_0$ is a constant function. By Lemma 8.1 (recall that $\hat{v} \in P_1$),

$$|\hat{\mathscr{E}}(\hat{a}\hat{v}\hat{\pi}_0\hat{w})| \leq \hat{C}|\hat{\pi}_0\hat{w}|_{0,\infty,\hat{K}}\left[|\hat{a}|_{1,\infty,\hat{K}}|\hat{v}|_{1,\infty,\hat{K}} + |\hat{a}|_{2,\infty,\hat{K}}|\hat{v}|_{0,\infty,\hat{K}}\right].$$

By the equivalence of the L^2 and L^∞ norms on P_0, and since the projection has norm less than that of the vector itself in any Hilbert space we have the chain of inequalities

$$|\hat{\pi}_0\hat{w}|_{0,\infty,\hat{K}} \leq \hat{C}|\hat{\pi}_0\hat{w}|_{0,\hat{K}} \leq \hat{C}|\hat{w}|_{0,\hat{K}}.$$

90

Similarly we may replace $|\hat{v}|_{1,\infty,\hat{K}}$ by $|\hat{v}|_{1,\hat{K}}$ and $|\hat{v}|_{0,\infty,\hat{K}}$ by $|\hat{v}|_{0,\hat{K}}$ since the L^2 and L^∞ norms are equivalent on P_1. Thus we get

$$|\hat{\mathscr{E}}(\hat{a}\hat{v}\hat{\pi}_0\hat{w})| \leq \hat{C}(|\hat{a}|_{1,\infty,\hat{K}}|\hat{v}|_{1,\hat{K}} + |\hat{a}|_{2,\infty,\hat{K}}|\hat{v}|_{0,\hat{K}})|\hat{w}|_{0,\hat{K}}. \tag{8.37}$$

[1]In this estimate, we do not use the "full" polynomial invariance of the quadrature scheme.

(ii) *Estimate for* $\hat{\mathscr{E}}(\hat{a}\hat{v}(\hat{w}-\hat{\pi}_0\hat{w}))$.

Let $\hat{w}\in P_1$ be fixed and let $\hat{\varphi}\in W^{2,\infty}(\hat{K})$. Then

$$\begin{aligned}|\hat{\mathscr{E}}(\hat{\varphi}(\hat{w}-\hat{\pi}_0\hat{w}))| &\leq \hat{C}|\hat{\varphi}(\hat{w}-\hat{\pi}_0\hat{w})|_{0,\infty,\hat{K}}\\ &\leq \hat{C}|\hat{\varphi}|_{0,\infty,\hat{K}}|\hat{w}-\hat{\pi}_0\hat{w}|_{0,\infty,\hat{K}}\\ &\leq \hat{C}|\hat{w}-\hat{\pi}_0\hat{w}|_{0,\infty,\hat{K}}||\hat{\varphi}||_{2,\infty,\hat{K}}.\end{aligned}$$

Thus the functional on $W^{2,\infty}(\hat{K})$ defined by $\hat{\varphi}\mapsto\hat{\mathscr{E}}(\hat{\varphi}(\hat{w}-\hat{\pi}_0\hat{w}))$ is continuous, linear with norm $\leq \hat{C}|\hat{w}-\hat{\pi}_0\hat{w}|_{0,\infty,\hat{K}}$. Since for $\hat{\varphi}\in P_1$, $\hat{\varphi}(\hat{w}-\hat{\pi}_0\hat{w})\in P_2$, we have that the functional vanishes on P_1. By Theorem 8.3,

$$|\hat{\mathscr{E}}(\hat{\varphi}(\hat{w}-\hat{\pi}_0\hat{w}))| \leq \hat{C}|\hat{w}-\hat{\pi}_0\hat{w}|_{0,\infty,\hat{K}}|\hat{\varphi}|_{2,\infty,\hat{K}}.$$

Set $\hat{\varphi}=\hat{a}\hat{v}$. Now,

$$|\hat{a}\hat{v}|_{2,\infty,\hat{K}} \leq \hat{C}(|\hat{a}|_{2,\infty,\hat{K}}|\hat{v}|_{0,\infty,\hat{K}}+|\hat{a}|_{1,\infty,\hat{K}}|\hat{v}|_{1,\infty,\hat{K}}).$$

Again we may use the equivalence between the L^∞-norms of $\hat{v}$ and $\hat{w}-\hat{\pi}_0\hat{w}$ and the L^2-norms of the same functions as in (i) since they belong to the finite dimensional space P_1. Also, by the triangle inequality,

$$|\hat{w}-\hat{\pi}_0\hat{w}|_{0,\hat{K}} \leq \hat{C}|\hat{w}|_{0,\hat{K}}.$$

Thus we get

$$|\hat{\mathscr{E}}(\hat{a}\hat{v}(\hat{w}-\hat{\pi}_0\hat{w}))| \leq \hat{C}(|\hat{a}|_{2,\infty,\hat{K}}|\hat{v}|_{0,\hat{K}}+|\hat{a}|_{1,\infty,\hat{K}}|\hat{v}|_{1,\hat{K}})|\hat{w}|_{0,\hat{K}}. \tag{8.38}$$

91

(iii) We can now complete the proof. Recall that $\mathscr{E}_K(avw)=(\det B_K)\,\hat{\mathscr{E}}(\hat{a}\hat{v}\hat{w})$. Also,

$$\begin{aligned}&|\hat{a}|_{m,\infty,\hat{K}} \leq C\,h_K^m|a|_{m,\infty,K}\\ &|\hat{v}|_{2-m,\hat{K}} \leq C\,h_K^{2-m}(\det B_K)^{-\frac{1}{2}}|v|_{2-m,K}.\\ &|\hat{w}|_{0,\hat{K}} \leq C(\det B_K)^{-\frac{1}{2}}|w|_{0,K},\end{aligned} \tag{8.39}$$

by Theorems 6.4 and 6.5. Combining (8.37), (8.38) and (8.39), we get

$$\begin{aligned}|\mathscr{E}_K(avw)| &\leq C\, h_K^2(|a|_{1,\infty,K}|v|_{1,K} + |a|_{2,\infty,K}|v|_{0,K})|w|_{0,K}\\ &\leq C\, h_K^2\|a\|_{2,\infty,K}\|v\|_{1,K}|w|_{0,K}.\end{aligned}$$

Setting $a = A_{ij}$, $v = \dfrac{\partial p}{\partial x_j}$, $w = \dfrac{\partial p'}{\partial x_i}$ we obtain (8.33), thus completing the proof. □

We leave the second stage as an exercise:

Exercise 8.5. Let $P_K = P_2$ and let the quadrature scheme be such that $\hat{\mathscr{E}}(\hat{\varphi}) = 0$ for all $\hat{\varphi} \in P_2$. Then show that for q such that $W^{2,q}(K) \hookrightarrow C^0(K)$, there exists C independent of K such that for all $f \in W^{2,q}(K)$ and all $p \in P_K$,

$$|\mathscr{E}_K(fp)| \leq C\, h_K^2(\det B_K)^{\frac{1}{2}-\frac{1}{q}}\|f\|_{2,q,K}\|p\|_{1,K}.$$

[*Hint:* If $\hat{\pi}_1$ is the orthogonal projection to P_1 in the L^2-sense then write

$$\hat{\mathscr{E}}(\hat{f}\hat{p}) = \mathscr{E}(\hat{f}\hat{\pi}_1\hat{p}) + \hat{\mathscr{E}}(\hat{f}(\hat{p} - \hat{\pi}_1\hat{p}))].$$

Remark 8.4. The inclusion $W^{2,q}(K) \hookrightarrow C^0(K)$ is true if, for instance, $2 - \frac{n}{q} > 0$, by the Sobolev imbedding theorem.

We now come to the final stage in the estimation of $\|u - u_h^*\|$.

Theorem 8.5. *Let* (t_h) *be a regular family of triangulations on* Ω *by n-simplices of type (2). Let us assume that the* V_h*-ellipticity is uniform with respect to h. Let* $\hat{\mathscr{E}}(\hat{\varphi}) = 0$ *for all* $\hat{\varphi} \in P_2$. *Then if* $u \in H^3(\Omega) \hookrightarrow C^0(\overline{\Omega})(n \leq 5)$, $a_{ij} \in W^{2,\infty}(\Omega)$ *and* $f \in W^{2,q}(\Omega)$ *for some* $q \geq 2$, *we have the estimate* **92**

$$\|u - u_h^*\|_{1,\Omega} \leq C\, h^2[\|u\|_{3,\Omega} + \|f\|_{2,q,\Omega}]. \tag{8.40}$$

Proof. We estimate the various quantities in (8.28). We have:

$|a(\pi_h u, w_h) - a_h(\pi_h u, w_h)| \leq$

$$\leq \sum_{K\in \mathfrak{t}_h} \sum_{i,j=1}^{n} |\mathscr{E}_K(a_{ij} \frac{\partial(\pi_h u|K)}{\partial x_j} \frac{\partial(w_h|K)}{\partial x_i}|$$

$$\leq \sum_{K\in \mathfrak{t}_h} \sum_{i,j=1}^{n} C\, h_K^2 \|a_{ij}\|_{2,\infty,K} \|\frac{\partial(\pi_h u|_K)}{\partial x_j}\|_{1,K} |\frac{\partial(w_h|K)}{\partial x_i}|_{0,K}$$

$$\leq h^2 \sum_{i,j=1}^{n} \|a_{ij}\|_{2,\infty,\Omega} \left(\sum_{K\in \mathfrak{t}_h} \|\frac{\partial(\pi_h u|K)}{\partial x_j}\|^2_{1,K}\right)^{\frac{1}{2}} \left(\sum_{K\in \mathfrak{t}_h} \|\frac{\partial(w_h|K)}{\partial x_i}\|^2_{0,K}\right)^{\frac{1}{2}}$$

(since $h_K \leq h$, and we may apply the Cauchy-Schwarz inequality)

$$\leq C\, h^2 \|\pi_h u\|_{2,\Omega} \|w_h\|_{1,\Omega}$$

Now,

$$\|\pi_h u\|_{2,\Omega} \leq \|u\|_{2,\Omega} + \|u - \pi_h u\|_{2,\Omega} \leq C\|u\|_{2,\Omega},$$

using Theorem 6.3 with $P_1 \subset P_K = P_2$. Therefore, for all $w_h \in V_h$, we have

(8.41) $$\frac{|a(\pi_h u, w_h) - a_h(\pi_h u, w_h)|}{\|w_h\|_{1,\Omega}} \leq Ch^2 \|u\|_{2,\Omega}.$$

Similarly, we have

$$|f(w_h) - f_h(w_h)| \leq \sum_{K\in \mathfrak{t}_h} |\mathscr{E}_K(f w_h|_K)|$$

$$\leq \sum_{K\in \mathfrak{t}_h} C\, h_K^2 (\text{meas}\, K)^{\frac{1}{2}-\frac{1}{q}} \|f\|_{2,q,K} \|w_h\|_{1,K}.$$

93

Since $q \geq 2$, $\frac{1}{2} - \dfrac{1}{q} \geq 0$ and by the general Hölder's inequality,

$$\sum_{K\in \mathfrak{t}_h} (\text{meas}\, K)^{\frac{1}{2}-\frac{1}{q}} \|f\|_{2,q,K} \|w_h\|_{1,K}$$

$$\leq \left(\sum_{K\in \mathfrak{t}_h} \text{meas}\, K\right)^{\frac{1}{2}-\frac{1}{q}} \left(\sum_{K\in \mathfrak{t}_h} \|f\|^q_{2,q,K}\right)^{\frac{1}{q}} \left(\sum_{K\in \mathfrak{t}_h} \|w_h\|^2_{1,K}\right)^{\frac{1}{2}}$$

$$= C\|f\|_{2,q,\Omega}\|w_h\|_{1,\Omega}.$$

Hence we get, fot all $w_h \in V_h$,

$$\frac{|f(w_h) - f_h(w_h)|}{\|w_h\|_{1,\Omega}} \leq Ch^2\|f\|_{2,q,\Omega}. \tag{8.42}$$

Combining (8.28), (8.29), (8.41) and (8.42) we get (8.40), thus completing the proof. □

Remark 8.5. The condition $n \leq 5$ (needed for the continuous inclusion $H^3(\Omega) \hookrightarrow C^0(\overline{\Omega})$) was already necessary for the definition of $\pi_h u$.

References: For a survey on numerical quadrature in general one may refer to Haber's survey article [13]. For application to the finite element method, the Sec. 4.2 of the book by Strang and Fix [22] or Chapter 2 of the forthcoming book of Ciarlet and Raviart [5].

Chapter 9

The Obstacle Problem

In Sec. 2 we cited the Obstacle Problem as an example of a non-linear 94
abstract problem of Sec. 1. Let us recall a few facts about this to start with.

Consider an elastic membrane (cf. Fig. 9.1) stretched over an open set $\Omega \subset \mathbb{R}^2$ and fixed along the boundary Γ which is assumed to be Lipschitz continuous. Let a force of density $F dx$ act on the membrane. Let us assume the existence of an obstacle given by $\chi(x)$, for $x \in \Omega$. Then vertical displacement given by u is the solution of the abstract problem where

$$\begin{cases} a(u, v) = \int_\Omega \sum_{i=1}^{2} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} dx \\ f(v) = \int_\Omega f v \, dx, \ f \in L^2(\Omega) \end{cases} \tag{9.1}$$

for $u, v \in V = H_0^1(\Omega)$, where $f = F/t$, t being the tension. The subset K is given by

$$K = \{v \in H_0^1(\Omega);\ v \geq \chi \ \text{ a.e. } \ in \Omega\}.$$

If v_1, v_2 are in K and $v_i < \chi$ in A_i with meas $A_i = 0$ for $i = 1, 2$, then $\lambda v_1 + (1 - \lambda) v_2 \geq \chi$ on $(A_1 \cup A_2)^c$ i.e. the complement of $A_1 \cup A_2$ and meas $(A_1 \cup A_2) = 0$. Thus K is convex. If $v \in \overline{K}$, let $v_n \in K$ such that $v_n \to v$ in $H_0^1(\Omega)$. Let $v_n \geq \chi$ in A_n^c, meas $A_n = 0$. Then all the v_n are $\geq \chi$ on $(\cup_n A_n)^c$ and meas $(\cup_n A_n) = 0$. Hence $v \geq \chi$ a.e. as well. Thus

$v \in K$ and K is closed as well. We have the regularity assumption that $\chi \in H^2(\Omega)$. Of course it is

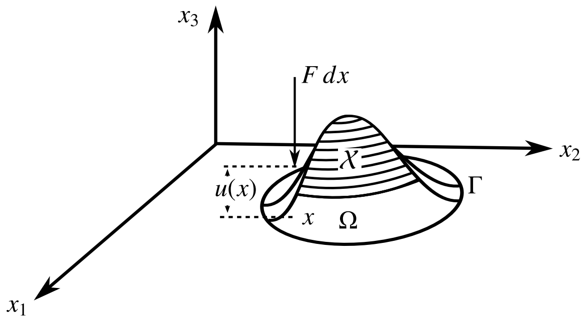

Figure 9.1:

95 The solution u satisfies

$$J(u) = \min_{v \in K} J(v), \tag{9.2}$$

where $J(v) = \frac{1}{2}a(v, v) - f(v)$ and is also characterized by the variational inequalities (cf. Theorem 1.1):

$$a(u, v - u) \geq f(v - u), \quad \text{for all} \quad v \in K. \tag{9.3}$$

We proposed as a problem to show that this problem is interpreted as the following classical problem (assuming $u \in H_0^1(\Omega) \cap H^2(\Omega)$).

$$\begin{cases} u \geq \chi \text{ in } \Omega, \\ -\Delta u = f \text{ where } u > \chi, \\ u = 0 \text{ on } \Gamma. \end{cases} \tag{9.4}$$

We have a few regularity results which are listed below:

(i) If Ω is convex and Γ is a C^2-boundary then $u \in H_0^1(\Omega) \cap H^2(\Omega)$.

(ii) If $f = 0$ and Ω a convex polygon then also $u \in H_0^1(\Omega) \cap H^2(\Omega)$.

(iii) The norm $||u||_{2,\Omega}$ is bounded above by a function of $||f||_{0,\Omega}$ and
96 $||\chi||_{2,\Omega}$ in cases (i) and (ii).

Our aim in this section is to use the finite element method to approximate this problem and obtain error estimates. We list our assumptions now:

Let Ω be a convex polygon, $f \in L^2(\Omega)$, $\chi \in H^2(\Omega)$ and let $u \in H^2(\Omega) \cap H_0^1(\Omega)$.

Remark 9.1. We cannot assume any more smoothness on u other than $H^2(\Omega)$. For instance in the 1-dimensional case if $f = 0$, and even if the function is very smooth the points of contact of u with χ will have discontinuous second derivatives in general; cf. Fig. 9.2.

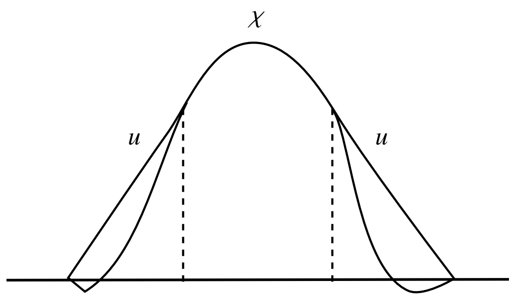

Figure 9.2:

With the above assumptions we proceed to the approximate problems, first in the abstract setting, as usual.

We have the problems (P_h) associated with the subspaces $V_h \subset V = H_0^1(\Omega)$. We now choose closed convex subsets $K_h \subset V_h$. One has to bear in mind that, in general, $K_h \not\subset K$ (we will see that this is the case in our approach, sub-sequently).

We find $u_h \in K_h$ such that for all $v_h \in K_h$,

$$a(u_h, v_h - u_h) \geq f(v_h - u_h). \tag{9.5}$$

The existence and uniqueness of the u_h follow from Theorem 1.1.

Let H be a Hilbert space with norm $|\cdot|$ and inner-product $(\cdot,\cdot)$. Let $(V, \|\cdot\|)$ be a subspace such that $V \hookrightarrow H$, $\overline{V} = H$. Then, as usual, if we identify H' and H, then H will be identified with a subspace of V'. (We **97** will take $V = H^1_0(\Omega)$ and $H = L^2(\Omega)$). Also as in Sec. 1 (cf. proof of Theorem 1.2), for all $u, v \in V$ we have

$$a(u, v) = (Au)(v), \tag{9.6}$$

where $A : V \to V'$ is a linear map. We now pass on to an abstract error bound.

Theorem 9.1 (FALK). *Assume that $f \in H$, $Au \in H$. Then there exists a constant C, independent of V_h and K_h, such that*

$$\|u - u_h\| \leq C \left[\inf_{v_h \in K_h} (\|u - v_h\|^2 + |u - v_h|) + \inf_{v \in K} |u_h - v| \right]^{\frac{1}{2}}. \tag{9.7}$$

(*Note:* The condition $Au \in H = L^2(\Omega)$ is satisfied if $u \in H^2(\Omega)$ since $Au = -\Delta u \in L^2(\Omega)$.)

Proof. Let α stand for the V_h-ellipticity constant. Then

$$\begin{aligned} \alpha\|u - u_h\|^2 &\leq a(u - u_h, u - u_h) \\ &= a(u, u) + a(u_h, u_h) - a(u, u_h) - a(u_h, u). \end{aligned} \tag{9.8}$$

For any $v \in K$ and any $v_h \in K_h$, by (9.3) and (9.5), we have

$$\begin{cases} a(u, u) \leq a(u, v) + f(u - v), \\ a(u_h, u_h) \leq a(u_h, v_h) + f(u_h - v_h). \end{cases} \tag{9.9}$$

Substituting in (9.8) we get

$$\begin{aligned} \alpha\|u - u_h\|^2 &\leq a(u, v) + f(u - v) + a(u_h, v_h) \\ &\qquad + f(u_h - v_n) - a(u, u_h) - a(u_h, u) \\ &= a(u, v - u_h) - f(v - u_h) + a(u, v_h - u) - f(v_h - u) \\ &\qquad + a(u_h - u, v_h - u) \\ &= (f - Au, u - v_h) + (f - Au, u_h - v) + a(u_h - u, v_h - u) \\ &\leq |f - Au|\,|u - v_h| + |f - Au|\,|u_h - v| + M\|u_h - u\|\,\|v_h - u\|. \end{aligned}$$

Notice that since $(\sqrt{\dfrac{\alpha}{M}}||u-u_h|| - \sqrt{\dfrac{M}{\alpha}}||u-v_h||)^2 \geq 0$, we have

$$||u-u_h||\,||u-v_h|| \leq \frac{1}{2}\left(\frac{\alpha}{M}||u-u_h||^2 + \frac{M}{\alpha}||u-v_h||^2\right),$$

and hence

$$\alpha||u-u_h||^2 \leq C[|u-v_h| + |u_h - v|] + \frac{\alpha}{2}||u-u_h||^2 + \frac{M^2}{2\alpha}||u-v_h||^2$$

Or,

$$||u-u_h||^2 \leq C[(|u-v_h| + ||u-v_h||^2) + |u_h - v|].$$

Varying $v_h \in K_h$ and $v \in K$ and extracting the square root after taking the infima we get (9.7). This completes the proof. □

Remark 9.2. If we have a linear problem then $f = Au$ gives the solution and we get the original bound (3.1).

Remark 9.3. From (9.8) we see that this estimate holds even if $a(\cdot,\cdot)$ is not symmetric.

We now apply this to the specific membrane problem. Maintaining our assumptions on Ω, let t_h be a triangulation by triangles of type (1), and let V_h be the corresponding subspace of $V = H_0^1(\Omega)$.

Remark 9.4. *It is of no practical use if we go to more sophisticated finite elements, unlike the linear problem.* Since $u \in H^2(\Omega)$ is the maximum smoothness, we may atmost use our abstract estimate theorems only on the spaces P_1.

One may be tempted to try for K_h those v_h which are $\geq \chi$ a.e. in Ω. However this is not of value from *numerical and computational points of view* for we do not easily know where exactly our piecewise linear **99**
solution functions would touch χ. We set instead

(9.10) $$K_h = \{v_h \in V_h;\ \text{At all nodes } b \text{ of } \mathrm{t}_h, v_h(b) \geq \chi(b)\}.$$

Remark 9.5.

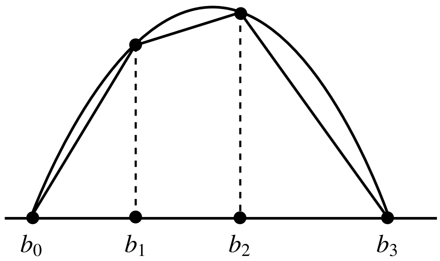

Figure 9.3:

As seen in Fig. 9.3, though for nodes b, $v_h(b) \geq \chi(b)$, it does not guarantee that $v_h \geq \chi$ a.e. Thus we see that $K_h \not\subset K$. Now the relation (9.10) is very easy to implement using the computer.

We now have our main result on the error bound.

Theorem 9.2 (FALK)**.** *There exists a constant C depending on $\|f\|_{0,\Omega}$ and $\|\chi\|_{2,\Omega}$ such that for a regular family of triangulations* (t_h) *as above we have*

$$\|u - u_h\|_{1,\Omega} \leq C\, h. \tag{9.11}$$

Remark 9.6. The order of convergence is therefore the same as that for the linear problems when we use piecewise linear approximations.

Proof. By Theorem 9.1,

$$\|u - u_h\|_{1,\Omega} \leq C \left[\inf_{v_h \in K_h} \left(\|u - v_h\|^2_{1,\Omega} + |u - v_h|_{0,\Omega} \right) + \inf_{v \in K} |u_h - v|_{0,\Omega} \right]^{\frac{1}{2}}.$$

100 (i) We first estimate the infimum over K_h. Note that if $v_h = \pi_h u$, then $v_h \in V_h$. Also for all nodes b, $v_h(b) = \pi_h u(b) = u(b) \geq \chi(b)$. Thus $v_h \in K_h$ as well. Thus,

$$\begin{aligned} \inf_{v_h \in K_h} \left(\|u - v_h\|^2_{1,\Omega} + |u - v_h|_{0,\Omega} \right) &\leq \|u - \pi_h u\|^2_{1,\Omega} + |u - \pi_h u|_{0,\Omega} \\ &\leq C\, h^2 \left(|u|^2_{2,\Omega} + |u|_{2,\Omega} \right). \end{aligned}$$

(ii) For the infimum over K, consider $v_1 = \max(u_h, \chi)$. Clearly $v_1 \geq \chi$ and v belongs to $H^1(\Omega)$ because u_h and χ also belong to $H^1(\Omega)$ (this is a non-trivial result which we assume here). Hence $v_1 \in K$, and thus,

$$\inf_{v\in K} |u_h - v|_{0,\Omega} \leq |u_h - v_1|_{0,\Omega}.$$

We have

$$|u_h - v_1|^2_{0,\Omega} = \int_{\Lambda_h} |u_h - \chi|^2 dx, \quad \text{where } \Lambda_h = \{x; \chi(x) \geq u_h(x)\}.$$

If $\pi_h\chi$ is the V_h-interpolate of χ, then for all nodes b, $u_h(b) \geq \chi(b) = \pi_h\chi(b)$. Since both u_h and $\pi_h\chi$ are piecewise linear, we may now assert that $u_h \geq \pi_h\chi$ everywhere. Thus $u_h - \pi_h\chi \geq 0$ on Ω. Thus for all $x \in \Lambda_h$, we have

$$\begin{aligned} 0 < |(\chi - u_h)(x)| = (\chi - u_h)(x) &\leq (\chi - \pi_h\chi)(x) \\ &\leq |(\chi - \pi_h\chi)(x)| \end{aligned}$$

and for $x \in \Omega - \Lambda_h$, $(\chi - u_h)(x) = 0$, so that

$$|u_h - v_1|_{0,\Omega} \leq \left(\int_{\Omega} |\chi - \pi_h\chi|^2 dx \right)^{\frac{1}{2}} = |\chi - \pi_h\chi|_{=0,\Omega} \leq C\, h^2 |\chi|_{2,\Omega}.$$

Hence

$$||u - u_h||_{1,\Omega} \leq C(h^2)^{\frac{1}{2}} = Ch,$$

where C depends on $|\chi|_{2,\Omega}$ and $|u|_{2,\Omega}$. However the regularity result (iii) **101**
helps us to bound $|u|_{2,\Omega}$ above by a constant C depending on $|f|_{0,\Omega}$ and $||\chi||_{2,\Omega}$ which completes the proof of the theorem. ◻

References: Two important references are Falk [11] and [12]. For regularity results refer Brezis and Stampacchia [3] and Lewy and Stampacchia [16].

Another references is Mosco and Strang [19].

Chapter 10

Conforming Finite Element Method for the Plate Problem

In Sec. 2 (cf. Example 2.4), as an example of fourth-order problem, we 102
described *the plate problem*. In abstract terms it is to find the solution of

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

where

(10.2)

$$\begin{cases} K = V = H_0^2(\Omega),\ \Omega \subset \mathbb{R}^2, \\ a(u,v) = \int\limits_\Omega \left(\Delta u \cdot \Delta v + (1-\sigma) \left\{ 2 \dfrac{\partial^2 u}{\partial x_1 \partial x_2} \dfrac{\partial^2 v}{\partial x_1 \partial x_2} - \dfrac{\partial^2 u}{\partial x_1^2} \dfrac{\partial^2 v}{\partial x_2^2} - \dfrac{\partial^2 u}{\partial x_2^2} \dfrac{\partial^2 v}{\partial x_2^2} \right\} \right) dx, \\ f(v) = \int\limits_\Omega f v \, dx,\ f \in L^2(\Omega). \end{cases}$$

The problem was interpreted as the classical boundary value problem

$$\begin{cases} \Delta^2 u = f \ \text{ in } \ \Omega, \\ u = \dfrac{\partial u}{\partial \nu} = 0 \ \text{ on } \ \Gamma = \partial\Omega. \end{cases} \tag{10.3}$$

Remark 10.1. It was commented in Sec. 2 that the second term of the integrand in the definition of $a(\cdot,\cdot)$ does not contribute to the differential equation. Our method here will be equally applicable to both the cases. viz. with the second term present (the plate problem) or with that term absent (as it happens in Hydro-dynamics). In our next section, on non-conforming methods, we will see that the second term is essential in order that we may apply that method.

We assume that Ω is a polygonal domain in $\mathbb{R}^2$. We saw in Sec. 3 that for fourth-order problems we need the inclusion $V_h \subset C^1(\overline{\Omega})$. (cf. Exercise 3.1). Thus we need to use finite elements of class C^1, such as the Argyris triangle, the Bogner-Fox-Schmidt rectangle and so on (cf. Sec. 4).

103 *When such finite elements can be imbedded in an* affine family, then we have the approximation theory, for regular families of triangulations, available to us. We show that this is the case for the Bogner-Fog-Schmidt rectangle. However for the Argyris triangle or for the 18-degree-of-freedom triangle such an imbedding is not possible and we have to modify the usual argument to obtain error estimates. The "minimal assumptions" for $0(h)$ convergence in the $\|\cdot\|_{2,\Omega}$ norm are that $P_2 \subset P$ and that $u \in H^3(\Omega) \cap H_0^2(\Omega)$. We have that if Ω is a convex polygon and if $f \in L^2(\Omega)$, then $u \in H^3(\Omega) \cap H_0^2(\Omega)$. This result is due to Knodratév.

We will go through the various examples of triangulations of class C^1 and study convergence in these cases.

Example 10.1. *The Bogner-Fog-Schmidt rectangle* (cf. Exercise 4.9). Let $P_K = Q_3(\dim P_K = 16)$. We then have (cf. Fig. 10.1):

$$\sum_K = \left\{ p(a_i), \frac{\partial p}{\partial x_1}(a_i),\ \frac{\partial p}{\partial x_2}(a_i),\ \frac{\partial^2 p}{\partial x_1 \partial x_2}(a_i);\ 1 \le i \le 4 \right\}. \tag{10.4}$$

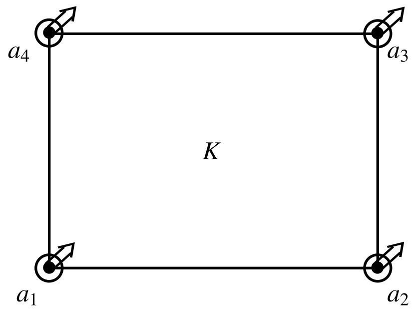

Figure 10.1:

Equivalently, one may also use the P_K-unisolvent set

$$\sum_K{}' = \{p(a_i), Dp(a_i)(a_{i+1} - a_i), Dp(a_i)(a_{i-1} - a_i),$$
$$D^2 p(a_i)(a_{i+1} - a_i, a_{i-1} - a_i);\ 1 \le i \le 4\} \tag{10.5}$$

(all indices being read modulo 4). **104**

Recall that for an affine family of finite elements, the degrees of freedom $p(a_i^0)$, $Dp(a_i^1)(\xi_{ik}^1)$, $(D^2 p(a_i^2)(\xi_{ik}^2, \xi_{il}^2)$ are such that (cf. Sec. 5):

$$\begin{cases} a_i^0 = F(\hat{a}_i^0), \ldots, a_i^2 = F(\hat{a}_i^2), \\ \xi_{i,k}^1 = B_K \hat{\xi}_{i,k}^1, \ldots, \xi_{i,1}^2 = B_K \hat{\xi}_{i,1}^2 \end{cases} \tag{10.6}$$

for then $\widehat{\pi_K v} = \hat{\pi}\hat{v}$ which is essentially what we need for the abstract error analysis.

In $\sum_K'$ note that

$$a_{i+1} - a_i = F(\hat{a}_{i+1}) - F(\hat{a}_i) = B_K(\hat{a}_{i+1} - \hat{a}_i), \tag{10.7}$$

and so on. Thus it is clear that this rectangle can be imbedded in an affine family of finite elements. Now $P_K \subset \hat{P} \subset Q_3$ for $k = 3$. By our abstract error analysis, we therefore have

$$||u - u_h||_{2,\Omega} \le Ch^2 |u|_{4,\Omega}. \tag{10.8}$$

assuming sufficient smoothness on u as usual.

A word about the boundary conditions. As in Exercise 3.1, we get that $V_h \subset H_0^2(\Omega)$ if $v = \dfrac{\partial v}{\partial \nu} = 0$ on Γ, for $v \in V_h$. Thus in choosing our basis functions we must assure ourselves that this condition is satisfied. This in turn depends on the values at the boundary nodes. Let b and c be two nodes on Γ such that the line joining them is parallel to (say) the x_1-axis. Since we need $v = 0$ on this line, and since v will be a polynomial in x_1 of degree ≤ 3 on this line we must have $v(b) = v(c) = 0$, $\dfrac{\partial v}{\partial x_1}(b) = \dfrac{\partial v}{\partial x_1}(c) = 0$. Also since we need $\dfrac{\partial v}{\partial \nu} = 0$ on this line and since $\dfrac{\partial v}{\partial \nu} = \dfrac{\partial v}{\partial x_2}$ is a polynomial in x_1 of degree ≤ 3, we need to set $\dfrac{\partial v}{\partial x_2}(b) = \dfrac{\partial v}{\partial x_2}(c) = 0$
105 and $\dfrac{\partial^2 v}{\partial x_1 \partial x_2}(b) = \dfrac{\partial^2 v}{\partial x, \partial x_2}(c) = 0$. Thus the degrees of freedom on all boundary nodes must be zero. The only "free" or "unknown" parameters are the degrees of freedom at the interior nodes. This takes care of the boundary conditions.

Let us now turn to the *Argyris triangle* (cf. Example 4.7).

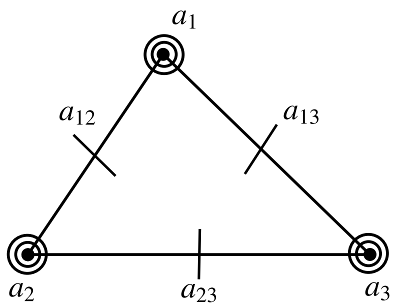

Figure 10.2:

We recall that $P_K = P_5$, $\dim P_K = 21$, and $\sum_K$ is given by (cf. Fig. 10.2)

(10.9)

$$\sum_K = \left\{ p(a_i), \frac{\partial p}{\partial x_1}(a_i), \ldots, \frac{\partial^3 p}{\partial x_2^2}(a_i), 1 \leq i \leq 3; \frac{\partial p}{\partial \nu}(a_{ij}), 1 \leq i < j \leq 3 \right\}.$$

We may replace the first and second derivative values at the vertices by $Dp(a_i)(a_{i+1} - a_i)$, $Dp(a_i)(a_{i-1} - a_i)$, $D^2p(a_i)(a_{i+1} - a_i, a_{i+1} - a_i)$, $D^2p(a_i)(a_{i+1} - a_i, a_{i-1} - a_i)$, $D^2p(a_i)(a_{i-1} - a_i, a_{i-1} - a_i)$ in order to get degrees of freedom for which the relations of the type (10.6) may be satisfied. However, *one cannot replace the normal derivatives* $\dfrac{\partial p}{\partial \nu}(a_{ij})$, $1 \leq i < j \leq 3$, *by such quantities since affine transformations do not preserve orthogonality.*

In order to estimate the errors we describe an "intermediary" finite element:

Example 10.2. The Hermite Triangle of Type (5).

Let $P_K = P_5 (\dim P_K = 21)$. Define

$$\begin{aligned}\sum_K = \Big\{ & p(a_i), Dp(a_i)(a_{i+1} - a_i), Dp(a_i)(a_{i-1} - a_i), \\ & D^2p(a_i)(a_{i+1} - a_i, a_{i+1} - a_i), D^2p(a_i)(a_{i+1} - a_i, a_{i-1} - a_i), \\ & D^2p(a_i)(a_{i-1} - a_i, a_{i-1} - a_i), 1 \leq i \leq 3; \\ & Dp(a_{ij})(a_k - a_{ij}), 1 \leq i < j \leq 3, k \neq 1, \neq j \Big\}.\end{aligned}$$

106

That is to say, the only change compared to the Argyris triangle is that we have replaced the normal derivatives at a_{ij} by the derivatives along the line joining a_{ij} to a_k, the opposite vertex.

Symbolically we can represent such a triangle as in Fig. 10.3.

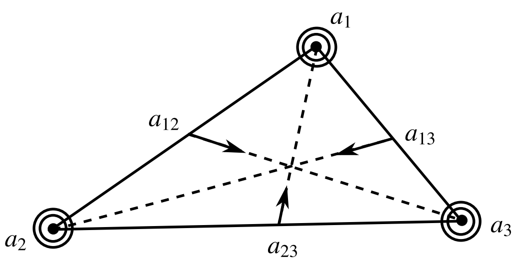

Figure 10.3:

This element can be put in an affine family as is readily seen. If Λ_K

is the associated interpolation operator, our error analysis yields

$$|v - \Lambda_K v|_{m,K} \leq C \frac{h_K^6}{\rho_K^m} |v|_{6,K},\ 0 \leq m \leq 6, \tag{10.10}$$

for $v \in H^6(K)$.

Remark 10.2. Though the Hermite triangle of type (5) yields an affine family, one cannot use it since $V_h \subset C^0(\overline{\Omega})$ but, in general, $V_h \not\subset C^1(\overline{\Omega})$ as is necessary for fourth-order problems. This is so because the adjacent triangles will not patch up, in general, in their derivatives along the medians; cf. 10.4.

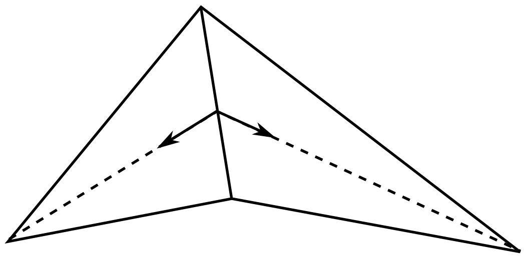

Figure 10.4:

107

Again we show how to take care of the boundary conditions in the Argyris triangle. We need again $v = \dfrac{\partial v}{\partial \nu} = 0$ on Γ. Let us have two nodes b, b', the vertices of a triangle lying on Γ with mid-point c. On this line v will be a polynomial of degree ≤ 5 in τ, an abscissa along this line. $\dfrac{\partial v}{\partial \nu}$ will be a polynomial in τ of degree ≤ 4 on this line. Hence for $v = 0$ on Γ we need to set, $v(b) = v(b') = 0$, $\dfrac{\partial v}{\partial \tau}(b) = \dfrac{\partial v}{\partial \tau}(b') = 0$, $\dfrac{\partial^2 v}{\partial \tau^2}(b) = \dfrac{\partial^2 v}{\partial \tau^2}(b') = 0$. For $\dfrac{\partial v}{\partial \nu} = 0$ on Γ we set, $\dfrac{\partial v}{\partial \nu}(b) = \dfrac{\partial v}{\partial \nu}(b') = \dfrac{\partial v}{\partial \nu}(c) = 0$, $\dfrac{\partial^2 v}{\partial \tau \partial \nu}(b) = \dfrac{\partial^2 v}{\partial \tau \partial \nu}(b') = 0$. Thus the only free or unknown parameters are $\dfrac{\partial^2 v}{\partial \nu^2}$ at vertices on Γ and the degrees of freedom at all interior nodes.

We now get an error estimate when we have triangulations of Argyris triangles. We use our usual terminology more loosely here[1]. By a *regular family* of triangulations made up of Argyris triangles we mean that all t_h consist only of Argyris triangles and that for all K, $\dfrac{h_K}{\rho_K} \leq \sigma$, a constant. We also assume that if $h = \max\limits_{K \in \mathrm{t}_h} h_K$, then $h \to 0$.

Theorem 10.1. *For a regular family* (t_h) *of triangulations made up of Argyris triangles*

$$|v - \pi_h v|_{m,\Omega} \leq C\, h^{6-m} |v|_{6,\Omega}, 0 \leq m \leq 6. \tag{10.11}$$

108

Proof. Let us denote the opposite vertex of $a_{ij}(i < j)$ by a_k. Let $\vec{\nu}_K$ be the unit outernormal at a_{ij} and $\vec{\tau}_K$ be the unit vector along the line $[a_i, a_j]$, at a_{ij} (cf. Fig. 10.5).

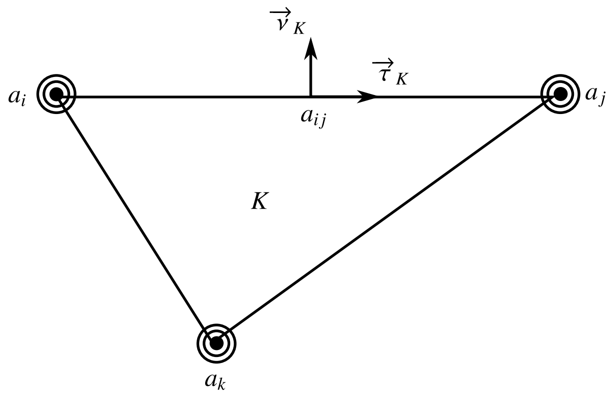

Figure 10.5:

Let π_K be the interpolation operator for the Argyris triangle K and let Λ_K be that for the corresponding Hermite triangle of type (5).

[1]Because we have to drop the assumption that all the finite elements are affine equivalent to a reference finite element.

Set $\delta = \pi_K v - \Lambda_K v$. Then $\delta \in P_5$. Now

$$\frac{\partial\delta}{\partial\nu_K}(a_{ij}) = \frac{\partial}{\partial\nu_K}(\pi_K v - \Lambda_K v)(a_{ij}) = \frac{\partial}{\partial\nu_K}(v - \Lambda_K v)(a_{ij}).$$

Also since $\pi_K v = \Lambda_K v$ along any side of K (since the values of these polynomials of degree 5 as well as those of their first and second derivatives agree at the end-points), we have $\dfrac{\partial\delta}{\partial\tau_K} = 0$.

Since

$$D\delta(a_{ij})(a_k - a_{ij}) = \frac{\partial\delta}{\partial\nu_K}(a_{ij})\langle a_k - a_{ij}, \vec{\nu}_K\rangle + \frac{\partial\delta}{\partial\tau_K}(a_{ij})\langle a_k - a_{ij}, \vec{\tau}_K\rangle,$$

where $\langle\cdot,\cdot\rangle$ is the Euclidean inner-product, substituting for $\dfrac{\partial\delta}{\partial\nu_K}$ and $\dfrac{\partial\delta}{\partial\tau_K}$ at a_{ij}, we get

$$D\delta(a_{ij})(a_k - a_{ij}) = \frac{\partial}{\partial\nu}(v - \Lambda_K v)(a_{ij})\langle a_k - a_{ij}, \vec{\nu}_K\rangle \tag{10.12}$$

109 Since $\delta \in P_5$, using the unisolvency in the Hermite triangle we may express δ in terms of its basis functions. Since all degrees of freedom except those of the type $D\delta(a_{ij})(a_k - a_{ij})$ are zero for δ, we have

$$\delta = \sum_{\substack{1\le i<j\le 3\\ k\ne i,k\ne j}} \frac{\partial}{\partial\nu_K}(v - \Lambda_K v)(a_{ij})\langle a_k - a_{ij}, \vec{\nu}_K\rangle p_{ijk}. \tag{10.13}$$

Now

$$|\langle a_k - a_{ij}, \vec{\nu}_K\rangle| \le \|a_k - a_{ij}\|\,\|\vec{\nu}_K\| \le h_K, \tag{10.14}$$

and

$$\begin{aligned} \left|\frac{\partial}{\partial\nu_K}(v - \Lambda_K v)(a_{ij})\right| &\le |v - \Lambda_K v|_{1,\infty,K} \\ &\le C(\text{meas}\, K)^{-\frac{1}{2}}\frac{h_K^6}{\rho_K}|v|_{6,K}. \end{aligned}$$

(Theorem 6.6 with $m = 1$, $k = 5$, $p = 2$, $q = \infty$). Also, meas $K \geq C\rho_K^2$, and we have

$$|\frac{\partial}{\partial \nu_K}(v - \Lambda_K v)(a_{ij})| \leq C\frac{h_K^6}{\rho_K^2}|v|_{6,K}. \tag{10.15}$$

Finally, by Theorem 6.4 and 6.5

$$|p_{ijk}|_{m,K} \leq C\frac{h_K}{\rho_K^m}|\hat{p}_{ijk}|_{m,\hat{K}}. \tag{10.16}$$

Combining (10.14), (10.15) and (10.16), we get

$$\begin{aligned}|\delta|_{m,K} &\leq \sum_{\substack{1\leq i<j\leq 3\\ k\neq i, k\neq j}} |\frac{\partial}{\partial \nu_K}(v - \Lambda_K v)(a_{ij})|\, |\langle a_k - a_{ij}, \overrightarrow{\nu}_K\rangle|\, |p_{ijk}|_{m,K} \\ &\leq C\frac{h_K^8}{\rho_K^{m+2}}|v|_{6,K}, \qquad (10.17)\end{aligned}$$

and hence,

$$\begin{aligned}|v - \pi_K v|_{m,K} &\leq |v - \Lambda_K v|_{m,K} + |\delta|_{m,K} \\ &\leq C\frac{h_K^6}{\rho_K^m}\left(1 + \frac{h_K^2}{\rho_K^2}\right)|v|_{6,K} \quad \text{(Using (10.10) and (10.17))} \\ &\leq C\, h^{6-m}|v|_{6,K} \quad (\text{since } h_K \leq h, \frac{h_K}{\rho_K} \leq \sigma).\end{aligned}$$

110

This on summing over K gives (10.11), thus completing the proof. □

Exercise 10.1. Perform the same analysis for the 18-degree-of freedom triangle. (cf. Exercise 4.7).

For the interpolation theory of the HCT-triangle (cf. Exercise 4.8), the normal derivatives are handled as in the present case. However the arbitrariness of the interior point is an obstacle to be overcome. For a discussion of this, see Ciarlet [4].

Another finite element, similar in its principle to the HCT-triangle used in the conforming finite element method for the plate problem is the *Fraeijs* de Veubeke and Sander Quadrilateral. See Ciavaldini and Nédélec [9].

These are all essentially the finite elements used in the "conforming" methods to approximate the plate problem (we will define such methods at the beginning of Sec. 11).

Chapter 11

Non-Conforming Methods for the Plate Problem

WE START WITH a brief classification of finite element methods. The 111
first class of methods are called *conforming methods*, which we have described upto now, except when we considered numerical integration. The second class consists of methods other than conforming. In the latter class we have the *Non-conforming* methods included:

Given the abstract problem, the *conforming methods* deal with the finding of subspaces $V_h \subset V$ and solving the problems

$$(P_h) \qquad a_h(u_h, v_h) = f_h(v_h), \quad \text{for all} \quad v_h \in V_h,$$

where $a_h = a$ and $f_h = f$ for all h and $u_h \in V_h$ is the required solution.

When we employ methods other than conforming we commit, in the terminology of G. Strang, "Variational Crimes". (See Strang and Fix [22]). These may occur in the following ways:

(i) When performing numerical integration, we may have a_h and f_h different from a and f respectively. However, V_h is a subspace of V;

(ii) The boundary Γ of Ω may be curved. In this case triangles lying in the interior will be triangles of straight edges while those meeting the boundary will have curved edges like parabolas. These are the so-called "isoparametric" finite elements. Hence if Ω_h is the union of the finite elements of the triangulation t_h, then, in general, $\Omega_h \neq \Omega$ and consequently $V_h \not\subset V$ (where V_h is a space of functions defined over Ω_h), $a_h \neq a$, $f_h \neq f$; for a discussion of these, see Ciarlet and Raviart [30], [31].

(iii) When employing *non-conforming methods* (which will be dealt with subsequently) though $\Omega_h = \Omega$, $f_h = f$, we will have $V_h \not\subset V$ and $a_h \neq a$.

112 (iv) One may employ any combination of the above three.

Let us return to the plate problem. For a conforming method we need the inclusion $V_h \subset H_0^2(\Omega)$ which essentially results from the inclusion $V_h \subset C^1(\overline{\Omega})$. Because of this necessity, when compared with second-order problems, we either have the dimension of P_K "large" (as in the case of the Argyris triangle) or that the structure of P_K is complicated (as in the HCT-triangle). Also one would like to have just $P_K = P_2$ since u is only in $H^3(\Omega)$ in most cases, but this is impossible by the Ženišek result (cf. Remark 4.3) which stresses that at least polynomials of degree 5 must be present in P_K.

Hence the desire to surmount these difficulties led to the devising of non-conforming methods, essentially developed by the Engineers.

Since the root of all trouble is the inclusion $V_h \subset H_0^2(\Omega)$, we drop this condition. Thus we start with $V_h \subset C^0(\overline{\Omega})$ and it is much easier from the computer programme view point. This of course, works only for a few finite elements, and we describe one of them.

Example 11.1. The Adini's rectangle; cf. Fig. 11.1.

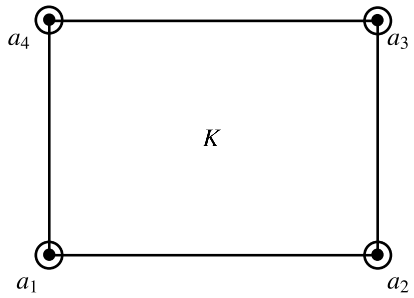

Figure 11.1:

The element K consists of a rectangle with vertices $\{a_i, 1 \leq i \leq 4\}$; the space P_K is given by $P_K = P_3 \oplus \{x_1 x_2^3\} \oplus \{x_1^3 x_2\}$, by which we mean polynomials of degree ≤ 4 whose only fourth-degree terms are those involving $x_1 x_2^3$ and $x_1^3 x_2$. Thus $P_3 \subset P_K$. We have the set of degrees of freedom:

$$\sum_K = \left\{ p(a_i), \frac{\partial p}{\partial x_1}(a_i), \ \frac{\partial p}{\partial x_2}(a_i), \ 1 \leq i \leq 4 \right\}.$$

113

Of course this element can be used only for plates with sides parallel to the coordinate axes, such as rectangular plates.

Exercise 11.1. Show that in Example 11.1, $\sum_K$ is P_K-unisolvent and that Adini's rectangle is a finite element of class C^0, and, in general, not of class C^1.

Thus we get 'a priori' that $V_h \subset H^1(\Omega)$. For the boundary condition, we set all degrees of freedom on the boundary nodes as zero. This gives us that $V_h \subset H_0^1(\Omega)$. Thus the only 'unknown' or 'free' parameters are the degrees of freedom at the interior nodes. Note that $\dfrac{\partial v_h}{\partial \nu}$ is zero only at the boundary nodes, in general.

In the abstract problem, we have $a(\cdot,\cdot)$ and $f(\cdot)$ given by

(11.1)
$$\begin{cases} a(u,v) = \int\limits_{\Omega}\left[\Delta u\Delta v + (1-\sigma)\left(2\dfrac{\partial^2 u}{\partial x_1\partial x_2}\dfrac{\partial^2 v}{\partial x_1\partial x_2} - \dfrac{\partial^2 u}{\partial x_1^2}\dfrac{\partial^2 v}{\partial x_2^2} - \dfrac{\partial^2 u}{\partial x_2^2}\dfrac{\partial^2 v}{\partial x_1^2}\right)\right]dx, \\ f(v) = \int\limits_{\Omega} fv dx,\ f \in L^2(\Omega). \end{cases}$$

The second integral is defined over V_h as well. Thus for the discrete problem (P_h) we may set $f_h = f$. However while for $u, v \in V_h$ the first integral is defined over each $K \in t_h$, we cannot define it over Ω, since we get Dirac measure-like terms along the boundary. To get over this, we now define

$$a_h(u_h,v_h) = \sum_{K\in t_h}\int\limits_K\left[\Delta u_h\Delta v_h + (1-\sigma)\left(2\frac{\partial^2 u_h}{\partial x_1\partial x_2}\frac{\partial^2 v_h}{\partial x_1\partial x_2} - \frac{\partial^2 u_h}{\partial x_1^2} - \frac{\partial^2 u_h}{\partial x_2^2}\frac{\partial^2 v_h}{\partial x_1^2}\right)\right]dx$$

(11.2)
$$= \sum_{K\in t_h}\int\limits_K\left[\sigma\Delta u_h\Delta v_h + (1-\sigma)\left(\frac{\partial^2 u_h}{\partial x_1^2}\frac{\partial^2 v_h}{\partial x_1^2} + \frac{\partial^2 u_h}{\partial x_2^2}\frac{\partial^2 v_h}{\partial x_2^2} + 2\frac{\partial^2 u_h}{\partial x_1\partial x_2}\frac{\partial^2 v_h}{\partial x_1\partial x_2}\right)\right]dx,$$

and we have the discrete problem (P_h): To find $u_h \in V_h$ such that for all $v_h \in V_h$

(11.3)
$$a_h(u_h,v_h) = f(v_h).$$

114

We now prove the existence and uniqueness of the solution u_h for (P_h). We define on V_h the seminorm

(11.4)
$$||v_h||_h = \left(\sum_{K\in t_h}|v_h|^2_{2,K}|^2_{2,K}\right)^{\frac{1}{2}}.$$

Notice that this may be defined over $V = H_0^2(\Omega)$ as well and for $v \in V$, $||v||_h = |v|_{2,\Omega}$. In the same way for $u, v \in V$, $a_h(u,v) = a(u,v)$.

We now show that the *seminorm* $||\cdot||_h$ *is indeed a norm* on V_h. Let $v_h \in V_h$ with $||v_h||_h = 0$. This gives that $\dfrac{\partial v_h}{\partial x_1}$ = constant over any K. But given adjacent finite elements the value of $\dfrac{\partial v_h}{\partial x_1}$ at the common vertices coincide and hence $\dfrac{\partial v_h}{\partial x_1}$ is constant over $\overline{\Omega}$. But this is zero on the boundary nodes. Hence $\dfrac{\partial v_h}{\partial x_1} = 0$ on $\overline{\Omega}$. Similarly $\dfrac{\partial v_h}{\partial x_2} = 0$ on $\overline{\Omega}$. Since $V_h \subset C^0(\Omega)$ and $v_h = 0$ on Γ, the above conditions give that $v_h = 0$ over $\overline{\Omega}$. Thus (11.4) defines a norm on V_h.

To show the existence and uniqueness of the solution of (P_h), we show that $a_h(\cdot,\cdot)$ is V_h-elliptic. In fact we do more than this. We show that the $a_h(\cdot,\cdot)$ are V_h-elliptic uniformly with respect to h.

Recall that from physical considerations, $0 < \sigma < \frac{1}{2}$ (see Sec. 2). Now

$$(11.5)\qquad \begin{aligned} a(v_h, v_h) &= \sum_{K\in t_h} \int\limits_K \sigma(\Delta v_h)^2 dx + (1-\sigma)\|v_h\|_h^2 \\ &\geq (1-\sigma)\|v_h\|_h^2. \end{aligned}$$

Remark 11.1. It was mentioned in passing in Sec. 10 that in order to apply non-conforming methods one needed the second term involving σ in the integral defining $a(\cdot,\cdot)$. The uniform V_h-ellipticity could not be got in the Hydrodynamical case where this term is absent.

We now proceed with the abstract error analysis. 115

Theorem 11.1 (STRANG)**.** *Let $a_h(\cdot,\cdot)$ be V_h-elliptic uniformly with respect to h with $\widetilde{\alpha} > 0$ so that for all $v_h \in V_h$*

$$(11.6)\qquad a_h(v_h, v_h) \geq \widetilde{\alpha}\|v_h\|_h.$$

Let in addition, there exist $\widetilde{M}$ such that for all $u_h, v_h \in V_h$

$$(11.7)\qquad |a(u_h, v_h)| \leq \widetilde{M}\|u_h\|_h\|v_h\|_h.$$

Assume that $a_h = a$ and $\|\cdot\|_h = \|\cdot\|$ on V. (These are needed to extend the definition of a_h and $\|\cdot\|_h$ to V). Then there exists a constant C, independent of h, such that

$$(11.8)\qquad \|u - u_h\|_h \leq C\left\{\inf_{v_h\in V_h}\|u - v_h\|_h + \sup_{w_h\in V_h}\frac{|f(w_h) - a_h(u, w_h)|}{\|w_h\|_h}\right\}$$

u_h being the solution of (P_h).

Proof. For all $v_h \in V_h$ we have

$$\|u - u_h\|_h \leq \|u - v_h\|_h + \|u_h - v_h\|_h. \tag{11.9}$$

Now for any $v_h \in V_h$, $f(u_h - v_h) = a_h(u_h, u_h - v_h)$, so that we may write

$$\begin{aligned}
\widetilde{\alpha}\|u_h - v_h\|_h^2 &\leq a_h(u_h - v_h, u_h - v_h) \\
&= a_h(u - v_h, u_h - v_h) + f(u_h - v_h) - a_h(u, u_h - v_h) \\
&\leq \widetilde{M}\|u - v_h\|_h\|u_h - v_h\|_h + |f(u_h - v_h) - a_h(u, u_h - v_h)|,
\end{aligned}$$

and thus,

$$\begin{aligned}
\|u_h - v_h\|_h &\leq \frac{\widetilde{M}}{\widetilde{\alpha}}\|u - v_h\|_h + \frac{1}{\widetilde{\alpha}}\frac{|f(u_h - v_h) - a_h(u, u_h - v_h)|}{\|u_h - v_h\|_h} \\
&\leq \frac{\widetilde{M}}{\widetilde{\alpha}}\|u - v_h\|_h + \frac{1}{\widetilde{\alpha}}\sup_{w_h \in V_h}\frac{|f(w_h) - a_h(u, w_h)|}{\|w_h\|_h}.
\end{aligned}$$

116

Substituting in (11.9) and varying $v_h \in V_h$ and taking the infimum we get (11.8). This completes the proof. ◻

Remark 11.2. In case the method is conforming, then $a_h(u, w_h) = a(u, w_h) = f(w_h)$ and the second term disappears in (11.8), leaving us with the original bound (3.1).

We note that for the $a_h(\cdot, \cdot)$ defined for the plate problem by (11.2), the conditions of Theorem 11.1 are satisfied. The condition (11.6) is embodied in (11.5). The condition (11.7) follows from the similar (continuity) condition on $a(\cdot, \cdot)$ and an application of the Cauchy-Schwarz inequality.

Exercise 11.2. Let (H) be a Hilbert space with innerproduct $(\cdot, \cdot)$ and norm $|\cdot|$. Let $(V, \|\cdot\|)$ be a subspace such that $V \hookrightarrow H$ and $\overline{V} = H$. Let $V_h \subset H$. Define

$$\mathscr{E}_h(u, v) = (f, v) - a_h(u, v), \quad \text{for all} \quad u, v \in V_h \cup V.$$

Then show that

$$|u - u_h| \leq \widetilde{M}\|u - u_h\|_h \left[\sup_{g\in H} \frac{1}{|g|} \inf_{\varphi_h\in V_h} \|\varphi - \varphi_h\|_h\right] + \sup_{g\in H}\left[\frac{1}{|g|} \inf_{\varphi_h\in V_h} (\mathscr{E}_h(u, \varphi - \varphi_h) + \mathscr{E}_h(\varphi, u - u_h))\right]$$

where for all $v \in V$, $a(v, \varphi) = g(v)$.

We now go on with the error analysis and study the order of convergence. We assume that $u \in H^3(\Omega) \cap H_0^2(\Omega)$ which is quite realistic from the regularity results.

Now, since $\pi_h u \in V_h$, we have that

$$\inf_{v_h\in V_h} \|u - v_h\|_h \leq \|u - \pi_h u\|_h.$$

117

Again applying error bounds for each K and then summing over all K we get

$$\|u - \pi_h u\|_h \leq Ch|u|_{3,\Omega}.$$

Thus

$$\inf_{v_h\in V_h} \|u - v_h\|_h \leq Ch|u|_{3,\Omega}. \tag{11.10}$$

Our aim is to get a similar estimate for the second term in (11.8). In fact we show that

$$\sup_{w_h\in V_h} \frac{|f(w_h) - a_h(u, w_h)|}{\|w_h\|_h} \leq Ch\|u\|_{3,\Omega}. \tag{11.11}$$

This entails more work. We define

$$\mathscr{E}_h(u, w_h) = f(w_h) - a_h(u, w_h) \tag{11.12}$$

for $u \in H^3(\Omega) \cap H_0^2(\Omega)$, $w_h \in V_h$. Since $w_h \in H_0^1(\Omega)$, there exists a sequence $\{w_h^n\}$ in $\mathscr{D}(\Omega)$ converging to w_h in $H_0^1(\Omega)$. Hence,

$$\int_\Omega f w_h^n dx = \int_\Omega \left[\Delta u \Delta w_h^n + (1 - \sigma)\left(2\frac{\partial^2 u}{\partial x_1 \partial x_2}\frac{\partial^2 w_h^n}{\partial x_1 \partial x_2} - \frac{\partial^2 u}{\partial x_1^2}\frac{\partial^2 w_h^n}{\partial x_2^2} - \frac{\partial^2 u}{\partial x_2^2}\frac{\partial^2 w_h^n}{\partial x_1^2}\right)\right] dx$$

$$= -\int_\Omega (\text{grad}\,\Delta u)(\text{grad}\, w_h^n)dx,$$

by Green's formula. The term involving $(1 - \sigma)$, by Lemma 2.2, can be converted to an integral over Γ. All integrals over Γ vanish since $w_h^n \in \mathscr{D}(\Omega)$. Since both sides of the above relation are continuous linear functionals on $H_0^1(\Omega)$, we can pass to the limit to obtain

$$f(w_h) = \int_\Omega f w_h dx = -\int_\Omega (\text{grad}\,\Delta u)(\text{grad}\, w_h)dx \tag{11.13}$$

for all $w_h \in V_h$. Now,

$$\begin{aligned}
a_h(u, w_h) &= \sum_{K\in t_h} \int_K \left(\Delta u \Delta w_h + (1-\sigma)\left[2\frac{\partial^2 u}{\partial x_1 \partial x_2}\frac{\partial^2 w_h}{\partial x_1 \partial x_2} - \frac{\partial^2 u}{\partial x_1^2}\frac{\partial^2 w_h}{\partial x_2^2} - \frac{\partial^2 u}{\partial x_2^2}\frac{\partial^2 w_h}{\partial x_1^2}\right]\right)dx \\
&= \sum_{K\in t_h}\Bigg[-\int_K \text{grad}\,\Delta u\,\text{grad}\, w_h dx \\
&\quad + \oint_{\partial K} \Delta u \frac{\partial w_h^K}{\partial \nu_K} d\gamma + (1-\sigma)\oint_{\partial K}\left(-\frac{\partial^2 u}{\partial \tau_K^2}\frac{\partial w_h}{\partial \nu_K} + \frac{\partial^2 u}{\partial \nu_K \partial \tau_K}\frac{\partial w_h}{\partial \tau_K}\right)d\gamma\Bigg],
\end{aligned} \tag{11.14}$$

118 by Green's formula (2.15) and Lemma 2.2 again. Notice however that by standard orientation arguments and the continuity of w_h over $\overline{\Omega}$,

$$\sum_{K\in t_h}\oint_{\partial K}\frac{\partial^2 u}{\partial \nu_K \partial \tau_K}\frac{\partial w_h}{\partial \tau_K}d\gamma = 0. \tag{11.15}$$

Using (11.13), (11.14) and (11.15), we substitute in (11.12) to get

$$\mathscr{E}_h(u, w_h) = \sum_{K\in t_h}\oint_{\partial K}\left(-\Delta u + (1-\sigma)\frac{\partial^2 u}{\partial \tau_K^2}\right)\frac{\partial w_h}{\partial \nu_K}d\gamma. \tag{11.16}$$

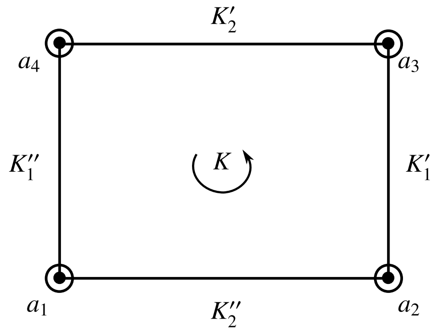

Figure 11.2:

Splitting the boundary into four parts as in Fig. 11.2, we get

$$\mathscr{E}_h(u, w_h) = \sum_{K \in \mathfrak{t}_h} \left(\Delta_{1,K}\left(u, \frac{\partial w_h}{\partial x_1}\right) + \Delta_{2,K}\left(u, \frac{\partial w_h}{\partial x_2}\right)\right), \tag{11.17}$$

where for $j = 1, 2$, we define

$$\begin{aligned} \Delta_{j,K}\left(u, \frac{\partial w_h}{\partial x_j}\right) &= \int\limits_{K_j'} \left(-\Delta u + (1-\sigma)\frac{\partial^2 u}{\partial \tau_K^2}\right)\left(\frac{\partial w_h}{\partial x_j} - \Lambda_K\left(\frac{\partial w_h}{\partial x_j}\right)\right) d\gamma \\ &\quad - \int\limits_{K_j''} \left(-\Delta u + (1-\sigma)\frac{\partial^2 u}{\partial \tau_K^2}\right)\left(\frac{\partial w_h}{\partial x_j} - \Lambda_K\left(\frac{\partial w_h}{\partial x_j}\right)\right) d\gamma, \end{aligned} \tag{11.18}$$

Λ_K being the Q_1-interpolation operator associated with the values at the **119**
four vertices.

Note that (11.17) is the same as (11.15). This is evident provided we show that the contribution of the terms involving Λ_K is zero. This is so because $w_h = 0$ on the boundary Γ and on the common boundaries, $\Lambda_K\left(\frac{\partial w_h}{\partial x_j}\right)$ is linear and equal in value for both adjacent finite elements since it agrees at the vertices, but occurs with opposite signs as is obvious from Fig. 11.3 ($K_1^{1'} = K_1^{2''}$).

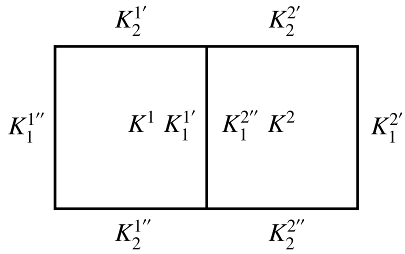

Figure 11.3:

We also record that $\left.\dfrac{\partial w_h}{\partial x_j}\right|_K \in \partial_j P_K$ where

$$\partial_j P_K = \left\{ \frac{\partial p}{\partial x_j} : K \to \mathbb{R};\ p \in P_K \right\}. \tag{11.19}$$

We now prove a result analogous to the Bramble-Hilbert lemma which will help us to estimate that $\Delta_{j,K}$'s and hence $\mathscr{E}_h(u, w_h)$.

Theorem 11.2 (BILINEAR LEMMA)**.** *Let $\Omega \subset \mathbb{R}^n$ be open with Lipschitz continuous boundary Γ. Let W be a subspace of $W^{l+1,q}(\Omega)$ such that $P_1 \subset W$. Let b be a continuous bilinear form over $W^{k+1,p}(\Omega) \times W$ such that*

$$\begin{cases} b(p, w) = 0 \quad \text{for all} \quad p \in P_k,\ w \in W \\ b(v, q) = 0 \quad \text{for all} \quad v \in W^{k+1,p}(\Omega),\ q \in P_1. \end{cases}$$

120 *Then there exists a constant $C = C(\Omega)$ such that for all $v \in W^{k+1,p}$ (Ω), $w \in W$,*

$$|b(v, w)| \leq C\|b\|\ |v|_{k+1,p,\Omega}|w|_{l+1,q,\Omega}. \tag{11.21}$$

Proof. For a given $w \in W$, $b(\cdot, w) : v \mapsto b(v, w)$ is a continuous linear form on $W^{k+1,p}(\Omega)$ vanishing on P_k. Hence by the Bramble-Hilbert lemma,

$$|b(v, w)| \leq C\|b(\cdot, w)\|^*_{k+1,p,\Omega}|v|_{k+1,q,\Omega}. \tag{11.22}$$

However for all $q \in P_1$, $b(v,w) = b(v, w+q)$ so that

$$|b(v,w)| \leq ||b||\; ||v||_{k+1,p,\Omega} ||w+q||_{l+1,q,\Omega}$$

and hence

$$\begin{aligned} |b(v,w)| &\leq ||b||\; ||v||_{k+1,p,\Omega} \inf_{q \in P_1} ||w+q||_{l+1,q,\Omega} \\ &\leq C||b||\; ||v||_{k+1,p,\Omega} |w|_{l+1,q,\Omega}, \quad \text{by the theorem 6.2,} \end{aligned}$$

so that

$$||b(\cdot,w)||^*_{k+1,p,\Omega} \leq C||b||\; |w|_{l+1,q,\Omega}.$$

and substituting in (11.22), we get (11.21), which completes the proof. □

We may now prove the theorem on our error estimate and order of convergence.

Theorem 11.3. *For a regular family* (t_h) *of triangulations made up of Adini's rectangles*

$$\text{(11.23)} \qquad ||u - u_h||_h \leq C\, h||u||_{3,\Omega}.$$

Proof. Let us first estimate $|\Delta_{j,K}|$ for $j = 1, 2$. Set

$$\begin{cases} \varphi = -\Delta u + (1-\sigma)\dfrac{\partial^2 u}{\partial \tau_K^2} \in H^1(K), \quad \text{since} \quad u \in H^3(\Omega), \\ v = \dfrac{\partial w_h}{\partial x_1} \in \partial_1 P_K. \end{cases}$$

Define 121

$$\text{(11.24)} \qquad \delta_{1,K}(\varphi, v) = \int\limits_{K_1'} \varphi(v - \Lambda_K v) d\gamma - \int\limits_{K_1''} \varphi(v - \Lambda_K V) d\gamma,$$

for $v \in \partial_1 P$, $\varphi \in H^1(K)$. If h_2 is the length of K_1' (and K_1'') and h_1 is that of K_2' (and K_2'') we have by a simple change of variable

$$\text{(11.25)} \qquad \delta_{1,K}(\varphi, v) = h_2 \delta_{1,\hat{K}}(\hat{\varphi}, \hat{v}),$$

where $\hat{K}$ is the reference finite element. Since $P_0 \subset Q_1$ which is preserved by Λ_K we have that for all $\hat{v} \in P_0$, $\hat{\varphi} \in H^1(\hat{K})$, $\delta_{1,\hat{K}}(\hat{\varphi}, \hat{v}) = 0$. Now let $\hat{\varphi} \in P_0$ and $\hat{v} \in \widehat{\partial_1 P}$. We wish to show that $\delta_{1,\hat{K}}(\hat{\varphi}, \hat{v}) = 0$:

We may take for $\hat{K}$, the unit square. Since $\hat{\varphi} \in P_0$, its value on $\hat{K}$ is a constant, say, b_0. Now let $\hat{v} \in \widehat{\partial_1 P}$. Then $\hat{v}$ is of the form

$$\hat{v} = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_1 x_2 + a_5 x_2^2 + a_6 x_1^2 x_2 + a_7 x_2^3. \tag{11.26}$$

Taking the values at the four vertices we get

$$\Lambda_K \hat{v} = a_0 + (a_1 + a_3)x_1 + (a_2 + a_5 + a_7)x_2 + (a_4 + a_6)x_1 x_2. \tag{11.27}$$

Now K_1' is the line $x_1 = 1$ and K_1'' is the line $x_1 = 0$. Thus

$$\begin{aligned} \hat{v} - \Lambda_K \hat{v}\Big|_{x_1=0} &= -(a_5 + a_7)x_2 + a_5 x_2^2 + a_7 x_2^3, \\ \hat{v} - \Lambda_K \hat{v}\Big|_{x_1=1} &= -(a_5 + a_7)x_2 + a_5 x_2^2 + a_7 x_2^3. \end{aligned}$$

Hence,

$$\begin{aligned} \int_{K_1'} \hat{\varphi}(\hat{v} - \Lambda_K \hat{v}) d\gamma &= \int_0^1 b_0(-(a_5 + a_7)x_2 + a_5 x_2^2 + a_7 x_2^3) dx_2 \\ &= \int_{K_1''} \hat{\varphi}(\hat{v} - \Lambda_K \hat{v}) d\gamma. \end{aligned} \tag{11.28}$$

122 Thus $\delta_{1,\hat{K}}(\hat{\varphi}, \hat{v}) = 0$ for $\hat{\varphi} \in P_0$, $\hat{v} \in \widehat{\partial_1 P}$.

Note further that the bilinear form $\delta_{1,\hat{K}}$ is continuous, for

$$\begin{aligned} |\delta_{1,\hat{K}}(\hat{\varphi}, \hat{v})| &\le C \|\hat{\varphi}\|_{L^2(\partial\hat{K})} \|\hat{v}\|_{L^2(\partial\hat{K})}. \\ &\le C \|\hat{\varphi}\|_{1,\hat{K}} \|\hat{v}\|_{1,\hat{K}}, \quad \text{by the Trace theorem (cf. Th. 2.3).} \end{aligned}$$

Thus we may apply the bilinear lemma to the bilinear form $\delta_{1,\hat{K}}$ with $l = k = 0$ to get

$$|\delta_{1,\hat{K}}(\hat{\varphi}, \hat{v})| \le C |\hat{\varphi}|_{1,\hat{K}} |\hat{v}|_{1,\hat{K}}. \tag{11.29}$$

We also have the relations

$$
\begin{cases} |\hat{\varphi}|_{1,\hat{K}} \leq C\|B_K\|\,|\det B_K|^{-\frac{1}{2}}|\varphi|_{1,K}, \\ |\hat{v}|_{1,\hat{K}} \leq C\|B_K\|\,|\det B_K|^{-\frac{1}{2}}|v|_{1,K}. \end{cases} \tag{11.30}
$$

Now $\|B_K\| \leq C\, h_K$ and $|\det B_K| = \text{meas}\, K/\,\text{meas}\, \hat{K} \geq C\rho_K^2$, and thus $\|B_K\|\,|\det B_K|^{-\frac{1}{2}} \leq \dfrac{C\, h_K}{\rho_K} \leq C$. Also $h_2 \leq h_K$, so that

$$
|\Delta_{1,K}(\varphi, v)| \leq C\, h_K\|u\|_{3,K}\|w_h\|_{2,K} \tag{11.31}
$$

where

$$
\begin{cases} \varphi = -\Delta u + (1-\sigma)\frac{\partial^2 u}{\partial \tau_K^2}, \\ v = \frac{\partial w_h}{\partial x_1}, \end{cases}
$$

and similarly,

$$
|\Delta_{2,K}(\varphi, v)| \leq C\, h_K\|u\|_{3,K}\|w_h\|_{2,K}. \tag{11.32}
$$

These inequalities lead us to the estimate

$$
|f(w_h) - a_h(u, w_h)| \leq Ch\|u\|_{3,\Omega}\|w_h\|_h, \tag{11.33}
$$

for a *regular family* of triangulations made up of Adini's rectangles. **123**
Thus varying w_h over V_h and taking the supremum, we get the estimate (11.11).

Using (11.10) and (11.11) and substituting in (11.8) we get the required estimate as given in (11.23). This completes the proof. □

Remark 11.3. By the Duality Argument, Lesaint and Lascaux [15] have proved that $\|u - u_h\|_{1,\Omega} \leq Ch^2\|u\|_{4,\Omega}$ assuming $u \in H^4(\Omega)$. They have also got an improved $0(h^2)$ convergence order in the $\|\cdot\|_h$ norm, when all the rectangles are equal - a "superconvergence" result.

We close this section with a brief description of other types of finite elements used in non-conforming methods.

Example 11.2. The Zienkiewicz triangle *(cf. Exercise 4.6)* cf. Fig. 11.4.

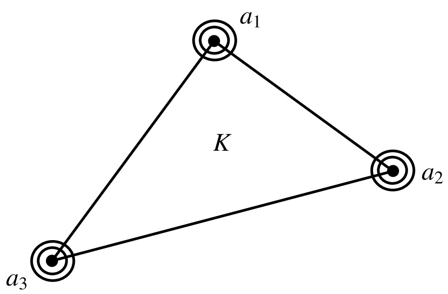

Figure 11.4:

We get $V_h \subset C^0(\overline{\Omega})$ only and hence the method is non-conforming. It does not always yield convergence. The method works if the sides of all triangles are parallel to three directions only, as in Fig. 11.5.

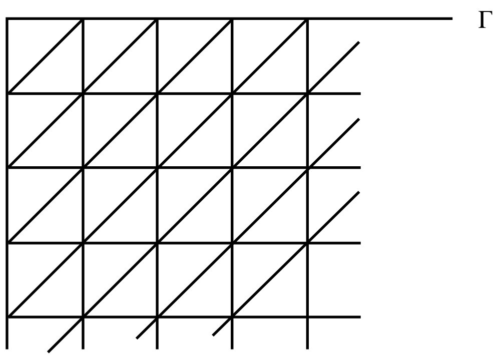

Figure 11.5:

124 This is not so if the number of directions is four, as in Fig. 11.6. (The Union Jack Problem).

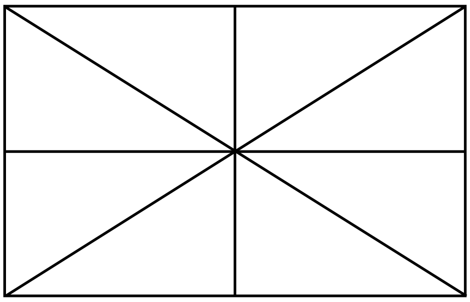

Figure 11.6:

Example 11.3. Morley's Triangle (cf. Fig. 11.7).

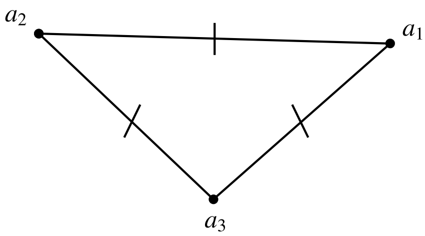

Figure 11.7:

Here $P_K = P_2$. We *always* get convergence for regular families, of course. In fact if $u \in H^4(\Omega)$, then

$$\|u - u_h\|_h \leq Ch\|u\|_{4,\Omega}. \tag{11.34}$$

What is astonishing is that this finite element is not even of class C^0.

Example 11.4. *Fraeijs de Veubeke triangle.* This finite element is again a triangle. Apart from the values of the polynomials at the vertices and the mid-points of the sides, we also take the average normal derivative along the sides. Here the space P_K, which we will not describe, satisfies the inclusion

$$P_2 \subset P_K \subset P_3,$$

and

$$\sum_K = \left\{ p(a_i), 1 \le i \le 3; p(a_{ij}), 1 \le i < j \le 3; \int_{K'_i} \frac{\partial p}{\partial \nu} d\gamma, 1 \le i \le 3 \right\}.$$

125 The finite element is shown symbolically in Fig. 11.8.

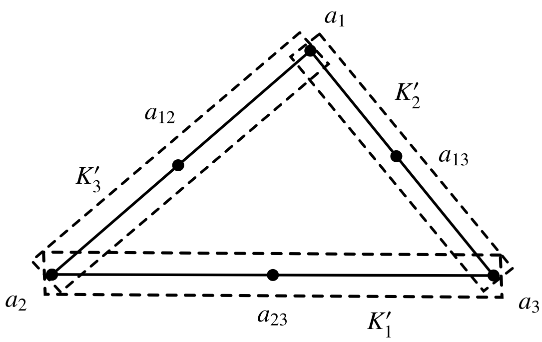

Figure 11.8:

Here also the finite element is not of class C^0 in general, but the method always yields convergence.

References: For general reference on non-conforming methods, see Strang and Fix [22], for the bilinear lemma, see Ciarlet [29]. For a detailed study of the Zienkiewicz triangle, Moreley's triangle and Fraeijs de Veubeke triangle, see Lascaux and Lesaint [15]. For a nonconforming method with penalty, see Babuska and Zlámal [26].

Bibliography

[1] BRAMBLE, J.H: THOMÉE, V. **126**

Interior maximum norm estimates for some simple finite element methods. *Revue FranÇaise d'Automatique, Informatique, Recherche Operationelle* (RAIRO), R-2, (1974), pp.-5-18.

[2] BRAMBLE, J.H; ZLÁMAL, M.

Triangular elements in the finite element methods, *Mathematics of Computation*, 24, (1970), pp. 809-820.

[3] BREZIS, H; STAMPACCHIA, G.

Sur la régularité de la solution d'inéquations elliptiques, *Bull. Soc. Math. France*, 96(1968), pp. 153-180.

[4] CIARLET, P.G.

Sur l'élément de Clough et Tocher, *RAIRO*, R-2, (1974), pp. 19-27.

[5] CIARLET, P.G; RAVIART, P.-A.

La Méthodes des Éléments Finis pour les Problèms aux Limites Elliptiques, (to appear).

[6] CIARLET, P.G; RAVIART, P.-A.

Maximum principle and uniform convergence for the finite element method, Computer Methods in Applied Mechanics and Engineering, 2, (1973), pp. 17-31.

[7] CIARLET, P.G; RAVIART, P.-A.

General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with applications to finite element methods, *Arch. Rational Mech. Anal.* 46, (1972), pp. 177-199.

[8] CIARLET, P.G; WAGSCHAL, C.

Multipoint Taylor formulas and applications to the finite element method, *Numer. Math.*, 17, (1971), pp. 84-100.

[9] CIAVALDINI, J.F; NEDELEC, J.C.

Sur l'élément de Fraeijs de Veubeke et Sander, *RAIRO*, R-2, (1974), pp. 29-45.

[10] DUVAUT, G; LIONS, J.-L.

Les Inéquations en Mécanique et en Physique, Dunod, Paris (1972).

[11] FALK, R.S.

Error estimates for the approximation of a class of variational inequalities, *Math. Comput*. (to appear).

127 [12] FALK, R.S.

Approximation of an elliptic boundary value problem with unilateral constraints, *RAIRO*, (to appear).

[13] HABER, S.

Numerical evaluation of multiple integrals, *SIAM Review*, 12, (1970), pp. 481-526.

[14] LANDAU, L; LIFCHITZ, E.

Théorie de l'Elasticité, Mir, Moscow (1967).

[15] LASCAUX, P; LESAINT, P.

Some non-conforming finite elements for the plate bending problem, *RAIRO*, (to appear).

[16] LEWY, H; STAMPACCHIA, G.

On the regularity of the solution of a variational inequality, *Comm. Pure Appl. Math.*, 22, (1969), pp. 153-188.

[17] LIONS, J.-L; MAGENES, E.

Problèmes aux Limites non Homogenes et Applications. Vol. 1, Dunod, Paris, (1968).

[18] LIONS, J.-L; STAMPACCHIA, G.

Variational inequalities, *Comm. Pure Appl. Math.*, 20, (1967). pp. 493-519.

[19] MOSCO, U; STRANG, G.

One-sided approximation and variational inequalities, *Bull. Amer. Math. Soc.* 80 (1974), pp. 308-312.

[20] NEČAS, J.

Les Méthodes Directes en Theorie des Equations Elliptiques, Masson, Paris, (1967).

[21] STRANG, G.

Approximation in the finite element method, *Number. Math.*, 19(1972), pp. 81-98.

[22] STRANG, G; FIX, G.J.

An Analysis of the Finite Element Method, Prentice-Hall, Englewood Cliffs, (1973).

[23] ŽENIŠEK, A.

Interpolation polynomials on the triangle, *Numer. Math.*, 15, (1970), pp. 283-296.

[24] ŽENIŠEK, A.

Polynomial approximation on tetrahedrons in the finite element method, *J. Approximation Theory*, 7, (1973), pp. 334-351.

[25] ZLÁMAL, M. **128**

On the finite element method, *Numer. Math.*, 12, (1968), pp. 394-409.

[26] BABUŠKA, I.; ZLÁMAL, M.

Nonconforming elements in the finite element method with penalty, *SIAM J. Numer. Anal.*, 10(1973), pp. 863-875.

[27] BRAMBLE, J.H.; HILBERT, S.R.

Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation, *SIAM J. Numer. Anal.*, 7, (1970), pp. 113-124.

[28] BRAMBLE, J.H.; HILBERT, S.R.

Bounds for a class of linear functionals with applications to Hermite interpolation, *Numer. Math.*, 16, (1971), pp. 362-369.

[29] CIARLET, P.G.

Conforming and nonconforming finite element methods for solving the plate problem, in *Conference on the Numerical Solution of Differential Equations* (G.A. WATSON, Editor), pp. 21-31, Lecture Notes in Mathematics, Vol. 363, Springer-Verlag, Berlin (1974).

[30] CIARLET, P.G.; RAVIART, P.-A.

Interpolation theory over curved elements, with applications to finite element methods, *Comp. Methods in Appl. Mech. and Engineering*, 1, (1972), pp. 217-249.

[31] CIARLET, P.G.; RAVIART, P.-A.

The combined effect of curved boundaries and numerical integration in isoparametric finite element methods, in *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations* (A.K. AZIZ, Editor), pp. 409-474, Academic Press, New York, (1972).

[32] ZLÁMAL, M.

A finite element procedure of the second order of accuracy, *Numer. Math.*, 16, (1970), pp. 394-402.

129 Additional References

Books (Additional References)

BABUSKA, I; AZIZ, A.K. First part of '*The Mathematical Foundations of the Finite Element Method with applications to Partial Differential Equatios*' (Symposium, University of Maryland, Baltimore, June-26-30, 1972) (Edited by A.K. AZIZ) Academic Press, New York, (1972).

NORRIE, D.H; DE VRIES, G. *The Finite Element Method - Fundamentals and Applications*, Academic Press, New York, (1973).

ODEN, J.T. *Finite Elements of Nonlinear Continue*, McGraw Hill, New York, (1972).

ZIENKIEWICZ, O.C. *The finite Element Method in Engineering Science*, McGraw Hill, London, (1971).

Journals (where articles dealing with the finite element method may be regularly found).

Revue Française d'Automatique, Informatique, Recherche Operationelle (RAIRO), "Red" Series

Computer Methods in Applied Mechanics and Engineering.

International Journal for Numerical Methods in Engineering.

Mathematics of Computation.

SIAM Journal on Numerical Analysis

Numerische Mathematik.

Proceedings of Conferences totally or partially devoted to the finite element method:

The Mathematics of Finite Elements and Applications, (Conference, Brunel University, 18-20 April, 1972), Edited by J.R. WHITEMAN, Academic Press, London, 1973.

Conference on the Numerical Solution of Differential Equations, (Conference, Dundee, 03-06 July 1973), Edited by G.A. WATSON, Lecture Notes in Mathematics, Vol. *363*, Springer-Verlag, New York, 1974.

130 *The Mathematical Foundations of the Finite Element Method with*

Applications to Partial Differential Equations—, (Symposium, University of Maryland, Baltimore, 26-30 June. 1972), Edited by A.K. AZIA, Academic Press. New York, 1972.

Computing Methods in Applied Science and Engineering, *Vol. 1 and 2 (International Symposium IRIA LABORIA, Versailles, 17-21 December, 1973) Edited by R. GLOWINSKI and J.L. LIONS, Lecture notes in Computer Science, Vol.,* 10 *and* 11, *Springer-Verlag, New York, 1974.*

Topics in Numerical Analysis *(Royal Irish Academy Conference on Numerical Analysis, Dublin, 1972), Edited by J.J.H. MILLER, Academic Press, New York, 1973.*

Mathematical Aspects of Finite Elements in Partial Differential Equations *(Symposium, 01-03 April, 1974, Mathematics Research Center, The University of Wisconsin, Madison), Edited by C. de BOOR, Academic Press, New York, 1974.*

Proceedings of the Second Conference on the Mathematics of Finite Elements and Applications, *(Brunel University, April 07-10, 1975, to appear.)*