CHAPTER 2

# INTRODUCTION TO THE FINITE ELEMENT METHOD

## Introduction

The basic scope of this chapter is to introduce the *finite element method* and to give a thorough *description* of the use of this method for approximating the solutions of second-order or fourth-order problems posed in variational form over a space $V$. A well-known approach for approximating such problems is *Galerkin's method*, which consists in defining similar problems, called *discrete problems*, over finite-dimensional subspaces $V_h$ of the space $V$. Then the *finite element method in its simplest form* is a Galerkin's method characterized by *three basic aspects* in the construction of the space $V_h$: First, a *triangulation* $\mathcal{T}_h$ is established over the set $\bar{\Omega}$, i.e., the set $\bar{\Omega}$ is written as a finite union of *finite elements* $K \in \mathcal{T}_h$. Secondly, the function $v_h \in V_h$ are *piecewise polynomials*, in the sense that for each $K \in \mathcal{T}_h$, the spaces $P_K = \{v_{h|K};\, v_h \in V_h\}$ consist of polynomials. Thirdly, there should exist a basis in the space $V_h$ whose functions have *small supports*. These three basic aspects are discussed in Section 2.1, where we also give simple criteria which insure the validity of inclusions such as $V_h \subset H^1(\Omega)$, $V_h \subset H_0^1(\Omega)$, etc... (Theorems 2.1.1 and 2.1.2). We also briefly indicate how the three basic aspects are still present in the more general finite element methods to be subsequently described. In this respect, we shall reserve the terminology *conforming finite element method* for the simplest such method (as described in this chapter).

In Section 2.2, we describe various examples of *finite elements*, which are either $n$-simplices (*simplicial* finite elements) or $n$-rectangles (*rectangular* finite elements), in which either all *degrees of freedom* are point values (*Lagrange* finite elements) or some degrees of freedom are *directional derivatives* (*Hermite* finite elements), which yield either the inclusion $X_h \subset H^1(\Omega)$ (finite elements *of class* $\mathscr{C}^0$) or the inclusion

$X_h \subset H^2(\Omega)$ (finite elements *of class* $\mathscr{C}^1$) when they are assembled in a *finite element space* $X_h$.

Then in Section 2.3, *finite elements* and *finite element spaces* are given general definitions, and we proceed to discuss their various properties. Of particular importance are the notion of an *affine family* of finite elements (where all the finite elements of the family can be obtained as images through affine mappings of a single *reference finite element*) and the notion of the *$P_K$-interpolation operator* (a basic relationship between these two notions is proved in Theorem 2.3.1). The $P_K$-interpolation operator and its global counterpart, the *$X_h$-interpolation operator* both play a fundamental role in the interpolation theory in Sobolev spaces that will be developed in the next chapter. We also show how to impose *boundary conditions* on functions in finite element spaces.

We conclude Section 2.3 by briefly indicating some reasons for which a particular finite element should be preferred to another one in practical computations.

In Section 2.4, we define the *convergence* and the *order of convergence* for a family of discrete problems. In this respect, *Céa's lemma* (Theorem 2.4.1) is crucial: The *error* $\|u - u_h\|$, i.e., the distance (measured in the norm of the space $V$) between the solution $u$ of the original problem and the solution $u_h$ of the discrete problem, is (up to a constant independent of the space $V_h$) bounded above by the distance $\inf_{v_h \in V_h} \|u - v_h\|$ between the function $u$ and the subspace $V_h$. Indeed, all subsequent convergence results will be variations on this theme!

## 2.1. Basic aspects of the finite element method

### *The Galerkin and Ritz methods*

Consider the linear abstract variational problem: Find $u \in V$ such that

$$\forall v \in V, \quad a(u, v) = f(v), \tag{2.1.1}$$

where the space $V$, the bilinear form $a(\cdot, \cdot)$, and the linear form $f$ are assumed to satisfy the assumptions of the Lax–Milgram lemma (Theorem 1.1.3). Then the *Galerkin method* for approximating the solution of such a problem consists in defining similar problems in *finite-dimensional subspaces* of the space $V$. More specifically, with any finite-dimensional subspace $V_h$ of $V$, we associate the *discrete problem*:

Find $u_h \in V_h$ such that

$$\forall v_h \in V_h, \quad a(u_h, v_h) = f(v_h). \tag{2.1.2}$$

Applying the Lax–Milgram lemma, we infer that such a problem has one and only one solution $u_h$, which we shall call a *discrete solution.*

**Remark 2.1.1.** In case the bilinear form is symmetric, the discrete solution is also characterized by the property (Theorem 1.1.2)

$$J(u_h) = \inf_{v_h \in V_h} J(v_h), \tag{2.1.3}$$

where the functional $J$ is given by $J(v) = \frac{1}{2}a(v, v) - f(v)$. This alternate definition of the discrete solution is known as the *Ritz method.* □

### *The three basic aspects of the finite element method. Conforming finite element methods*

Let us henceforth assume that the abstract variational problem (2.1.1) corresponds to a second-order or to a fourth-order elliptic boundary value problem posed over an open subset $\Omega$ of $\mathbf{R}^n$, with a Lipschitz-continuous boundary $\Gamma$. Typical examples of such problems have been studied in Section 1.2.

In order to apply Galerkin method, we face, by definition, the problem of constructing finite-dimensional subspaces $V_h$ of spaces $V$ such as $H_0^1(\Omega)$, $H^1(\Omega)$, $H_0^2(\Omega)$, etc...

*The finite element method, in its simplest form, is a specific process of constructing subspaces* $V_h$, which shall be called *finite element spaces.* This construction is characterized by *three basic aspects,* which for convenience shall be recorded as (FEM 1), (FEM 2) and (FEM 3), respectively, and which shall be described in this section.

(FEM 1) *The first aspect, and certainly the most characteristic, is that a triangulation $\mathcal{T}_h$ is established over the set $\bar{\Omega}$, i.e., the set $\bar{\Omega}$ is subdivided into a finite number of subsets K, called finite elements, in such a way that the following properties are satisfied:*

($\mathcal{T}_h$1) $\bar{\Omega} = \bigcup_{K \in \mathcal{T}_h} K$.

($\mathcal{T}_h$2) For each $K \in \mathcal{T}_h$, the set $K$ is closed and the interior $\mathring{K}$ is non empty.

($\mathcal{T}_h$3) For each distinct $K_1$, $K_2 \in \mathcal{T}_h$, one has $\mathring{K}_1 \cap \mathring{K}_2 = \phi$.

($\mathcal{T}_h$4) For each $K \in \mathcal{T}_h$, the boundary $\partial K$ is Lipschitz-continuous.

**Remark 2.1.2.** A fifth condition ($\mathcal{T}_h5$) relating "adjacent" finite elements, will be introduced in the next section. □

Once such a triangulation $\mathcal{T}_h$ is established over the set $\bar{\Omega}$, one defines a ***finite element space*** $X_h$ through a specific process, which will be illustrated by many examples in the next section and subsequently. We shall simply retain for the moment that $X_h$ is a ***finite-dimensional*** space of functions defined over the set $\bar{\Omega}$ (we shall deliberately ignore at this stage instances of finite element spaces whose "functions" may have two definitions across "adjacent" finite elements; see Section 2.3).

Given a finite element space $X_h$, we define the (finite-dimensional) spaces

$$P_K = \{v_{h|K};\quad v_h \in X_h\}$$

spanned by the restrictions $v_{h|K}$ of the functions $v_h \in X_h$ to the finite elements $K \in \mathcal{T}_h$. Without specific assumptions concerning the spaces $P_K$, $K \in \mathcal{T}_h$, there is no reason for an inclusion such as $X_h \subset H^1(\Omega)$ – let alone an inclusion such as $X_h \subset H^2(\Omega)$ – to hold.

In order to obtain such inclusions, we need additional conditions of a particularly simple nature, as we show in the next theorems (converses of these results hold, as we shall show in Theorems 4.2.1 and 6.2.1).

**Remark 2.1.3.** Here and subsequently, we shall comply with the use of the notation $H^m(K)$, in lieu of $H^m(\mathring{K})$. □

**Theorem 2.1.1.** *Assume that the inclusions* $P_K \subset H^1(K)$ *for all* $K \in \mathcal{T}_h$ *and* $X_h \subset \mathscr{C}^0(\bar{\Omega})$ *hold. Then the inclusions*

$$X_h \subset H^1(\Omega),$$
$$X_{oh} = \{v_h \in X_h;\quad v_h = 0 \quad \text{on} \quad \Gamma\} \subset H_0^1(\Omega),$$

*hold.*

**Proof.** Let a function $v \in X_h$ be given. We already know that it is in the space $L^2(\Omega)$. Therefore, by definition of the space $H^1(\Omega)$, we must find for each $i = 1, \ldots, n$, a function $v_i \in L^2(\Omega)$ such that

$$\forall \phi \in \mathscr{D}(\Omega), \int_\Omega v_i\phi \,\mathrm{d}x = -\int_\Omega v\partial_i\phi \,\mathrm{d}x.$$

For each $i$, a natural candidate is the function whose restriction to each finite element $K$ is the function $\partial_i(v|_K)$. Since each finite element $K$ has a Lipschitz-continuous boundary $\partial K$, we may apply Green's formula (1.2.4): For each $K \in \mathcal{T}_h$,

$$\int_K \partial_i(v|_K)\phi \, \mathrm{d}x = -\int_K v|_K \partial_i \phi \, \mathrm{d}x + \int_{\partial K} v|_K \phi \nu_{i,K} \, \mathrm{d}\gamma,$$

where $\nu_{i,K}$ is the $i$-th component of the unit outer normal vector along $\partial K$. By summing over all finite elements, we obtain

$$\int_\Omega v_i \phi \, \mathrm{d}x = -\int_\Omega v \partial_i \phi \, \mathrm{d}x + \sum_{K \in \mathcal{T}_h} \int_{\partial K} v|_K \phi \nu_{i,K} \, \mathrm{d}\gamma,$$

and the proof follows if we notice that the sum $\Sigma_{K \in \mathcal{T}_h} \int_{\partial K} v|_K \phi \nu_{i,K} \, \mathrm{d}\gamma$ vanishes: Either a portion of $\partial K$ is a portion of the boundary $\Gamma$ of $\Omega$ in which case $\phi = 0$ along this portion, or the contribution of adjacent elements is zero.

The boundary $\Gamma$ being Lipschitz-continuous by assumption, the second inclusion follows from the characterization

$$H_0^1(\Omega) = \{v \in H^1(\Omega), \quad v = 0 \quad \text{on} \quad \Gamma\},$$

which was mentioned in Section 1.2. □

Assuming Theorem 2.1.1 applies, we shall therefore use the finite element space $V_h = X_{oh}$ if we are solving a second-order homogeneous Dirichlet problem, or $V_h = X_h$ if we are solving a second-order homogeneous or nonhomogeneous Neumann problem.

The proof of the next theorem is similar to that of Theorem 2.1.1 and, for this reason, is left to the reader as an exercise (Exercise 2.1.1).

**Theorem 2.1.2.** *Assume that the inclusions $P_K \subset H^2(K)$ for all $K \in \mathcal{T}_h$ and $X_h \subset \mathscr{C}^1(\bar{\Omega})$ hold. Then the inclusions*

$$X_h \subset H^2(\Omega),$$
$$X_{oh} = \{v_h \in X_h; \quad v_h = 0 \quad \text{on} \quad \Gamma\} \subset H^2(\Omega) \cap H_0^1(\Omega),$$
$$X_{ooh} = \{v_h \in X_h; \quad v_h = \partial_\nu v_h = 0 \quad \text{on} \quad \Gamma\} \subset H_0^2(\Omega),$$

*hold.* □

Thus if we are to solve a simply supported plate problem, or a clamped plate problem, we shall use the finite element space $V_h = X_{oh}$, or the finite element space $V_h = X_{ooh}$, respectively, as given in the previous theorem.

Let us return to the description of the finite element method.

(FEM 2) *The second basic aspect* of the finite element method is that *the spaces $P_K$, $K \in \mathcal{T}_h$, contain polynomials, or, at least, contain functions which are "close to" polynomials.* At this stage, we cannot be too specific about the underlying reasons for this aspect of the method but at least, we can say that

(i) it is the key to all convergence results as we shall see, and

(ii) it yields simple computations of the coefficients of the resulting linear system (see (2.1.4) below).

Let us now briefly examine how the discrete problem (2.1.2) is solved in practice. Let $(w_k)_{k=1}^M$ be a basis in the space $V_h$. Then *the solution $u_h = \sum_{k=1}^M u_k w_k$ of problem* (2.1.2) *is such that the coefficients $u_k$ are solutions of the linear system*

$$\sum_{k=1}^{M} a(w_k, w_l)u_k = f(w_l), \quad 1 \leq l \leq M, \tag{2.1.4}$$

*whose matrix is always invertible,* since the bilinear form, being assumed to be $V$-elliptic, is *a fortiori* $V_h$-elliptic. By reference to the elasticity problem, the matrix $(a(w_k, w_l))$ and the vector $(f(w_l))$ are often called the *stiffness matrix* and the *load vector,* respectively.

In the choice of the basis $(w_k)_{k=1}^M$, it is of paramount importance, *from a numerical standpoint,* that *the resulting matrix possess as many zeros as possible.*

For *all* the examples which were considered in Section 1.2 the coefficients $a(w_k, w_l)$ are *integrals* of a specific form: For instance, in the case of the first examples, one has

$$a(w_k, w_l) = \int_\Omega \left(\sum_{i=1}^{n} \partial_i w_k \partial_i w_l + a w_k w_l\right) dx,$$

so that a coefficient $a(w_k, w_l)$ vanishes whenever the $dx$-measure of the intersection of the supports of the basis functions $w_k$ and $w_l$ is zero.

(FEM 3) As a consequence, we shall consider as *the third basic aspect* of the finite element method that *there exists at least one "canonical" basis in the space $V_h$ whose corresponding basis functions have supports which are as "small" as possible, it being implicitly understood that these basis functions can be easily described.*

**Remark 2.1.4.** When the bilinear form is symmetric, the matrix $(a(w_k, w_l))$ is symmetric and positive definite, which is an advantage for the numerical solution of the linear system (2.1.4). By contrast, this is not generally the case for standard finite-difference methods, except for rectangular domains.

Assuming again the symmetry of the bilinear form, one could conceivably start out with any given basis, and, using the Gram–Schmidt orthonormalization procedure, construct a new basis $(w_k^*)_{k=1}^M$ which is orthonormal with respect to the inner product $a(\cdot, \cdot)$. This is indeed an efficient way of getting a sparse matrix since the corresponding matrix $(a(w_k^*, w_l^*))$ is the identity matrix! However, this process is not recommended from a *practical* standpoint: For comparable computing times, it yields worse results than the solution by standard methods of the linear system corresponding to the "canonical" basis. □

It was mentioned earlier that the three basic aspects were characteristic of the finite element method *in its simplest form*. Indeed, *there are more general finite element methods*:

(i) One may start out with *more general variational problems*, such as variational inequalities (see Section 5.1) or various nonlinear problems (see Sections 5.2 and 5.3), or different variational formulations (see Chapter 7).

(ii) The space $V_h$, in which one looks for the discrete solution, may no longer be a subspace of the space $V$. This may happen when the boundary of the set $\Omega$ is curved, for instance. Then it cannot be exactly triangulated in general by standard finite elements and thus it is replaced by an approximate set $\Omega_h$ (see Section 4.4). This also happens when the functions in the space $V_h$ lack the proper continuity across adjacent finite elements (see the "nonconforming" methods described in Section 4.2 and Section 6.2).

(iii) Finally, the bilinear form and the linear form may be approximated. This is the case for instance when numerical integration is used for computing the coefficients of the linear system (2.1.4) (see Section 4.1), or for the shell problem (see Section 8.2).

Nevertheless, *it is characteristic of all these more general finite element methods that the three basic aspects are again present.*

To conclude these general considerations, we shall reserve the terminology *conforming finite element methods* for the finite element methods described at the beginning of this section, i.e., for which $V_h$ is a subspace of the space $V$, and the bilinear form and the linear form of the discrete problem are identical to the original ones.

*Exercises*

*2.1.1.* Prove Theorem 2.1.2.

*2.1.2.* The purpose of this problem is to give another proof of the Lax–Milgram lemma (Theorem 1.1.3; see also Exercise 1.1.2) in case the Hilbert space $V$ is separable. Otherwise the bilinear form and the linear form satisfy the same assumptions as in Theorem 1.1.3.

(i) Let $V_h$ be any finite-dimensional subspace of the space $V$, and let $u_h$ be the discrete solution of the associated discrete problem (2.1.2). Show that there exists a constant $C$ independent of the subspace $V_h$ such that $\|u_h\| \leq C$ (as usual, there is a simpler proof when the bilinear form is symmetric).

(ii) The space $V$ being separable, there exists a nested sequence $(V_\nu)_{\nu \in N}$ of finite-dimensional subspaces such that $(\bigcup_{\nu \in N} V_\nu)^- = V$. Let $(u_\nu)_{\nu \in N}$ be the sequence of associated discrete solutions. Show that there exists a subsequence of the sequence $(u_\nu)_{\nu \in N}$ which weakly converges to a solution $u$ of the original variational problem.

(iii) Show that the whole sequence converges in the norm of $V$ to the solution $u$.

(iv) Show that the Sobolev spaces $H^m(\Omega)$ are separable.

## 2.2. Examples of finite elements and finite element spaces

### *Requirements for finite element spaces*

Throughout this section, we assume that we are using a *conforming finite element method* for solving a second-order or a fourth-order boundary value problem. Let us first summarize the various requirements that a finite element space $X_h$ must satisfy, according to the discussion made in the previous section. Such a space is associated with a triangulation $\mathcal{T}_h$ of a set $\bar{\Omega} = \bigcup_{K \in \mathcal{T}_h} K$ (FEM 1), and for each finite element $K \in \mathcal{T}_h$, we define the space

$$P_K = \{v_h|_K; \quad v_h \in X_h\}. \tag{2.2.1}$$

Then the requirements are the following:

(i) For each $K \in \mathcal{T}_h$, the space $P_K$ should consist of functions which are polynomials or "nearly polynomials" (FEM 2).

(ii) By Theorems 2.1.1 and 2.1.2, inclusions such as $X_h \subset \mathscr{C}^0(\bar{\Omega})$ or $X_h \subset \mathscr{C}^1(\bar{\Omega})$ should hold, depending upon whether we are solving a second-order or a fourth-order problem. For the time being, *we shall ignore boundary conditions*, which we shall take into account in the next section only.

(iii) Finally, we must check that there exists one canonical basis in the space $X_h$, whose functions have "small" supports and are easy to describe (FEM 3).

In this section, we shall describe various finite elements $K$ which are all polyedra in $\mathbf{R}^n$, and which are sometimes called *straight finite elements*. As a consequence, *we have to restrict ourselves to problems which are posed over sets $\bar{\Omega}$ which are themselves polyedra*, in which case we shall say that the set $\Omega$ is *polygonal*.

*First examples of finite elements for second order problems: n-Simplices of type* ($k$), (3')

*We begin by examining examples for which the inclusion $X_h \subset \mathscr{C}^0(\bar{\Omega})$ holds, and which are the most commonly used by engineers for solving second-order problems with conforming finite element methods.* Inasmuch as such problems are most often found in mechanics of continua, it is clear that the value to be assigned in practice to the dimension $n$ in the forthcoming examples is either 2 or 3 (see the examples given in Section 1.2).

We equip the space $\mathbf{R}^n$ with its canonical basis $(e_i)_{i=1}^n$. For each integer $k \geq 0$, we shall denote by $P_k$ the space of all polynomials of degree $\leq k$ in the variables $x_1, x_2, \ldots, x_n$, i.e., a polynomial $p \in P_k$ is of the form

$$\begin{aligned} p : x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n &\to p(x) \\ &= \sum_{\sum_{i=1}^n \alpha_i \leq k} \gamma_{\alpha_1 \alpha_2 \cdots \alpha_n} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}, \end{aligned}$$

for appropriate coefficients $\gamma_{\alpha_1 \alpha_2 \cdots \alpha_n}$, or using the multi-index notation,

$$p : x \in \mathbf{R}^n \to p(x) = \sum_{|\alpha| \leq k} \gamma_\alpha x^\alpha .$$

The dimension of the space $P_k$ is given by

$$\dim P_k = \binom{n+k}{k}. \qquad (2.2.2)$$

If $\Phi$ is a space of functions defined over $\mathbf{R}^n$, and if $A$ is any subset of $\mathbf{R}^n$, we shall denote by $\Phi(A)$ the space formed by the restrictions to the set $A$ of the functions in the space $\Phi$. Thus, for instance, we shall let

$$P_k(A) = \{p|_A;\quad p \in P_k\}. \tag{2.2.3}$$

Notice that the dimension of the space $P_k(A)$ is the same as that of the space $P_k = P_k(\mathbf{R}^n)$ as long as the interior of the set $A$ is not empty.

In $\mathbf{R}^n$, a (nondegenerate) *$n$-simplex* is the convex hull $K$ of $(n+1)$ points $a_j = (a_{ij})_{i=1}^n \in \mathbf{R}^n$, which are called the *vertices* of the $n$-simplex, and which are such that the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \cdots & a_{1,n+1} \\ a_{21} & a_{22} \cdots & a_{2,n+1} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} \cdots & a_{n,n+1} \\ 1 & 1 \cdots & 1 \end{pmatrix} \tag{2.2.4}$$

is regular (equivalently, the $(n+1)$ points $a_j$ are not contained in a hyperplane). Thus, one has

$$K = \left\{x = \sum_{j=1}^{n+1} \lambda_j a_j;\quad 0 \leq \lambda_j \leq 1,\quad 1 \leq j \leq n+1,\quad \sum_{j=1}^{n+1} \lambda_j = 1\right\}. \tag{2.2.5}$$

Notice that a 2-simplex is a triangle and that a 3-simplex is a tetrahedron.

For any integer $m$ with $0 \leq m \leq n$, an *$m$-face* of the $n$-simplex $K$ is any $m$-simplex whose $(m+1)$ vertices are also vertices of $K$. In particular, any $(n-1)$-face is simply called a *face*, any 1-face is called an *edge*, or a *side*.

The *barycentric coordinates* $\lambda_j = \lambda_j(x)$, $1 \leq j \leq n+1$, of any point $x \in \mathbf{R}^n$, *with respect to the* $(n+1)$ *points* $a_j$, are the (unique) solutions of the linear system

$$\begin{cases} \sum_{j=1}^{n+1} a_{ij}\lambda_j = x_i, & 1 \leq i \leq n, \\ \sum_{j=1}^{n+1} \lambda_j = 1, \end{cases} \tag{2.2.6}$$

whose matrix is precisely the matrix $A$ of (2.2.4). By inspection of the linear system (2.2.6), one sees that *the barycentric coordinates are affine* *functions of* $x_1, x_2, \ldots, x_n$ (i.e., they belong to the space $P_1$):

$$\lambda_i = \sum_{j=1}^{n} b_{ij}x_j + b_{in+1}, \quad 1 \leq i \leq n+1, \tag{2.2.7}$$

where the matrix $B = (b_{ij})$ is the inverse of the matrix $A$.

The *barycenter*, or *center of gravity*, of an $n$-simplex $K$ is the point of $K$ whose all barycentric coordinates are equal to $1/(n+1)$.

To describe the first finite element, we need to prove that *a polynomial* $p: x \to \Sigma_{|\alpha|\leq 1} \gamma_\alpha x^\alpha$ *of degree* 1 *is uniquely determined by its values at the* $(n+1)$ *vertices* $a_j$ *of any* $n$*-simplex in* $\mathbf{R}^n$. To see this, it suffices to show that the linear system

$$\sum_{|\alpha|\leq 1} \gamma_\alpha (a_j)^\alpha = \mu_j, \quad 1 \leq j \leq n+1,$$

has one and only one solution $(\gamma_\alpha, |\alpha| \leq 1)$ for all right-hand sides $\mu_j$, $1 \leq j \leq n+1$. Since $\dim P = \operatorname{card}(\bigcup_{j=1}^{n+1}\{a_j\}) = n+1$, the matrix of this linear system is *square*, and therefore it suffices to prove either *uniqueness* or *existence*. In this case, the existence is clear: The barycentric coordinates verify $\lambda_i(a_j) = \delta_{ij}$, $1 \leq i, j \leq n+1$, and thus the polynomial

$$x \in \mathbf{R}^n \to \sum_{i=1}^{n+1} \mu_i \lambda_i(x)$$

has the desired property. As a consequence, we have the *identity*

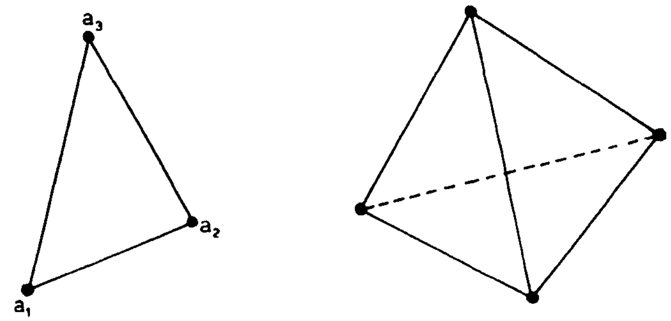
$$\forall p \in P_1, p = \sum_{i=1}^{n+1} p(a_i)\lambda_i. \tag{2.2.8}$$

*Although we shall not repeat this argument in the sequel, it will be often implicitly used.*

A polynomial $p \in P_1$ being completely determined by its values $p(a_i)$, $1 \leq i \leq n+1$, we can now define the simplest finite element, which we shall call *$n$-simplex of type* (1): The set $K$ is an $n$-simplex with vertices $a_i$, $1 \leq i \leq n+1$, the space $P_K$ is the space $P_1(K)$, and the *degrees of freedom of the finite element*, i.e., those parameters which uniquely define a function in the space $P_K$, consist of the values at the vertices. Denoting by $\Sigma_K$ the corresponding *set of degrees of freedom*, we shall write symbolically

$$\Sigma_K = \{p(a_i), \quad 1 \leq i \leq n+1\}.$$

In Fig. 2.2.1, we have recorded the main characteristics of this finite element for arbitrary $n$, along with the figures in the special cases $n = 2$ and 3. In case $n = 2$, this element is also known as *Courant's triangle* (see the section "Bibliography and Comments").



$\left\{\begin{array}{l}\textit{Courant's triangle, or}\\ \textit{triangle of type } (1)\\ \dim P_K = 3\end{array}\right.$

$\left\{\begin{array}{l}\textit{tetrahedron of type } (1)\\ \dim P_K = 4\end{array}\right.$

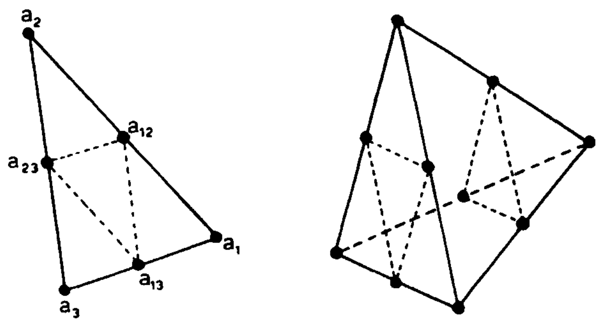
| *n-simplex of type* (1) |
|---|
| $P_K = P_1(K)$; $\dim P_K = (n+1)$;<br>$\Sigma_K = \{p(a_i),\ 1 \leqslant i \leqslant n+1\}$. |

Fig. 2.2.1

Let us call $a_{ij} = \frac{1}{2}(a_i + a_j)$, $1 \leqslant i < j \leqslant n+1$, the mid-points of the edges of the $n$-simplex $K$. Since $\lambda_k(a_{ij}) = \frac{1}{2}(\delta_{ki} + \delta_{kj})$, $1 \leqslant i < j \leqslant n+1$, $1 \leqslant k \leqslant n+1$, we obtain the identity (where, here and subsequently, indices $i, j, k, \ldots$, are always assumed to take all possible values in the set $\{1, 2, \ldots, n\}$ whenever this fact is not specified)

$$\forall p \in P_2, \quad p = \sum_i \lambda_i(2\lambda_i - 1)p(a_i) + \sum_{i<j} 4\lambda_i\lambda_j p(a_{ij}), \tag{2.2.9}$$

which yields the definition of a finite element, called the *n-simplex of type* (2): the space $P_K$ is $P_2(K)$, and the set $\Sigma_K$ consists of the values at the vertices and at the mid-points of the edges (Fig. 2.2.2).



$\begin{cases}\textit{triangle of type } (2)\\ \dim P_K = 6\end{cases}$ $\begin{cases}\textit{tetrahedron of type } (2)\\ \dim P_K = 10\end{cases}$

| *n-simplex of type* (2) |
|---|
| $P_K = P_2(K)$; $\dim P_K = \dfrac{(n+1)(n+2)}{2}$; |
| $\Sigma_K = \{p(a_i),\ 1 \leq i \leq n+1;\ p(a_{ij}),\ 1 \leq i < j \leq n+1\}$. |

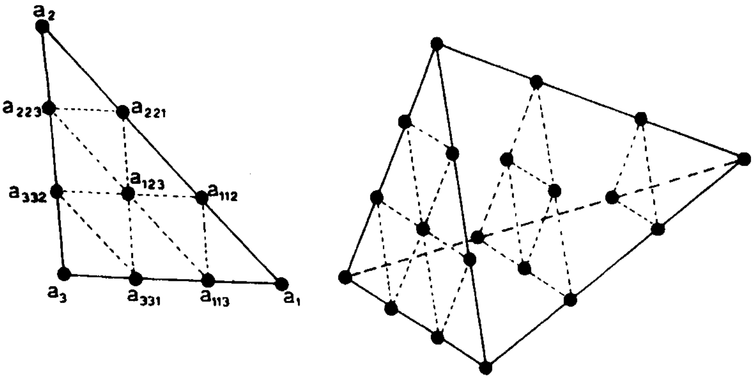
Fig. 2.2.2

Let $a_{iij} = \frac{1}{3}(2a_i + a_j)$ for $i \neq j$, and $a_{ijk} = \frac{1}{3}(a_i + a_j + a_k)$ for $i < j < k$. From the identity

$$\forall p \in P_3,\quad p = \sum_i \frac{\lambda_i(3\lambda_i - 1)(3\lambda_i - 2)}{2} p(a_i) + \sum_{i \neq j} \frac{9\lambda_i\lambda_j(3\lambda_i - 1)}{2} p(a_{iij}) + \sum_{i<j<k} 27\lambda_i\lambda_j\lambda_k p(a_{ijk}), \tag{2.2.10}$$

we deduce the definition of the *n-simplex of type* (3) (Fig. 2.2.3).

One may define analogous finite elements with polynomials of arbitrary degree, but they are not often used. In this respect, we leave to the reader the proof of the following theorem (Exercise 2.2.2), from which for any integer $k$, the definition of the *n-simplex of type* ($k$) can be easily derived.



$\left\{\begin{array}{l}\textit{triangle of type } (3)\\ \dim P_K = 10\end{array}\right.$ $\left\{\begin{array}{l}\textit{tetrahedron of type } (3)\\ \dim P_K = 20\end{array}\right.$

*n-simplex of type* (3)

$P_K = P_3(K)$; $\dim P_K = \dfrac{(n+1)(n+2)(n+3)}{6}$;

$\Sigma_K = \{p(a_i),\ 1 \leqslant i \leqslant n+1;\ p(a_{iij}),\ 1 \leqslant i, j \leqslant n+1,\ i \neq j;$
$p(a_{ijk}),\ 1 \leqslant i < j < k \leqslant n+1\}.$

Fig. 2.2.3

**Theorem 2.2.1.** *Let $K$ be an $n$-simplex with vertices $a_j$, $1 \leqslant j \leqslant n+1$. Then for a given integer $k \geqslant 1$, any polynomial $p \in P_k$ is uniquely determined by its values on the set*

$$L_k(K) = \left\{x = \sum_{j=1}^{n+1} \lambda_j a_j;\ \sum_{j=1}^{n+1} \lambda_j = 1,\right.$$
$$\left.\lambda_j \in \left\{0, \frac{1}{k}, \ldots, \frac{k-1}{k}, 1\right\},\quad 1 \leqslant j \leqslant n+1\right\}. \tag{2.2.11}$$

□

Let us now examine a modification of the $n$-simplex of type (3), in which the degrees of freedom $p(a_{ijk})$ are no longer present, and which is often preferred by the engineers to the previous element. To describe the corresponding finite element, we need the following result.

**Theorem 2.2.2.** *For each triple $(i, j, k)$ with $i < j < k$, let*

$$\phi_{ijk}(p) = 12p(a_{ijk}) + 2 \sum_{l=i,j,k} p(a_l) - 3 \sum_{\substack{l,m=i,j,k \\ l \neq m}} p(a_{llm}). \tag{2.2.12}$$

*Then any polynomial in the space*

$$P_3' = \{p \in P_3;\quad \phi_{ijk}(p) = 0,\quad 1 \leq i < j < k \leq n+1\} \tag{2.2.13}$$

*is uniquely determined by its values at the vertices $a_i$, $1 \leq i \leq n+1$, and at the points $a_{iij}$, $1 \leq i, j \leq n+1$, $i \neq j$. In addition, the inclusion*

$$P_2 \subset P_3' \tag{2.2.14}$$

*holds.*

**Proof.** The $\binom{n+1}{3}$ degrees of freedom $\phi_{ijk}$ are linearly independent (since $\phi_{ijk}(p) = 12p(a_{ijk}) + \cdots$) and thus, the dimension of the space $P_3'$ is

$$\dim P_3' = \dim P_3 - \binom{n+1}{3} = (n+1)^2,$$

i.e., precisely the number of degrees of freedom. Using the identity (2.2.10), and arguing as before, we obtain the identity

$$\forall p \in P_3',\quad p = \sum_i \Big(\frac{1}{2}\lambda_i(3\lambda_i - 1)(3\lambda_i - 2) - \frac{9}{2}\lambda_i \sum_{\substack{j<k \\ j,k \neq i}} \lambda_j\lambda_k\Big)p(a_i) + \\ + \sum_{i \neq j} \Big(\frac{9}{2}\lambda_i\lambda_j(3\lambda_i - 1) + \frac{27}{4}\lambda_i\lambda_j \sum_{k \neq i,j} \lambda_k\Big)p(a_{iij}), \tag{2.2.15}$$

which proves the first part of the theorem.

To prove that the inclusion (2.2.14) holds, let $p$ be a polynomial of degree $\leq 2$ and let $A \in \mathscr{L}_2(\mathbf{R}^n; \mathbf{R})$ be its second derivative (which is constant). From the expansions

$$p(a_l) = p(a_{ijk}) + Dp(a_{ijk})(a_l - a_{ijk}) + \tfrac{1}{2}A(a_l - a_{ijk})^2,\quad l \in I,$$

valid for any triple $I = \{i, j, k\}$ with $i < j < k$, we deduce

$$\sum_{l \in I} p(a_l) = 3p(a_{ijk}) + \frac{1}{2}\sum_{l \in I} A(a_l - a_{ijk})^2,$$

since $\Sigma_{l\in I}(a_l - a_{ijk}) = 0$. Likewise, from the expansions

$$p(a_{llm}) = p(a_{ijk}) + Dp(a_{ijk})(a_{llm} - a_{ijk}) + \tfrac{1}{2}A(a_{llm} - a_{ijk})^2, \qquad l, m \in I, \quad l \neq m,$$

we deduce

$$\sum_{\substack{l,m\in I\\ l\neq m}} p(a_{llm}) = 6p(a_{ijk}) + \frac{1}{2}\sum_{\substack{l,m\in I\\ l\neq m}} A(a_{llm} - a_{ijk})^2,$$

taking into account that $a_{ijk} = \frac{1}{2}(a_{iij} + a_{kkj}) = \frac{1}{2}(a_{jjk} + a_{iik}) = \frac{1}{2}(a_{kki} + a_{jji})$. Because $A$ is a linear mapping, and because

$$a_{llm} - a_{ijk} = \tfrac{1}{3}(2(a_l - a_{ijk}) + (a_m - a_{ijk})),$$

we can write

$$\sum_{l\in I} A(a_l - a_{ijk})^2 - \frac{3}{2}\sum_{\substack{l,m\in I\\ l\neq m}} A(a_{llm} - a_{ijk})^2 = -\frac{2}{3}A\Big(\sum_{l\in I}(a_l - a_{ijk})\Big)^2 = 0,$$

and the proof is complete. □

From Theorem 2.2.2 we deduce the definition of the *n-simplex of type* (3′) (Fig. 2.2.4).

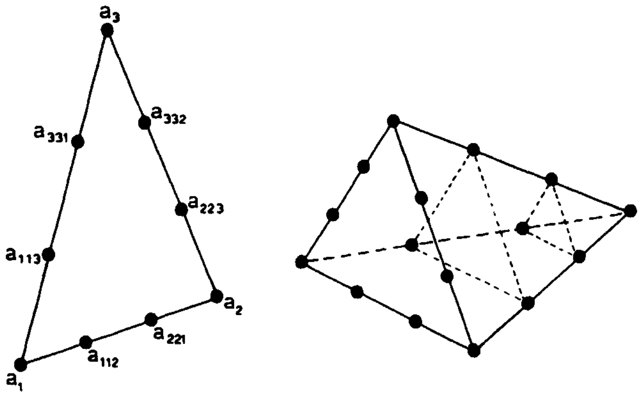
### *Assembly in triangulations. The associated finite element spaces*

Next we examine the question of constructing triangulations, using anyone of the finite elements previously described. Being non degenerate $n$-simplices, these have non empty interiors and Lipschitz-continuous boundaries, and therefore properties $(\mathcal{T}_h2)$ and $(\mathcal{T}_h4)$ are satisfied. To construct triangulations in the sense understood in Section 2.2, we shall write $\bar{\Omega} = \bigcup_{K\in\mathcal{T}_h} K$ in such a way that the $n$-simplices have piecewise disjoint interiors (cf. properties $(\mathcal{T}_h1)$ and $(\mathcal{T}_h3)$). In order to satisfy inclusions such as $X_h \subset \mathscr{C}^0(\bar{\Omega})$ (and $X_h \subset \mathscr{C}^1(\bar{\Omega})$ later on), we shall impose a fifth condition on a triangulation made up of $n$-simplices, namely:

$(\mathcal{T}_h5)$ *Any face of any n-simplex $K_1$ in the triangulation is either a subset of the boundary $\Gamma$, or a face of another n-simplex $K_2$ in the triangulation.*

In the second case, the $n$-simplices $K_1$ and $K_2$ are said to be *adjacent*.

An example of a triangulation for $n = 2$ is given in Fig. 2.2.5, while



$\begin{cases}\textit{triangle of type } (3') \\ \dim P_K = 9.\end{cases}$

$\begin{cases}\textit{tetrahedron of type } (3') \\ \dim P_K = 16.\end{cases}$

| *n-simplex of type* (3') |
|---|
| $P_K = P_3'(K)$ (cf. (2.2.13)); $\dim P_K = (n+1)^2$;<br>$\Sigma_K = \{p(a_i),\ 1 \leqslant i \leqslant n+1,\ p(a_{iij}),\ 1 \leqslant i,j \leqslant n+1,\ i \neq j\}.$ . |

Fig. 2.2.4

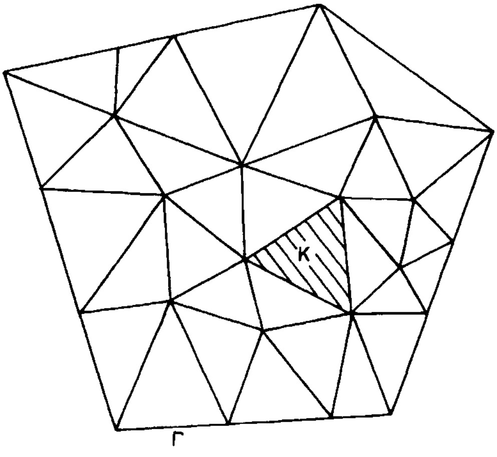


Fig. 2.2.5

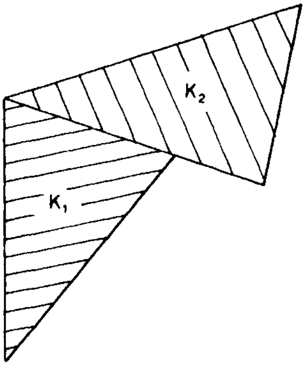


Fig. 2.2.6.

Fig. 2.2.6 shows an example of a "forbidden situation" since the intersection of $K_1$ and $K_2$ is not an edge of $K_2$.

Given a triangulation $\mathcal{T}_h$, we associate in a natural way a finite element space $X_h$ of functions $v_h$: $\bar{\Omega} \rightarrow \mathbf{R}$ with each type of finite element:

With $n$-simplices of type (1), a function $v_h \in X_h$

(i) is such that each restriction $v_h|_K$ is in the space $P_K = P_1(K)$ for each $K \in \mathcal{T}_h$, and

(ii) is completely determined by its values at all the vertices of the triangulation.

Likewise, with $n$-simplices of type (2), a function of $X_h$

(i) is in the space $P_K = P_2(K)$ for each $K \in \mathcal{T}_h$, and

(ii) is completely determined by its values at all the vertices and all the mid-points of the edges of the triangulation.

Similar constructions hold for $n$-simplices of type (3) or (3').

In all cases, a function $v_h$ in the space $X_h$ is seen to be determined by *degrees of freedom* which make up a set of the form

$$\Sigma_h = \{v_h(b); \quad b \in \mathcal{N}_h\}, \tag{2.2.16}$$

where $\mathcal{N}_h$ is a finite subset of $\bar{\Omega}$. The set $\Sigma_h$ is the *set of degrees of freedom of the finite element space* $X_h$.

One should observe that *if there is no ambiguity in the definition of the degrees of freedom across adjacent finite elements, it is precisely because we have satisfied requirement* ($\mathcal{T}_h$5). This requirement also plays a crucial role in the proof of the following result.

**Theorem 2.2.3.** *Let $X_h$ be the finite element space associated with n-simplices of type (k) for any integer $k \geqslant 1$ or with n-simplices of type (3'). Then the inclusion*

$$X_h \subset \mathscr{C}^0(\bar{\Omega}) \cap H^1(\Omega)$$

*holds.*

**Proof.** We shall give the proof in case $n = 2$ and for triangles of type (2), leaving the other cases as a problem (Exercise 2.2.3). Given a function $v_h$ in the space $X_h$, consider the two functions $v_h|_{K_1}$ and $v_h|_{K_2}$ along the common side $K' = [b_i, b_j]$ of two adjacent triangles $K_1$ and $K_2$ (Fig. 2.2.7). Let $t$ denote an abscissa along the axis containing the segment $K'$. Considered as functions of $t$, the two functions $v_h|_{K_1}$ and $v_h|_{K_2}$ are quadratic polynomials along $K'$, whose values coincide at the three points $b_i$, $b_j$, $b_{ij} = (b_i + b_j)/2$. Therefore these polynomials are identical, and the inclusion $X_h \subset \mathscr{C}^0(\bar{\Omega})$ holds. Finally the inclusion $X_h \subset H^1(\Omega)$ is a consequence of Theorem 2.1.1. □

It remains to verify requirement (FEM 3), i.e., that there is indeed a canonical choice for basis functions with small supports. In each case,

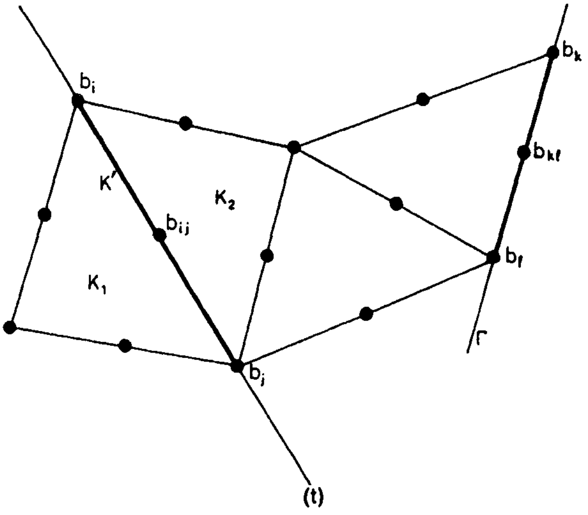


Fig. 2.2.7

the set $\Sigma_h$ of degrees of freedom of the space is of the form

$$\Sigma_h = \{(b_k); \quad 1 \leq k \leq M\}. \tag{2.2.17}$$

If we define functions $w_k$, $1 \leq k \leq M$, by the conditions

$$w_k \in X_h \quad \text{and} \quad w_k(b_l) = \delta_{kl}, \quad 1 \leq k,l \leq M, \tag{2.2.18}$$

it is seen that (i) such functions form a basis of the space $X_h$ and that (ii) they have "small" supports. In Fig. 2.2.8, we have represented the three types of supports which are encountered when triangles of type (3) are employed, for instance.

### *n-Rectangles of type (k). Rectangles of type* (2'), (3'). *Assembly in triangulations*

Before we turn to a second category of finite elements, we need a few definitions. For each integer $k \geq 0$, we shall denote by $Q_k$ the space of all polynomials which are of degree $\leq k$ with respect to each one of the $n$ variables $x_1, x_2, \ldots, x_n$, i.e., a polynomial $p \in Q_k$ is of the form

$$p: x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n \to p(x) = \sum_{\substack{\alpha_i \leq k,\\ 1 \leq i \leq n}} \gamma_{\alpha_1\alpha_2\cdots\alpha_n} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n},$$

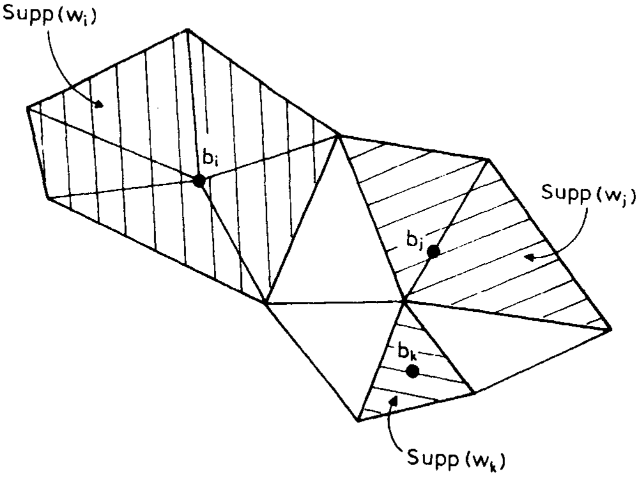


Fig. 2.2.8

for appropriate coefficients $\gamma_{\alpha_1\alpha_2\cdots\alpha_n}$. The dimension of the space $Q_k$ is given by

$$\dim Q_k = (k+1)^n \tag{2.2.19}$$

and the inclusions

$$P_k \subset Q_k \subset P_{nk} \tag{2.2.20}$$

hold.

Notice that the dimension of the space $Q_k(A)$ is the same as that of the space $Q_k = Q_k(\mathbf{R}^n)$ as long as the interior of the set $A \subset \mathbf{R}^n$ is not empty.

**Theorem 2.2.4.** *A polynomial $p \in Q_k$ is uniquely determined by its values on the set*

$$\hat{M}_k = \left\{x = \left(\frac{i_1}{k}, \frac{i_2}{k}, \ldots, \frac{i_n}{k}\right) \in \mathbf{R}^n;\quad i_j \in \{0, 1, \ldots, k\},\quad 1 \leq j \leq n\right\}. \tag{2.2.21}$$

**Proof.** It suffices to use the identity

$$\forall p \in Q_k,\quad p = \sum_{\substack{0 \leq i_j \leq k \\ 1 \leq j \leq n}} \prod_{j=1}^{n} \Biggl(\prod_{\substack{i'_j=0 \\ i'_j \neq i_j}}^{k} \frac{kx_j - i'_j}{i_j - i'_j}\Biggr) p\left(\frac{i_1}{k}, \frac{i_2}{k}, \ldots, \frac{i_n}{k}\right). \qquad \square \tag{2.2.22}$$

In $\mathbf{R}^n$, an *$n$-rectangle*, or simply a *rectangle* if $n = 2$, is a set of the form

$$K = \prod_{i=1}^{n} [a_i, b_i] = \{x = (x_1, x_2, \ldots, x_n);\quad a_i \leq x_i \leq b_i,\quad 1 \leq i \leq n\}, \tag{2.2.23}$$

with $-\infty < a_i < b_i < +\infty$ for each $i$, i.e., it is a product of compact intervals with non-empty interiors. A *face* of $K$ is any one of the sets

$$\{a_j\} \times \prod_{\substack{i=1 \\ i \neq j}}^{n} [a_i, b_i] \quad \text{or} \quad \{b_j\} \times \prod_{\substack{i=1 \\ i \neq j}}^{n} [a_i, b_i],\quad 1 \leq j \leq n,$$

while an *edge* of $K$, also called a *side*, is any one of the sets

$$[a_j, b_j] \times \prod_{\substack{i=1 \\ i \neq j}}^{n} \{c_i\},$$

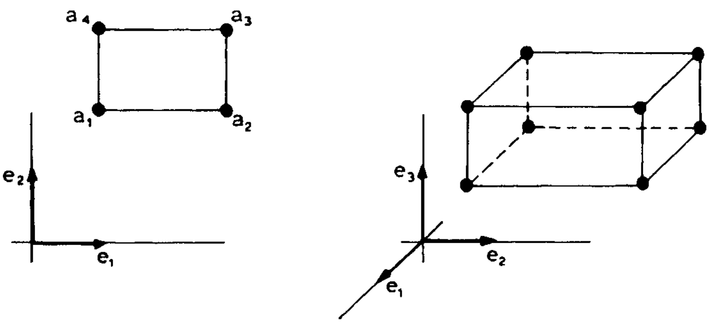
with $c_i = a_i$ or $b_i$, $1 \leqslant i \leqslant n$, $i \neq j$, $1 \leqslant j \leqslant n$. A *vertex* of $K$ is any point $x = (x_1, x_2, \dots, x_n)$ of $K$ with $x_i = a_i$ or $b_i$, $1 \leqslant i \leqslant n$.

Observe that the set $\hat{M}_k$ of (2.2.21) is a subset of a particular $n$-rectangle, namely the *unit hypercube* $[0, 1]^n$. Then, given any $n$-rectangle $K$, we infer that a polynomial $p \in Q_k$ is uniquely determined by its values on the subset

$$M_k(K) = F_K(\hat{M}_k) \tag{2.2.24}$$

of the $n$-rectangle $K$, where $F_K$ is a *diagonal affine mapping*, i.e., of the form $F_K: x \in \mathbf{R}^n \to F_K(x) = B_K x + b_K$, with $b_K$ a vector in $\mathbf{R}^n$ and $B_K$ an $n \times n$ diagonal matrix, such that $K = F_K([0, 1]^n)$. From this, we deduce the definition of finite elements, called *$n$-rectangles of type* $(k)$.

Just as in the case of $n$-simplices, the values $k = 1, 2$ or $3$ are the most commonly encountered. In Fig. 2.2.9, 2.2.10 and 2.2.11, the corresponding elements are represented for $n = 2$ and 3, and the numbering of the points occurring in the sets of degrees of freedom is also indicated for $n = 2$.

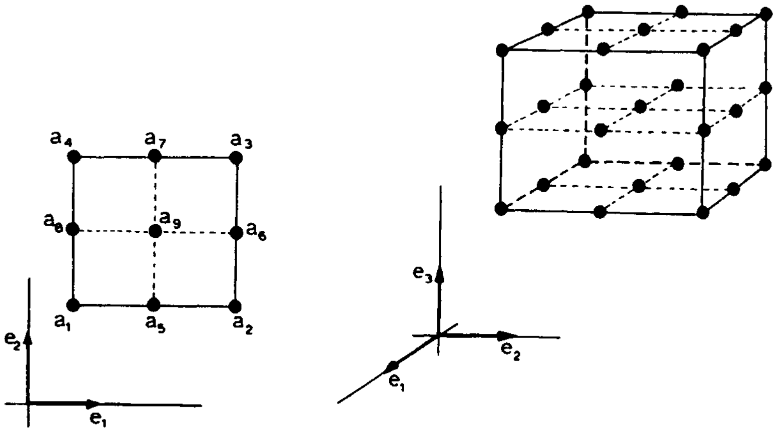


$\begin{cases}\textit{rectangle of type } (1)\\ \dim P_K = 4\end{cases}$

$\begin{cases}\textit{3-rectangle of type } (1)\\ \dim P_K = 8\end{cases}$

| *n-rectangle of type* (1) |
|---|
| $P_K = Q_1(K)$; $\dim P_K = 2^n$;<br>$\Sigma_K = \{p(a);\ a \in M_1(K)\}$ (cf. (2.2.24)) |

Fig. 2.2.9



$\begin{cases}\textit{rectangle of type } (2)\\ \dim P_K = 9\end{cases}$ $\begin{cases}3\textit{-rectangle of type } (2)\\ \dim P_K = 27\end{cases}$

| *n-rectangle of type* (2) |
|---|
| $P_K = Q_2(K)$; $\dim P_K = 3^n$;<br>$\Sigma_K = \{p(a),\ \ a \in M_2(K)\}$ (cf. (2.2.24)). |

Fig. 2.2.10

For the numbering of the nodes when $n = 2$, we have followed this rule: Assuming, without loss of generality, that the set $K$ is the *unit square* $[0, 1]^2$, four points are consecutively numbered if they are the vertices of a square centered at the point $(\frac{1}{2}, \frac{1}{2})$. This rule allows for particularly simple expressions of the corresponding functions $p_i$ appearing in identities of the form

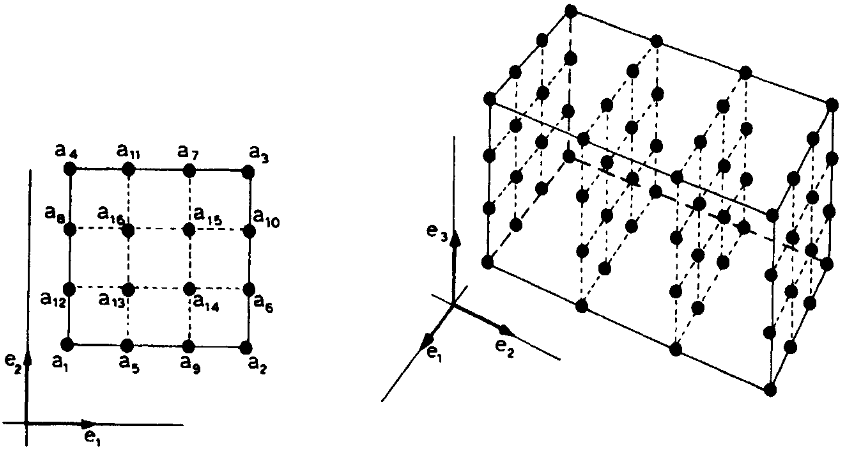
$$\forall p \in Q_k, \quad p = \sum_i p(a_i)p_i,$$

which are special cases (for $k = 1, 2, 3$ and $n = 2$) of the identity (2.2.22). Notice that the coordinates of a given point with respect to the four vertices $a_i$, $1 \leq i \leq 4$, of the unit square are

$$(x_1, x_2), \quad (x_2, 1 - x_1), \quad (1 - x_1, 1 - x_2), \quad (1 - x_2, x_1),$$

respectively. Then, if we introduce the variables

$$x_3 = 1 - x_1, \quad x_4 = 1 - x_2, \tag{2.2.25}$$



$\begin{cases}\textit{rectangle of type } (3)\\ \dim P_K = 16\end{cases}$ $\begin{cases}3\textit{-rectangle of type } (3)\\ \dim P_K = 64\end{cases}$

| *n-rectangle of type* (3) |
|---|
| $P_K = Q_3(K)$; $\dim P_K = 4^n$:<br>$\Sigma_K = \{p(a);\ \ a \in M_3(K)\}$ (cf. (2.2.24)). |

Fig. 2.2.11

the four functions $p_i$ are obtained through circular permutations of the variables $x_1, x_2, x_3, x_4$ (such permutations correspond to rotations of $+\pi/2$ around the point $(\frac{1}{2}, \frac{1}{2})$).

Corresponding to the unit square of type (1) (recall that $K = [0, 1]^2$), we have the identity

$$\forall p \in Q_1, \quad p = \sum_{i=1}^{4} p(a_i)p_i,$$

with

$$p_1 = (1 - x_1)(1 - x_2), \quad p_2 = x_1(1 - x_2), \quad p_3 = x_1x_2,$$
$$p_4 = (1 - x_1)x_2.$$

We may thus condense these expressions in

$$p_1 = x_3x_4, \ldots. \tag{2.2.26}$$

Likewise, corresponding to the unit square of type (2), we have the identity

$$\forall p \in Q_2, \quad p = \sum_{i=1}^{9} p(a_i)p_i,$$

with

$$\left.\begin{aligned} p_1 &= x_3(2x_3-1)x_4(2x_4-1), \ldots \\ p_5 &= -4x_3(x_3-1)x_4(2x_4-1), \ldots \\ p_9 &= 16x_1x_2x_3x_4, \end{aligned}\right\} \qquad (2.2.27)$$

using the above rule. Finally, corresponding to the unit square of type (3), we have

$$\left.\begin{aligned} p_1 &= \frac{1}{4}x_3(3x_3-1)(3x_3-2)x_4(3x_4-1)(3x_4-2), \ldots \\ p_5 &= -\frac{9}{4}x_3(3x_3-1)(x_3-1)x_4(3x_4-1)(3x_4-2), \ldots \\ p_9 &= \frac{9}{4}x_3(3x_3-2)(x_3-1)x_4(3x_4-1)(3x_4-2), \ldots \\ p_{13} &= \frac{81}{4}x_3(3x_3-1)(x_3-1)x_4(3x_4-1)(x_4-1), \ldots . \end{aligned}\right\} \qquad (2.2.28)$$

**Remark 2.2.1.** The inconsistency for the notations $a_i$, $5 \leq i \leq 9$, between the rectangles of type (2) and (3), avoids the introduction of a new letter. □

In analogy with the $n$-simplices of type (3'), one can derive two finite elements, in which the "internal" values of the rectangle of type (2) or (3) are no longer degrees of freedom (for simplicity, we shall restrict ourselves to the case $n = 2$). The existence of these finite elements is a consequence of the following two theorems.

**Theorem 2.2.5.** *Let the points $a_i$, $1 \leq i \leq 9$, be as in Fig. 2.2.10. Then any polynomial in the space*

$$Q_2' = \left\{p \in Q_2;\quad 4p(a_9) + \sum_{i=1}^{4} p(a_i) - 2\sum_{i=5}^{8} p(a_i) = 0\right\} \qquad (2.2.29)$$

*is uniquely determined by its values at the points $a_i$, $1 \leq i \leq 8$. In addition,* *the inclusion*

$$P_2 \subset Q_2' \tag{2.2.30}$$

*holds.*

**Proof.** The first part of the proof is similar to the first part of the proof of Theorem 2.2.2. In particular, we have the identity

$$\forall p \in Q_2', \quad p = \sum_{i=1}^{8} p(a_i)p_i,$$

with

$$\left.\begin{aligned} p_1 &= x_3x_4(2x_3 + 2x_4 - 3), \ldots \\ p_5 &= -4x_3x_4(x_3 - 1), \ldots . \end{aligned}\right\} \tag{2.2.31}$$

To prove the inclusion (2.2.30), let $p$ be a polynomial of degree 2, and let $A$ denote its (constant) second derivative. From the expansions

$$p(a_i) = p(a_9) + Dp(a_9)(a_i - a_9) + \frac{1}{2}A(a_i - a_9)^2, \quad 1 \leq i \leq 8,$$

we deduce

$$\sum_{i=1}^{4} p(a_i) = 4p(a_9) + \frac{1}{2}\sum_{i=1}^{4} A(a_i - a_9)^2,$$

$$\sum_{i=5}^{8} p(a_i) = 4p(a_9) + \frac{1}{2}\sum_{i=5}^{8} A(a_i - a_9)^2,$$

since

$$\sum_{i=1}^{4} (a_i - a_9) = \sum_{i=5}^{8} (a_i - a_9) = 0.$$

Because the mapping $A$ is bilinear, and because $a_5 = (a_1 + a_2)/2, \ldots$, we obtain

$$\sum_{i=5}^{8} A(a_i - a_9)^2 = \frac{1}{2}\sum_{i=1}^{4} A(a_i - a_9)^2.$$

Combining the previous relations, we deduce that

$$4p(a_9) + \sum_{i=1}^{4} p(a_i) - 2\sum_{i=5}^{8} p(a_i) = 0,$$

and the proof is complete. □

**Theorem 2.2.6.** *Let the points $a_i$, $1 \leqslant i \leqslant 16$, be as in Fig.* 2.2.11. *Define the space*

$$Q_3' = \{p \in Q_3;\quad \psi_i(p) = 0,\quad 1 \leqslant i \leqslant 4\}, \tag{2.2.32}$$

*where*

$$\begin{aligned}\psi_1(p) = {} & 9p(a_{13}) + 4p(a_1) + 2p(a_2) + p(a_3) + 2p(a_4) \\ & -6p(a_5) - 3p(a_6) - 3p(a_{11}) - 6p(a_{12}),\end{aligned} \tag{2.2.33}$$

*and $\psi_2(p)$, $\psi_3(p)$, and $\psi_4(p)$ are derived by circular permutations in the sets $\bigcup_{i=1}^{4}\{a_i\}$, $\bigcup_{i=5}^{8}\{a_i\}$, $\bigcup_{i=9}^{12}\{a_i\}$ and $\bigcup_{i=13}^{16}\{a_i\}$. Then any polynomial in the space $Q_3'$ is uniquely determined by its values at the points $a_i$, $1 \leqslant i \leqslant 12$. In addition, the inclusion*

$$P_3 \subset Q_3' \tag{2.2.34}$$

*holds.*

**Proof.** The proof is left as a problem (Exercise 2.2.5). We shall only record the identity

$$\forall p \in Q_3', \quad p = \sum_{i=1}^{12} p(a_i)p_i,$$

with

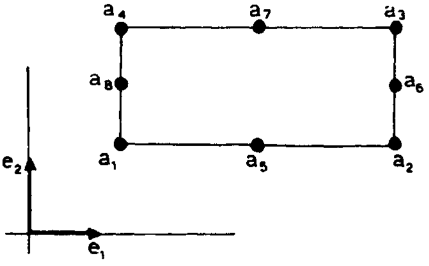
$$\left.\begin{aligned} p_1 &= x_3x_4\left\{1 + \frac{9}{2}x_3(x_3-1) + \frac{9}{2}x_4(x_4-1)\right\}, \ldots \\ p_5 &= -\frac{9}{2}x_3(x_3-1)(3x_3-1)x_4, \ldots \\ p_9 &= \frac{9}{2}x_3(x_3-1)(3x_3-2)x_4, \ldots \end{aligned}\right\} \tag{2.2.35}$$

□

From these two theorems, we derive the definition of the *rectangle of type* (2′) (Fig. 2.2.12) and of the *rectangle of type* (3′) (Fig. 2.2.13).

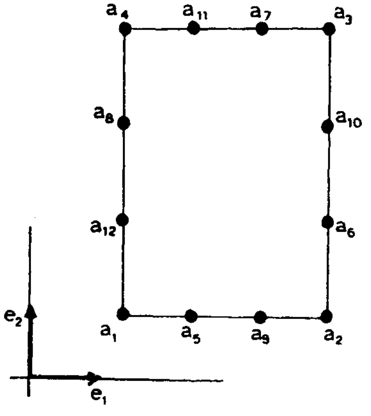
If it happens that the set $\bar{\Omega} \subset \mathbf{R}^n$ is *rectangular*, i.e., it is either an $n$-rectangle or a finite union of $n$-rectangles, it can be conveniently "triangulated" by finite elements which are themselves $n$-rectangles: The fifth condition ($\mathcal{T}_h5$) on a triangulation now reads:

($\mathcal{T}_h5$) *Any face of any $n$-rectangle $K_1$ in the triangulation is either a subset of the boundary $\Gamma$, or a face of another $n$-rectangle $K_2$ in the triangulation.*



| *rectangle of type* (2′) |
|---|
| $P_K = Q_2'(K)$ (cf. (2.2.29)); dim $P_K = 8$;<br>$\Sigma_K = \{p(a_i),\quad 1 \leqslant i \leqslant 8\}$. |

Fig. 2.2.12



| *rectangle of type* (3′) |
|---|
| $P_K = Q_3'(K)$ (cf. (2.2.32)); dim $P_K = 12$;<br>$\Sigma_K = \{p(a_i);\quad 1 \leqslant i \leqslant 12\}$. |

Fig. 2.2.13

In the second case, the $n$-rectangles $K_1$ and $K_2$ are said to be *adjacent.*

An example of a triangulation made up of rectangles is given in Fig. 2.2.14.

*With such a triangulation, we may associate in a natural way a finite element space $X_h$ with each type of the rectangular finite elements which we just described.* Since the discussion is almost identical to the one concerning $n$-simplices, we shall be very brief. In particular, one can prove the following analog of Theorem 2.2.3.

**Theorem 2.2.7.** *Let $X_h$ be the finite element space associated with $n$-rectangles of type $(k)$ for any integer $k \geq 1$ or with rectangles of type $(2')$ or $(3')$. Then the inclusion*

$$X_h \subset \mathscr{C}^0(\bar{\Omega}) \cap H^1(\Omega) \tag{2.2.36}$$

*holds.* □

Finally, arguing as before, it is easily seen that such finite element spaces possess a basis whose functions have "small" support (FEM 3).

*First examples of finite elements with derivatives as degrees of freedom: Hermite n-simplices of type* (3), (3'). *Assembly in triangulations*

So far, the degrees of freedom of each finite element $K$ have been "point values", i.e., of the form $p(a)$, for some points $a \in K$. We shall next introduce finite elements in which some degrees of freedom are partial derivatives, or, more generally, *directional derivatives,* i.e.,

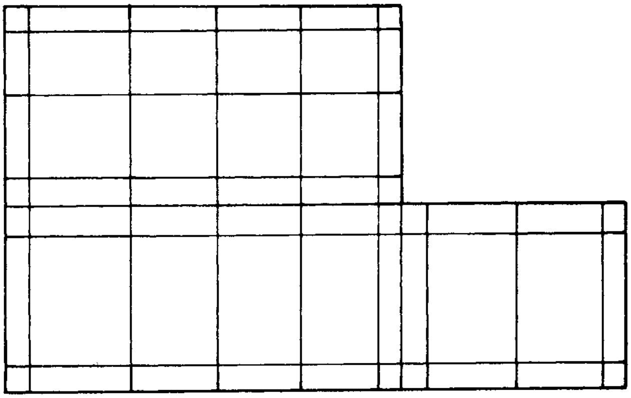


Fig. 2.2.14

expressions such as $Dp(a)b$, $D^2p(a)(b,c)$, etc..., where $b, c$ are vectors in $\mathbf{R}^n$.

The first example of this type of finite element is based on the following theorem.

**Theorem 2.2.8.** *Let $K$ be an $n$-simplex with vertices $a_i$, $1 \leqslant i \leqslant n+1$, and let $a_{ijk} = \frac{1}{3}(a_i + a_j + a_k)$, $1 \leqslant i < j < k \leqslant n+1$. Then any polynomial in the space $P_3$ is uniquely determined by its values and the values of its $n$ first partial derivatives at the vertices $a_i$, $1 \leqslant i \leqslant n+1$, and its values at the points $a_{ijk}$, $1 \leqslant i < j < k \leqslant n+1$.*

**Proof.** It suffices to argue as usual so as to obtain the following identity:

$$\begin{aligned}\forall p \in P_3, \quad p = {} & \sum_i \Big(-2\lambda_i^3 + 3\lambda_i^2 - 7\lambda_i \sum_{\substack{j<k \\ j\neq i,\, k\neq i}} \lambda_j\lambda_k\Big) p(a_i) \\ & + 27 \sum_{i<j<k} \lambda_i\lambda_j\lambda_k p(a_{ijk}) \\ & + \sum_{i\neq j} \lambda_i\lambda_j(2\lambda_i + \lambda_j - 1) Dp(a_i)(a_j - a_i). \end{aligned} \tag{2.2.37}$$

The only novelty is that one needs to use the derivatives of the barycentric coordinates in order to show that $Dp(a_i) = D\tilde{p}(a_i)$, $1 \leqslant i \leqslant n+1$, denoting momentarily by $\tilde{p}$ the right-hand side of (2.2.37). By differentiating the polynomial $\tilde{p}$, we obtain

$$D\tilde{p}(a_i) = \sum_{j\neq i} \{Dp(a_i)(a_j - a_i)\} D\lambda_j.$$

To show that the above expression is equal to $Dp(a_i)$, it is equivalent to show that

$$D\tilde{p}(a_i)(a_k - a_i) = Dp(a_i)(a_k - a_i), \quad 1 \leqslant k \leqslant n+1, \quad k \neq i.$$

These last relations are in turn consequences of the relations

$$D\lambda_j(a_k - a_i) = \delta_{jk} - \lambda_j(a_i), \quad 1 \leqslant k \leqslant n+1, \quad k \neq i,$$

which we now establish. Denoting by $B$ the inverse matrix of the matrix $A$ of (2.2.4), we obtain

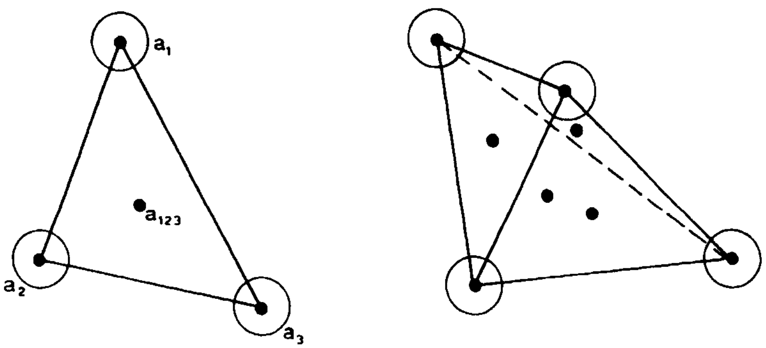
$$\partial_l\lambda_j = b_{jl}, \quad 1 \leqslant j \leqslant n+1, \quad 1 \leqslant l \leqslant n$$

(cf. (2.2.7)). Therefore we have

$$D\lambda_j(a_k - x) = \sum_{l=1}^{n} b_{jl}a_{lk} - \sum_{l=1}^{n} b_{jl}x_l = \delta_{jk} - \lambda_j(x)$$

for any $x \in \mathbf{R}^n$, and in particular for $x = a_i$. □

From this theorem, we deduce the definition of a finite element, which is called the *Hermite n-simplex of type* (3) (Fig. 2.2.15), where the directional derivatives $Dp(a_i)(a_j - a_i)$ are degrees of freedom. Of course, the knowledge of these $n$ directional derivatives at a vertex $a_i$ is equivalent to the knowledge of the first derivative $Dp(a_i)$. Such a knowledge is indicated graphically by one small circle, or sphere, centered at the point $a_i$. Since the first derivative $Dp(a_i)$ is equally well determined by the partial derivatives $\partial_j p(a_i)$, $1 \leq j \leq n$, another possible set of degrees of freedom for this element is the set $\Sigma'_K$ indicated in Fig. 2.2.15.



$\left\{\begin{array}{l}\textit{Hermite triangle of type } (3)\\ \dim P_K = 10\end{array}\right.$

$\left\{\begin{array}{l}\textit{Hermite tetrahedron of type } (3)\\ \dim P_K = 20\end{array}\right.$

| *Hermite n-simplex of type* (3) |
|---|
| $P_K = P_3(K)$; $\dim P_K = \dfrac{(n+1)(n+2)(n+3)}{6}$;<br>$\Sigma_K = \{p(a_i),\ 1 \leq i \leq n+1;\ p(a_{ijk}),\ 1 \leq i < j < k \leq n+1;$<br>$Dp(a_i)(a_j - a_i),\ 1 \leq i, j \leq n+1,\ j \neq i\}.$<br>$\Sigma'_K = \{p(a_i),\ 1 \leq i \leq n+1;\ p(a_{ijk}),\ 1 \leq i < j < k \leq n+1;$<br>$\partial_j p(a_i),\ 1 \leq i \leq n+1,\ 1 \leq j \leq n\}.$ |

Fig. 2.2.15

The derivation of a related element without the degrees of freedom $p(a_{ijk})$, $i<j<k$, is based on the following theorem, whose proof is left to the reader (Exercise 2.2.6).

**Theorem 2.2.9.** *For each triple* $(i, j, k)$ *with* $i<j<k$, *let*

$$\psi_{ijk}(p) = 6p(a_{ijk}) - 2\sum_{l=i,j,k} p(a_l) + \sum_{l=i,j,k} Dp(a_l)(a_l - a_{ijk}). \tag{2.2.38}$$

*Then any polynomial in the space*

$$P_3'' = \{p \in P_3;\quad \psi_{ijk}(p) = 0,\quad 1 \leq i < j < k \leq n+1\} \tag{2.2.39}$$

*is uniquely determined by its values and the values of its* $n$ *first partial derivatives at the vertices* $a_i$, $1 \leq i \leq n+1$. *In addition, the inclusion*

$$P_2 \subset P_3'' \tag{2.2.40}$$

*holds.* □

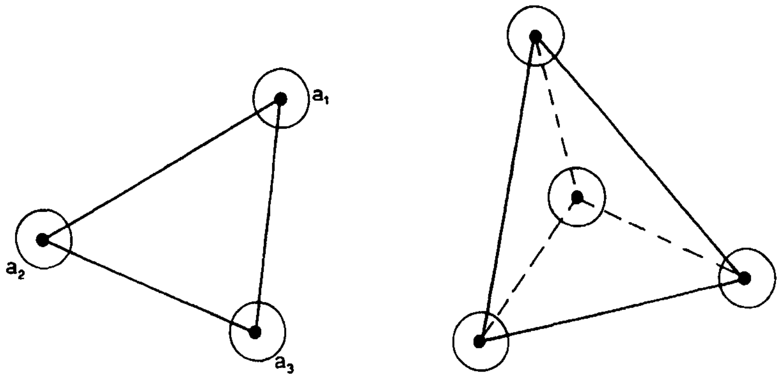
From this theorem, one deduces the definition of the *Hermite n-simplex of type* (3′), which, in case $n = 2$, is also called the *Zienkiewicz triangle* (Fig. 2.2.16).

Given a triangulation made up of $n$-simplices, we associate in a natural way a finite element space $X_h$ with either type of finite elements. To be specific, assume we are using Hermite $n$-simplices of type (3), the case of Hermite $n$-simplices of type (3′) being quite similar. Then a function $v_h$ is in the space $X_h$ if (i) each restriction $v_h|_K$ is in the space $P_K = P_3(K)$ for each $K \in \mathcal{T}_h$, and (ii) it is defined by its values at all the vertices of the triangulation, its values at the centers of gravity of all triangles found as 2-faces of the $n$-simplices $K \in \mathcal{T}_h$, and the values of its $n$ first partial derivatives at all the vertices of the triangulation. The corresponding set of degrees of freedom of the space $X_h$ is thus of the form

$$\Sigma_h = \{v_h(b);\quad b \in \mathcal{N}_v \cup \mathcal{N}_c;\quad \partial_j v_h(b),\quad b \in \mathcal{N}_v;\quad 1 \leq j \leq n\},$$

where $\mathcal{N}_v$ denotes the set of all the vertices of the $n$-simplices of the triangulation and $\mathcal{N}_c$ denotes the set of all centers of gravity of all 2-faces of the $n$-simplices found in the triangulation.

When a finite element space is constructed with $n$-simplices of type (3) or (3′), the sets $\Sigma_K'$ are preferred to the sets $\Sigma_K$ (cf. Figs. 2.2.15 and



*Zienkiewicz triangle, or*
*Hermite triangle of type* (3')
dim $P_K = 9$

*Hermite tetrahedron of type* (3')
dim $P_K = 16$

| *Hermite n-simplex of type* (3') |
|---|
| $P_K = P_3''(K)$ (cf. (2.2.39)), dim $P_K = (n+1)^2$;<br>$\Sigma_K = \{p(a_i),\ 1 \leqslant i \leqslant n+1;\ Dp(a_i)(a_j - a_i),\ 1 \leqslant i, j \leqslant n+1,\ i \neq j\}$.<br>$\Sigma'_K = \{p(a_i),\ 1 \leqslant i \leqslant n+1;\ \partial_j p(a_i),\ 1 \leqslant i \leqslant n+1,\ 1 \leqslant j \leqslant n\}$. |

Fig. 2.2.16

2.2.16) inasmuch as they directly correspond to the set $\Sigma_h$, but this observation is of a purely practical nature.

Again, *requirement* ($\mathcal{T}_h 5$) *insures that the degrees of freedom are unambiguously defined across adjacent finite elements*, and it is also the basis for the following theorem.

**Theorem 2.2.10.** *Let $X_h$ be the finite element space associated with Hermite n-simplices of type* (3), *or with Hermite n-simplices of type* (3'). *Then the inclusion*

$$X_h \subset \mathscr{C}^0(\bar{\Omega}) \cap H^1(\Omega) \tag{2.2.41}$$

*holds.*

**Proof.** Arguing as in Theorem 2.2.3, it suffices to derive the inclusion $X_h \subset \mathscr{C}^0(\bar{\Omega})$: Along any side common to two adjacent triangles, there is a unique polynomial of degree 3 in one variable which takes on prescribed values and prescribed first derivatives at the end points of the side.

This argument easily extends to the $n$-dimensional case. □

To verify requirement (FEM 3), let us assume for definiteness that we are considering Hermite triangles of type (3), so that the associated set of degrees of freedom of the space is of the form

$$\begin{aligned}\Sigma_h = \{v(b_k),\quad & \partial_1 v(b_k),\quad \partial_2 v(b_k),\quad 1 \leq k \leq J;\\ & v(b_k),\quad J+1 \leq k \leq L\}.\end{aligned}\tag{2.2.42}$$

Then if we define functions $w_k$, $w_k^1$, $w_k^2 \in X_h$ by the conditions

$$\left.\begin{aligned}
&w_k(b_l) = \delta_{kl},\quad 1 \leq k,l \leq L,\quad \partial_1 w_k(b_l) = \partial_2 w_k(b_l) = 0,\\
&\qquad 1 \leq k \leq L,\quad 1 \leq l \leq J,\\
&w_k^1(b_l) = 0,\quad 1 \leq k \leq J,\quad 1 \leq l \leq L,\quad \partial_1 w_k^1(b_l) = \delta_{kl},\\
&\qquad \partial_2 w_k^1(b_l) = 0,\quad 1 \leq k, l \leq J,\\
&w_k^2(b_l) = 0,\quad 1 \leq k \leq J,.\quad 1 \leq l \leq L,\quad \partial_1 w_k^2(b_l) = 0,\\
&\qquad \partial_2 w_k^2(b_l) = \delta_{kl},\quad 1 \leq k, l \leq J,
\end{aligned}\right\}\tag{2.2.43}$$

it is easily seen that these functions have "small" supports.

*First examples of finite elements for fourth-order problems: The Argyris and Bell triangles, the Bogner–Fox–Schmit rectangle. Assembly in triangulations*

Finally, we examine some examples of finite elements which yield the inclusion $X_h \subset \mathscr{C}^1(\bar{\Omega})$, and which may therefore be used for solving fourth-order problems. It is legitimate to restrict ourselves to the case where $n = 2$, in view of the examples given in Section 1.2. Our first example is based on the following result.

**Theorem 2.2.11.** *Let $K$ be a triangle with vertices $a_i$, $1 \leq i \leq 3$, and let $a_{ij} = \frac{1}{2}(a_i + a_j)$, $1 \leq i < j \leq 3$, denote the mid-points of the sides. Then any polynomial $p$ of degree 5 is uniquely determined by the following set of 21 degrees of freedom:*

$$\Sigma_K = \{\partial^\alpha p(a_i),\quad |\alpha| \leq 2,\ 1 \leq i \leq 3;\quad \partial_\nu p(a_{ij}),\quad 1 \leq i < j \leq 3\},\tag{2.2.44}$$

*where $\partial_\nu$ denotes the normal derivative operator along the boundary of $K$.*

**Proof.** Given a set of degrees of freedom, finding the corresponding polynomial of degree 5 amounts to solving a linear system with a square matrix, for which existence and uniqueness for all right-hand sides are equivalent properties, as we already observed. We shall prove the latter property, i.e., that any polynomial $p \in P_5$ such that

$$\partial^\alpha p(a_i) = 0, \quad |\alpha| \leqslant 2, \quad 1 \leqslant i \leqslant 3, \quad \partial_\nu p(a_{ij}) = 0, \quad 1 \leqslant i < j \leqslant 3,$$

is identically zero.

Let $t$ denote an abscissa along the axis which contains the side $K' = [a_1, a_2]$. Then the restriction $p|_{K'}$, considered as a function $q$ of $t$, is a polynomial of degree 5 which satisfies

$$q(a_1) = q'(a_1) = q''(a_1) = q(a_2) = q'(a_2) = q''(a_2) = 0,$$

since, if $\tau$ is a unit vector on the axis containing the side $K'$, we have

$$q'(a_1) = \partial_\tau p(a_1), \quad q''(a_1) = \partial_{\tau\tau} p(a_1), \text{ etc.}\ldots,$$

and thus $q = 0$.

Likewise, considered as a function $r$ of $t$, the normal derivative $\partial_\nu p$ along $K'$ is a polynomial of degree 4 which satisfies

$$r(a_1) = r'(a_1) = r(a_{12}) = r(a_2) = r'(a_2) = 0,$$

since

$$r(a_1) = \partial_\nu p(a_1), \quad r'(a_1) = \partial_{\nu\tau} p(a_1), \quad r(a_{12}) = \partial_\nu p(a_{12}), \text{ etc.}\ldots,$$

and, thus, $r = 0$.

Since we have $\partial_\tau p = 0$ along $K'$ ($p = 0$ along $K'$), we have proved that $p$ and its first derivative $Dp$ vanish identically along $K'$. This implies that the polynomial $\lambda_3^2$ is a factor of $p$, as we now show: After using an appropriate affine mapping if necessary, we may assume without loss of generality that $\lambda_3(x_1, x_2) = x_1$. We can write

$$p(x_1, x_2) = \sum_{i=0}^{5} x_1^i p_i(x_2)$$

where $p_i$, $0 \leqslant i \leqslant 5$, are polynomials of degree $(5 - i)$ in the variable $x_2$. Therefore

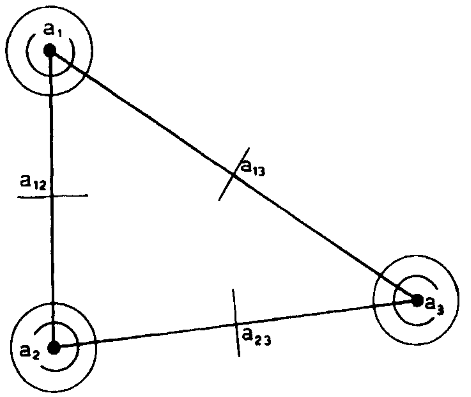
$$\forall x_2 \in \mathbf{R}, \quad p(0, x_2) = p_0(x_2) = 0,$$
$$\forall x_2 \in \mathbf{R}, \quad \partial_1 p(0, x_2) = p_1(x_2) = 0,$$

which proves our assertion.

Similar arguments hold for the other sides, and we find that the polynomial $(\lambda_1^2\lambda_2^2\lambda_3^2)$ is a factor of $p$. Since the $\lambda_i$ are polynomials of degree 1 which do not reduce to constants, it necessarily follows that $p = 0$. □

With Theorem 2.2.11, we can define a finite element, the 21-*degree of freedom triangle*, also known as *Argyris triangle* (Fig. 2.2.17).

Fig. 2.2.17 is self-explanatory as regards the graphical symbols used for representing the various degrees of freedom. We observe that at each vertex $a_i$, the first and second derivatives $Dp(a_i)$ and $D^2p(a_i)$ are known. With this observation in mind, we see that other possible definitions for the set of degrees of freedom are the sets $\Sigma_K'$ and $\Sigma_K''$



*Argyris triangle, or*
21-*degree of freedom triangle*

$P_K = P_5(K)$, $\dim P_K = 21$;

$\Sigma_K = \{p(a_i),\ \partial_1 p(a_i),\ \partial_2 p(a_i),\ \partial_{11} p(a_i),\ \partial_{12} p(a_i),\ \partial_{22} p(a_i),\ 1 \leqslant i \leqslant 3;$
$\quad \partial_\nu p(a_{ij}),\ 1 \leqslant i < j \leqslant 3\}$

$\Sigma_K' = \{p(a_i),\ 1 \leqslant i \leqslant 3;$
$\quad Dp(a_i)(a_j - a_i),\ 1 \leqslant i, j \leqslant 3,\ j \neq i;$
$\quad D^2p(a_i)(a_j - a_i, a_k - a_i),\ 1 \leqslant i, j, k \leqslant 3,\ j \neq i,\ k \neq i,$
$\quad \partial_\nu p(a_{ij}),\ 1 \leqslant i < j \leqslant 3\}$

$\Sigma_K'' = \{p(a_i),\quad Dp(a_i)(a_{i-1} - a_i),\quad Dp(a_i)(a_{i+1} - a_i),\quad 1 \leqslant i \leqslant 3;$
$\quad D^2p(a_i)(a_{j+1} - a_j)^2,\quad 1 \leqslant i, j \leqslant 3;$
$\quad Dp(a_{ij})\nu_k,\quad \{i, j, k\} = \{1, 2, 3\},\quad i < j\}.$

Fig. 2.2.17

indicated in Fig. 2.2.17. In the expression of the set $\Sigma''_K$, the indices are numbered modulo 3, and each vector $\nu_i$, $1 \le i \le 3$, is the height originating at the point $a_i$.

It may be desirable to dispose of the degrees of freedom $\partial_\nu p(a_{ij})$, $1 \le i < j \le 3$. This reduction will be a consequence of the following result.

**Theorem 2.2.12.** *Any polynomial in the space*

$$P_5'(K) = \{p \in P_5(K);\quad \partial_\nu p \in P_3(K') \quad \text{for each side } K' \text{ of } K\} \tag{2.2.45}$$

*is uniquely determined by the following set of* 18 *degrees of freedom*:

$$\Sigma_K = \{\partial^\alpha p(a_i),\ |\alpha| \le 2,\ 1 \le i \le 3\}. \tag{2.2.46}$$

*The space $P_5'(K)$ satisfies the inclusion*

$$P_4(K) \subset P_5'(K). \tag{2.2.47}$$

**Proof.** By writing $\partial_\nu p \in P_3(K')$ in definition (2.2.45), it is of course meant that, considered as a function of an abscissa along an axis containing the side $K'$, the normal derivative $\partial_\nu p$ is a polynomial of degree 3. The inclusion (2.2.47) being obvious, it remains to prove the first part of the theorem.

To begin with, we prove a preliminary result: *Let $K' = [a_i, a_j]$ be a segment in $\mathbf{R}^n$, with mid-point $a_{ij}$, and let $v$ be a function such that $v|_{K'} \in P_4(K')$. Then we have $v|_{K'} \in P_3(K')$ if and only if $\chi_{ij}(v) = 0$, where*

$$\begin{aligned}\chi_{ij}(v) = {} & 4(v(a_i) + v(a_j)) - 8v(a_{ij}) + Dv(a_i)(a_j - a_i) \\ & + Dv(a_j)(a_i - a_j).\end{aligned} \tag{2.2.48}$$

To see this, let, for any $x \in K'$, $\alpha_4 = D^4 v(x)\tau^4$, where $\tau$ is a unit vector along $K'$, so that $\alpha_4$ is a constant. Then we have

$$\begin{aligned} v(a_i) = {} & v(a_{ij}) + Dv(a_{ij})(a_i - a_{ij}) + \frac{1}{2} D^2 v(a_{ij})(a_i - a_{ij})^2 \\ & + \frac{1}{6} D^3 v(a_{ij})(a_i - a_{ij})^3 + \frac{\alpha_4}{24} \|a_i - a_{ij}\|^4, \\ v(a_j) = {} & v(a_{ij}) + Dv(a_{ij})(a_j - a_{ij}) + \frac{1}{2} D^2 v(a_{ij})(a_j - a_{ij})^2 \\ & + \frac{1}{6} D^3 v(a_{ij})(a_j - a_{ij})^3 + \frac{\alpha_4}{24} \|a_j - a_{ij}\|^4, \end{aligned}$$

from which we deduce $(a_i - a_{ij} = -(a_j - a_{ij}))$:

$$v(a_i) + v(a_j) = 2v(a_{ij}) + \frac{1}{2}\{D^2v(a_{ij})(a_i - a_{ij})^2 + D^2v(a_{ij})(a_j - a_{ij})^2\}$$
$$+ \frac{\alpha_4}{24}\{\|a_i - a_{ij}\|^4 + \|a_j - a_{ij}\|^4\}.$$

Likewise,

$$Dv(a_i)(a_i - a_{ij}) = D^2v(a_{ij})(a_i - a_{ij})^2 + \frac{1}{2}D^3v(a_{ij})(a_i - a_{ij})^3$$
$$+ \frac{\alpha_4}{6}\|a_i - a_{ij}\|^4,$$

$$Dv(a_j)(a_j - a_{ij}) = D^2v(a_{ij})(a_j - a_{ij})^2 + \frac{1}{2}D^3v(a_{ij})(a_j - a_{ij})^3$$
$$+ \frac{\alpha_4}{6}\|a_j - a_{ij}\|^4,$$

and therefore,

$$D^2v(a_{ij})(a_i - a_{ij})^2 + D^2v(a_{ij})(a_j - a_{ij})^2$$
$$= Dv(a_i)(a_i - a_{ij}) + Dv(a_j)(a_j - a_{ij})$$
$$- \frac{\alpha_4}{6}\{\|a_i - a_{ij}\|^4 + \|a_j - a_{ij}\|^4\}.$$

Combining our previous relations, we get

$$2v(a_{ij}) = v(a_i) + v(a_j) + \frac{1}{4}\{Dv(a_i)(a_j - a_i) + Dv(a_j)(a_i - a_j)\}$$
$$+ \frac{\alpha_4}{96}\|a_i - a_j\|^4,$$

and the assertion is proved.

As a consequence of this preliminary result, the space $P_5'(K)$ may be also defined as

$$P_5'(K) = \{p \in P_5(K);\quad \chi_{ij}(\partial_\nu p) = 0,\quad 1 \leq i < j \leq 3\}, \tag{2.2.49}$$

i.e., in view of relations (2.2.48), we have characterized the space $P_5'(K)$ by the property that each normal derivative $\partial_\nu p(a_{ij})$ is expressed as a linear combination of the parameters $\partial^\alpha p(a_i)$, $\partial^\alpha p(a_j)$, $|\alpha| = 1, 2$. Then the proof is completed by combining the usual argument with the result of Theorem 2.2.11. □

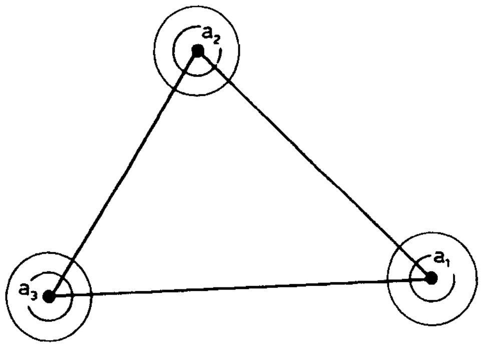
From Theorem 2.2.12, we deduce the definition of a finite element, called the 18-*degree of freedom triangle*, or, preferably, *Bell's triangle*. See Fig. 2.2.18, where we have indicated three possible sets of degrees of freedom which parallel those of the Argyris triangle.

Given a triangulation made up of triangles, we associate a finite element space $X_h$ with either type of finite elements. We leave it to the reader to derive the associated set of degrees of freedom of the space $X_h$ and to check that the canonical basis is again composed of functions with "small" support. We shall only prove the following result.

**Theorem 2.2.13.** *Let $X_h$ be the finite element space associated with Argyris triangles or Bell's triangles. Then the inclusion*

$$X_h \subset \mathscr{C}^1(\bar{\Omega}) \cap H^2(\Omega) \tag{2.2.50}$$

*holds.*



| *Bell's triangle or*<br>18-*degree of freedom triangle* |
|---|
| $P_K = P_5'(K)$ (cf. (2.2.45)); $\dim P_K = 18$;<br>$\Sigma_K = \{p(a_i),\ \partial_1 p(a_i),\ \partial_2 p(a_i),\ \partial_{11} p(a_i),\ \partial_{12} p(a_i),\ \partial_{22} p(a_i),\ 1 \leqslant i \leqslant 3\}$<br>$\Sigma_K' = \{p(a_i),\ 1 \leqslant i \leqslant 3;\quad Dp(a_i)(a_j - a_i),\ 1 \leqslant i, j \leqslant 3,\ j \neq i;$<br>$\quad D^2 p(a_i)(a_j - a_i, a_k - a_i),\ 1 \leqslant i, j, k \leqslant 3,\ j \neq i,\ k \neq i\}$<br>$\Sigma_K'' = \{p(a_i),\ Dp(a_i)(a_{i-1} - a_i),\ Dp(a_i)(a_{i+1} - a_i),\ 1 \leqslant i \leqslant 3;$<br>$\quad D^2 p(a_i)(a_{j+1} - a_j)^2,\ 1 \leqslant i, j \leqslant 3\}$ |

Fig. 2.2.18

**Proof.** By Theorem 2.1.2, it suffices to show that the inclusion $X_h \subset \mathscr{C}^1(\bar{\Omega})$ holds.

Let $K_1$ and $K_2$ be two adjacent triangles with a common side $K' = [b_i, b_j]$ (Fig. 2.2.19) and let $v_h$ be a function in the space $X_h$ constructed with Argyris triangles. Considered as functions of an abscissa $t$ along an axis containing the side $K'$, the functions $v_h|_{K_1}$ and $v_h|_{K_2}$ are, along $K'$, polynomials of degree 5 in the variable $t$. Call these polynomials $q_1$ and $q_2$. Since, by definition of the space $X_h$, we have

$$q(b_i) = q'(b_i) = q''(b_i) = q(b_j) = q'(b_j) = q''(b_j) = 0,$$

with $q = q_1 - q_2$, it follows that $q = 0$ and hence the inclusion $V_h \subset \mathscr{C}^0(\bar{\Omega})$ holds. Likewise, call $r_1$ and $r_2$, the restrictions to the side $K'$ of the functions $\partial_\nu v_h|_{K_1}$ and $-\partial_\nu v_h|_{K_2}$. Then both $r_1$ and $r_2$ are polynomials of degree 4 in the variable $t$ and, again by definition of the space $X_h$, we have

$$r(b_i) = r'(b_i) = r(b_{ij}) = r(b_j) = r'(b_j) = 0,$$

with $r = r_1 - r_2$, so that $r = 0$. We have thus proved the continuity of the normal derivative which, combined with the continuity of the tangential derivative ($q = 0$ along $K'$ implies $q' = 0$ along $K'$), shows that the first derivatives are also continuous on $\bar{\Omega}$.

If the space $X_h$ is constructed with Bell's triangles,'the argument is

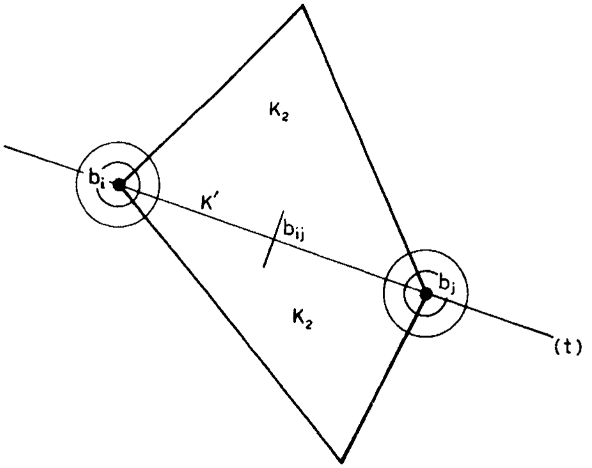


Fig. 2.2.19

identical for the difference $q = q_1 - q_2$. The difference $r = r_1 - r_2$ vanishes because it is a polynomial of degree 3 in the variable $t$ which is such that

$$r(b_i) = r'(b_i) = r(b_j) = r'(b_j) = 0. \qquad \square$$

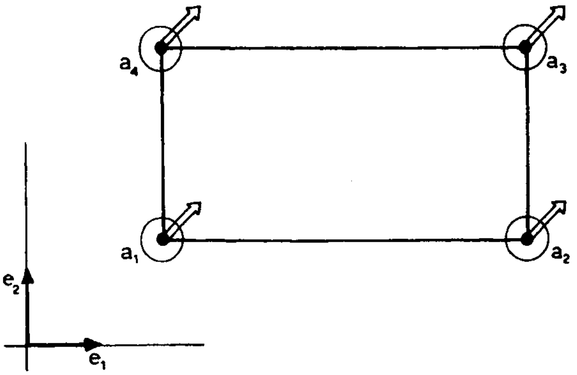
To conclude, we give one instance of a rectangular finite element which may be used for solving fourth-order problems posed over rectangular domains. Its existence depends upon the following theorem, whose proof is left to the reader (Exercise 2.2.8).

**Theorem 2.2.14.** *Let $K$ denote a rectangle with vertices $a_i$, $1 \leq i \leq 4$. Then a polynomial $p \in Q_3$ is uniquely determined by the following set of degrees of freedom*:

$$\Sigma_K = \{p(a_i),\quad \partial_1 p(a_i),\quad \partial_2 p(a_i),\quad \partial_{12} p(a_i),\quad 1 \leq i \leq 4\}. \qquad \square \tag{2.2.51}$$

The resulting finite element is the *Bogner–Fox–Schmit rectangle.* See Fig. 2.2.20, which is again self-explanatory for the graphical symbols.

The proof of the next result is also left to the reader (Exercise 2.2.8):



| *Bogner–Fox–Schmit rectangle* |
|---|
| $P_K = Q_3$; dim $P_K = 16$;<br>$\Sigma_K = \{p(a_i),\ \partial_1 p(a_i),\ \partial_2 p(a_i),\ \partial_{12} p(a_i),\ 1 \leq i \leq 4\}$ |

Fig. 2.2.20

**Theorem 2.2.15.** *Let $X_h$ be the finite element space associated with Bogner–Fox–Schmit rectangles. Then the inclusion*

$$X_h \subset \mathscr{C}^1(\bar{\Omega}) \cap H^2(\Omega) \tag{2.2.52}$$

*holds.* □

Finally, the reader should check, using the standard construction, that a finite element space constructed with any one of the last three finite elements indeed possesses canonical bases whose functions have "small" supports (FEM 3).

*Exercises*

***2.2.1.*** (i) Prove that the dimension of the space $P_k$, resp. $Q_k$, is $\binom{n+k}{k}$, resp. $(k+1)^n$.

(ii) Prove that $\dim P_k(A) = \dim P_k$, resp. $\dim Q_k(A) = \dim Q_k$, if the interior of the set $A \subset \mathbf{R}^n$ is not empty.

***2.2.2.*** Let $K$ be an $n$-simplex with vertices $a_j$, $1 \leqslant j \leqslant n+1$. For a given integer $k \geqslant 1$, show that a polynomial of degree $\leqslant k$ is uniquely defined by its values on the set $L_k(K)$ defined in (2.2.11) (Nicolaides (1972)). The set $L_k(K)$ is called the *principal lattice of order $k$* of the $n$-simplex $K$.

***2.2.3.*** Complete the proof of Theorem 2.2.3 so as to cover all cases.

***2.2.4.*** Give another proof of Theorem 2.2.4 (i.e., without recurring to identity (2.2.22)), by showing that if a polynomial of $Q_k$ vanishes on the set $\hat{M}_k$ defined in (2.2.21), then it is identically zero.

***2.2.5.*** Prove Theorem 2.2.6.

***2.2.6.*** Prove Theorem 2.2.9. Are the spaces $P'_3$ and $P''_3$ (cf. (2.2.13) and (2.2.39), respectively) identical?

***2.2.7.*** Given a triangle with vertices $a_i$, $1 \leqslant i \leqslant 3$, and mid-points $a_{ij} = \frac{1}{2}(a_i + a_j)$, $1 \leqslant i < j \leqslant 3$, show that a polynomial $p \in P_4$ is completely determined by the following degrees of freedom (Ženíšek (1974)):

$$\left.\begin{array}{ll} p(a_i),\quad \partial_{11}p(a_i),\quad \partial_{12}p(a_i),\quad \partial_{22}p(a_i), & 1 \leqslant i \leqslant 3, \\ p(a_{ij}),\quad 1 \leqslant i < j \leqslant 3. & \end{array}\right\}$$

Does this element yield the inclusion $X_h \subset \mathscr{C}^0(\bar{\Omega})$, resp. the inclusion $X_h \subset \mathscr{C}^1(\bar{\Omega})$?

***2.2.8.*** (i) Given a rectangle with vertices $a_i$, $1 \leqslant i \leqslant 4$, show that a polynomial $p \in Q_3$ is completely determined by the following degrees of freedom:

$$p(a_i),\quad \partial_1 p(a_i),\quad \partial_2 p(a_i),\quad \partial_{12} p(a_i),\quad 1 \leq i \leq 4.$$

(ii) Show that the corresponding space $X_h$ satisfies the inclusion

$$X_h \subset \mathscr{C}^1(\bar{\Omega}) \cap H^2(\Omega).$$

**2.2.9.** Consider the finite element space $X_h$ constructed with (Hermite) triangles of type (3) and let $w_k^1$, $1 \leq k \leq M_1$, be the basis functions of the space $X_h$ associated with the values at the barycenters of all the triangles of the triangulation, so that the discrete solution takes the form

$$u_h = \sum_{k=1}^{M} u_k w_k = \sum_{k=1}^{M_1} u_k^1 w_k^1 + \sum_{k=1}^{M_2} u_k^2 w_k^2.$$

Show that in this case the solution of the linear system (2.1.4) amounts, in fact, to solving a smaller linear system, in the unknowns $u_k^2$, $1 \leq i \leq M_2$, only.

This process, known as the *static condensation of the degrees of freedom*, is of course to be distinguished from the use of (Hermite) triangles of type (3').

## 2.3. General properties of finite elements and finite element spaces

### *Finite elements as triples $(K, P, \Sigma)$. Basic definitions. The $P$-interpolation operator*

Let us begin by giving the general definition of a finite element. A *finite element* in $\mathbf{R}^n$ is a triple $(K, P, \Sigma)$ where:

(i) $K$ is a closed subset of $\mathbf{R}^n$ with a non empty interior and a Lipschitz-continuous boundary,

(ii) $P$ is a space of real-valued functions defined over the set $K$,

(iii) $\Sigma$ is a finite set of linearly independent linear forms $\phi_i$, $1 \leq i \leq N$, defined over the space $P$ (in order to avoid ambiguities, the forms $\phi_i$ need to be defined over a larger space; we shall examine this point later; cf. Remark 2.3.3). By definition, it is assumed that the set $\Sigma$ is *$P$-unisolvent* in the following sense: given any real scalars $\alpha_i$, $1 \leq i \leq N$, there exists a unique function $p \in P$ which satisfies

$$\phi_i(p) = \alpha_i,\quad 1 \leq i \leq N. \tag{2.3.1}$$

Consequently, there exist functions $p_i \in P$, $1 \leq i \leq N$, which satisfy

$$\phi_j(p_i) = \delta_{ij}, \quad 1 \leq j \leq N. \tag{2.3.2}$$

Since we have

$$\forall p \in P, \quad p = \sum_{i=1}^{N} \phi_i(p)p_i, \tag{2.3.3}$$

Of course this implies that the space $P$ is finite-dimensional and that $\dim P = N$.

The linear forms $\phi_i$, $1 \leq i \leq N$, are called the *degrees of freedom of the finite element*, and the functions $p_i$, $1 \leq i \leq N$, are called the *basis functions of the finite element.*

Whenever we find it convenient, we shall use the notations $P_K$, $\Sigma_K$, $\phi_{i,K}$ and $p_{iK}$ in lieu of $P$, $\Sigma$, $\phi_i$ and $p_i$.

**Remark 2.3.1.** The set $K$ itself is often called a *finite element*, as we did in the previous section, and as we shall occasionally do in the sequel. □

**Remark 2.3.2.** The $P$-unisolvence of the set $\Sigma$ is equivalent to the fact that the $N$ linear forms $\phi_i$ form a basis in the dual space of $P$. As a consequence, one may view the bases $(\phi_i)_{i=1}^{N}$ and $(p_i)_{i=1}^{N}$ as being *dual bases*, in the algebraic sense.

In the light of the definition of a finite element, let us briefly review the examples given in the previous section.

We have seen examples for which the set $K$ is either an $n$-simplex, in which case the finite element is said to be *simplicial*, or *triangular* if $n = 2$, or *tetrahedral* if $n = 3$, or an $n$-rectangle in $\mathbf{R}^n$, in which case the finite element is said to be *rectangular*. As we already mentioned, these are all special cases of *straight finite elements*, i.e., for which the set $K$ is a polyhedron in $\mathbf{R}^n$. Other polygonal shapes are found in practice, such as *quadrilaterals* (see Section 4.3 and Section 6.1) or "*prismatic*" *finite elements* (see Remark 2.3.6). We shall also describe (Section 4.3) "*curved*" *finite elements*, i.e., whose boundaries are composed of "curved" faces.

*The main characteristic of the various spaces $P$ encountered in the examples is that they all contain a "full" polynomial space $P_k(K)$ for* *some integer* $k \geq 1$, a property that will be shown in subsequent chapters to be crucial as far as convergence properties are concerned.

In all the examples described previously, the degrees of freedom were of some of the following forms:

$$\left.\begin{aligned} &p \to p(a_i^0), \\ &p \to Dp(a_i^1)\xi_{ik}^1, \\ &p \to D^2p(a_i^2)(\xi_{ik}^2, \xi_{il}^2), \end{aligned}\right\} \tag{2.3.4}$$

where the points $a_i^r$, $r = 0, 1, 2$, belong to the finite element, and the (non zero) vectors $\xi_{ik}^1$, $\xi_{ik}^2$, $\xi_{il}^2$ are either constructed from the geometry of the finite element (e.g., $Dp(a_i)(a_j - a_i)$, $\partial_\nu p(a_{ij})$, etc...) or fixed vectors of $\mathbf{R}^n$ (e.g., $\partial_i p(a_j)$, $\partial_{ij} p(a_k)$). The points $a_i^r$, $r = 0, 1, 2$, are called the *nodes of the finite element* and make up a set which shall be denoted $\mathcal{N}_K$ in general.

Whereas only directional derivatives of order 1 or 2 occurred in the examples, one could conceivably consider degrees of freedom which would be partial derivatives of arbitrarily high order, but these are seldom used in practice. As we shall see later, however, (Section 4.2 and Section 6.2) there are practical instances of degrees of freedom which are not attached to nodes: They are instead *averages* (over the finite element or over one of its faces) of some partial derivative.

When all the degrees of freedom of a finite element are of the form $p \to p(a_i)$, we shall say that the associated finite element is a *Lagrange finite element* while if at least one directional derivative occurs as a degree of freedom, the associated finite element is said to be a *Hermite finite element.*

As the examples in the previous section have shown, there are essentially two methods for proving that a given set $\Sigma$ of degrees of freedom is $P$-unisolvent: *After it has been checked that* $\dim P = \operatorname{card}(\Sigma)$, one either

(i) exhibits the basis functions, or

(ii) shows that if all the degrees of freedom are set equal to zero, then the only corresponding function in the space $P$ is identically zero.

We have used method (i) for all the examples, except for the Argyris triangle where we used method (ii).

Given a finite element $(K, P, \Sigma)$, and given a function $v = K \to \mathbf{R}$, sufficiently smooth so that the degrees of freedom $\phi_i(v)$, $1 \leq i \leq N$, are well defined, we let

$$\Pi v = \sum_{i=1}^{N} \phi_i(v)p_i \tag{2.3.5}$$

denote the *P-interpolant* of the function $v$, which is unambiguously defined since the set $\Sigma$ is $P$-unisolvent. Indeed, the $P$-interpolant, also denoted $\Pi_K v$, is equivalently characterized by the conditions

$$\Pi v \in P, \quad \text{and} \quad \phi_i(\Pi v) = \phi_i(v), \quad 1 \leq i \leq N. \tag{2.3.6}$$

Whenever the degrees of freedom are of the form (2.3.4), let $s$ denote the maximal order of derivatives occurring in the definition of the set $\Sigma$. Then, for all finite elements of this type described here, the inclusion $P \subset \mathscr{C}^s(K)$ holds. Consequently, we shall usually consider that *the domain* dom $\Pi$ *of the P-interpolation operator* $\Pi$ *is the space*

$$\operatorname{dom} \Pi = \mathscr{C}^s(K). \tag{2.3.7}$$

This being the case, it follows that *over the space* $P \subset \operatorname{dom} \Pi$, *the interpolation operator reduces to the identity*, i.e.,

$$\forall p \in P, \quad \Pi p = p. \tag{2.3.8}$$

**Remark 2.3.3.** In order that the $P$-interpolation operator be unambiguously defined, *it is necessary that the forms* $\phi_i$ *be also defined on the space* $\mathscr{C}^s(K)$, for the following reason. Assume again that the space $P$ is contained in the space $\mathscr{C}^s(K)$. Then if the domain of the operator $\Pi$ were only the space $P$, infinitely many extensions to the space $\mathscr{C}^s(K)$ would exist. Let us give one simple example of such a phenomenon: Let $K$ be an $n$-simplex with barycenter $a$. Then the linear form $p \in \mathscr{C}^0(K) \to 1/(\operatorname{meas}(K)) \int_K p \, dx$ is one possible extension of the form $p \in P_1(K) \to p(a)$.

Of course, these considerations are usually omitted inasmuch as when one considers a degree of freedom such as $\partial_i p(a_j)$ for instance, it is implicitly understood that this form is the usual one, i.e., defined over the space $\mathscr{C}^1(K)$, not any one of its possible extensions from the space $P$ to the space $\mathscr{C}^1(K)$. For another illustration of this circumstance, see the description of *Wilson's brick*, in Section 4.2. □

Whereas for a Lagrange finite element, the set of degrees of freedom is unambiguously defined – indeed, it can be conveniently identified with the set of nodes - there are always several possible definitions for the degrees of freedom of a Hermite finite element which correspond to the "same" finite element. More precisely, *we shall say that two finite elements* $(K, P, \Sigma)$ *and* $(L, Q, \Xi)$ *are equal if we have*

$$K = L, \quad P = Q \quad \text{and} \quad \Pi_K = \Pi_L. \tag{2.3.9}$$

As an example, let us consider the Hermite $n$-simplex of type (3') with the two sets of degrees of freedom (cf. Fig. 2.2.16):

$$\begin{aligned} \Sigma = \{p(a_i), \quad & 1 \leqslant i \leqslant n+1; \quad Dp(a_i)(a_j - a_i), \\ & 1 \leqslant i,\ j \leqslant n+1, \quad i \neq j\}, \\ \Sigma' = \{p(a_i), \quad & 1 \leqslant i \leqslant n+1; \quad \partial_k p(a_i), \quad 1 \leqslant i \leqslant n+1, \\ & 1 \leqslant k \leqslant n\}. \end{aligned}$$

Let us denote by $\Pi$ and $\Pi'$ the corresponding $P_3(K)$-interpolation operators. Then, for any function $v \in \mathscr{C}^1(K) = \operatorname{dom} \Pi = \operatorname{dom} \Pi'$, we have, with obvious notations,

$$\begin{aligned} \Pi v &= \sum_i v(a_i) p_i + \sum_{i,j} Dv(a_i)(a_j - a_i) p_{ij}, \\ \Pi' v &= \sum_i v(a_i) p'_i + \sum_{i,k} \partial_k v(a_i) p'_{ik}. \end{aligned}$$

One has, for each pair $(i, j)$, $Dv(a_i)(a_j - a_i) = \sum_{k=1}^n \mu_{ijk} \partial_k v(a_i)$ for appropriate coefficients $\mu_{ijk}$. To conclude that $\Pi = \Pi'$, it suffices to observe that for each polynomial $p \in P_K$, one also has $Dp(a_i)(a_j - a_i) = \sum_{k=1}^n \mu_{ijk} \partial_k p(a_i)$ with the *same* coefficients $\mu_{ijk}$.

### *Affine families of finite elements*

We now come to an essential idea, which we shall first illustrate by an example.

Suppose we are given a family $(K, P_K, \Sigma_K)$ of triangles of type (2). Then *our aim is to describe such a family as simply as possible.*

Let $\hat{K}$ be a triangle with vertices $\hat{a}_i$, and mid-points of the sides $\hat{a}_{ij} = (\hat{a}_i + \hat{a}_j)/2$, $1 \leqslant i < j \leqslant 3$, and let

$$\hat{\Sigma} = \{p(\hat{a}_i), \quad 1 \leqslant i \leqslant 3; \qquad p(\hat{a}_{ij}), \quad 1 \leqslant i < j \leqslant 3\},$$

so that the triple $(\hat{K}, \hat{P}, \hat{\Sigma})$ with $\hat{P} = P_2(\hat{K})$ is also a triangle of type (2).

Given any finite element $K$ in the family (Fig. 2.3.1), *there exists a unique invertible affine mapping*

$$F_K: \hat{x} \in \mathbf{R}^2 \to F_K(\hat{x}) = B_K\hat{x} + b_K,$$

i.e., with $B_K$ an invertible $2 \times 2$ matrix and $b_K$ a vector of $\mathbf{R}^2$, such that

$$F_K(\hat{a}_i) = a_i, \quad 1 \leq i \leq 3.$$

Then *it automatically follows that*

$$F_K(\hat{a}_{ij}) = a_{ij}, \quad 1 \leq i < j \leq 3,$$

since the property for a point to be the mid-point of a segment is preserved by an affine mapping (likewise, the points which we called $a_{iij}$ or $a_{ijk}$ keep their geometrical definition through an affine mapping).

Once we have established a bijection $\hat{x} \in \hat{K} \to x = F_K(\hat{x}) \in K$ between the points of the sets $\hat{K}$ and $K$, it is natural to associate the space

$$P_K^* = \{p: K \to \mathbf{R}; \quad p = \hat{p} \cdot F_K^{-1}, \quad \hat{p} \in \hat{P}\}$$

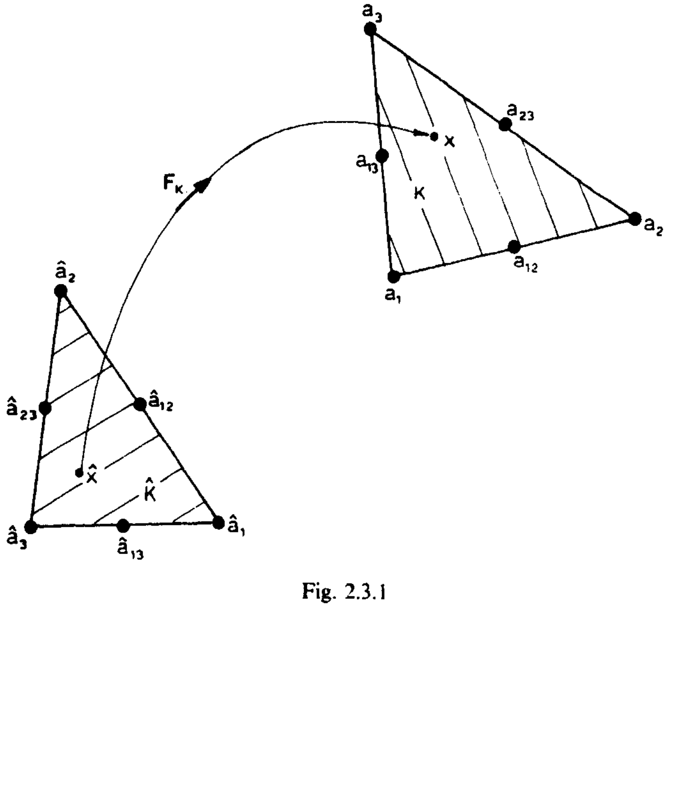


Fig. 2.3.1

with the space $P$. *Then it follows that*

$$P_K^* = P_2(K) = P_K,$$

because the mapping $F_K$ is affine.

In other words, *rather than prescribing such a family by the data* $K$, $P_K$ *and* $\Sigma_K$, *it suffices to give one reference finite element* $(\hat{K}, \hat{\Sigma}, \hat{P})$ *and the affine mappings* $F_K$. Then the generic finite element $(K, P_K, \Sigma_K)$ in the family is such that

$$\left.\begin{aligned} &K = F_K(\hat{K}),\\ &P_K = \{p\colon K \to \mathbf{R};\quad p = \hat{p}\cdot F^{-1},\quad \hat{p} \in \hat{P}\},\\ &\Sigma_K = \{p\{F_K(\hat{a}_i)),\quad 1 \leq i \leq 3;\quad p(F_K(\hat{a}_{ij})),\quad 1 \leq i < j \leq 3\}.\end{aligned}\right\}$$

With this example in mind, we are in a position to give the general definition: Two finite elements $(\hat{K}, \hat{P}, \hat{\Sigma})$ and $(K, P, \Sigma)$, with degrees of freedom of the form (2.3.4), are said to be *affine-equivalent* if there exists an *invertible affine mapping*:

$$F\colon \hat{x} \in \mathbf{R}^n \to F(\hat{x}) = B\hat{x} + b \in \mathbf{R}^n, \tag{2.3.10}$$

such that the following relations hold:

$$K = F(\hat{K}), \tag{2.3.11}$$

$$P = \{p\colon K \to \mathbf{R};\quad p = \hat{p}\cdot F^{-1},\quad \hat{p} \in \hat{P}\}, \tag{2.3.12}$$

$$a_i^r = F(\hat{a}_i^r),\quad r = 0, 1, 2, \tag{2.3.13}$$

$$\xi_{ik}^1 = B\hat{\xi}_{ik}^1,\quad \xi_{ik}^2 = B\hat{\xi}_{ik}^2,\quad \xi_{il}^2 = B\hat{\xi}_{il}^2, \tag{2.3.14}$$

whenever the nodes $a_i^r$, resp. $\hat{a}_i^r$, and vectors $\xi_{ik}^1$, $\xi_{ik}^2$, $\xi_{il}^2$, resp. $\hat{\xi}_{ik}^1$, $\hat{\xi}_{ik}^2$, $\hat{\xi}_{il}^2$, occur in the definition of the set $\Sigma$, resp. $\hat{\Sigma}$.

**Remark 2.3.4.** The justification of the relations (2.3.14) will become apparent in the proof of Theorem 2.3.1. □

With this definition of affine-equivalence, let us return to the examples given in Section 2.2 (the reader should check for oneself the various statements to come).

To begin with, it is clear that two $n$-simplices of type $k$ for a given integer $k \geq 1$, are affine equivalent, and that this is also the case for $n$-simplices of type (3′), in view of the definition (2.2.13) of the associated space $P_K$. Likewise, two $n$-rectangles of type $(k)$ for a given integer $k \geq 1$, or two rectangles of type (2′) or (3′) are affine equivalent through diagonal affine mappings. In other words, *any two identical Lagrange finite elements that we considered are affine-equivalent.*

When we come to Hermite finite elements, the situation is less simple. Consider for example two Hermite $n$-simplices of type (3) with sets of degrees of freedom in the form $\Sigma_K$ (Fig. 2.2.15). Then it is clear that they are affine-equivalent because the relations

$$a_j - a_i = F(\hat{a}_j) - F(\hat{a}_i) = B(\hat{a}_j - \hat{a}_i), \quad 1 \leq i, j \leq n+1, \quad j \neq i,$$

hold, among other things. However, had we taken the sets of degrees of freedom in the form $\Sigma'_K$, it would not have been clear to decide whether the two finite elements were affine-equivalent, and yet these two sets of degrees of freedom correspond to the *same* finite element, as we already pointed out.

The same analysis and conclusion apply to the Hermite $n$-simplex of type (3′) or to the Bogner–Fox–Schmit rectangle. In this last case, it suffices to observe that this finite element can also be defined by the following set of degrees of freedom (the index $i$ is counted modulo 4)

$$\begin{aligned} \Sigma'_K = \{p(a_i),\quad & Dp(a_i)(a_{i-1} - a_i),\quad Dp(a_i)(a_{i+1} - a_i), \\ & D^2p(a_i)(a_{i-1} - a_i,\quad a_{i+1} - a_i),\quad 1 \leq i \leq 4\}, \end{aligned} \tag{2.3.15}$$

for which relations (2.3.14) hold.

*There are counter-examples.* For instance, consider a finite element where some degrees of freedom are normal derivatives at some nodes. Then two such finite elements are not in general affine equivalent: The property for a vector to be normal to a hyperplane is not in general preserved through an affine mapping. Thus two Argyris triangles are not affine equivalent in general, except for instance if they happen to be both equilateral triangles. The case of Bell's triangles is left as a problem (Exercise 2.3.4).

Let us return to the general case. We shall constantly use the correspondences

$$\hat{x} \in \hat{K} \to x = F(\hat{x}) \in K, \tag{2.3.16}$$

$$\hat{p} \in \hat{P} \to p = \hat{p} \cdot F^{-1} \in P, \tag{2.3.17}$$

between the points $\hat{x} \in \hat{K}$ and $x \in K$, and the functions $\hat{p} \in \hat{P}$ and $p \in P$ corresponding to two affine-equivalent finite elements. As a consequence of the correspondences (2.3.16) and (2.3.17), notice that we have

$$\hat{p}(\hat{x}) = p(x) \quad \text{for all} \quad \hat{x} \in \hat{K}, \quad \hat{p} \in \hat{P}. \tag{2.3.18}$$

*We next prove a crucial relationship between the $\hat{P}$-interpolation operator $\hat{\Pi}$ and the $P$-interpolation operator $\Pi$* associated with affine-equivalent finite elements. This relationship will be itself a consequence of the fact that the basis functions are also in the correspondence (2.3.17).

**Theorem 2.3.1.** *Let $(\hat{K}, \hat{P}, \hat{\Sigma})$ and $(K, P, \Sigma)$ be two affine-equivalent finite elements with degrees of freedom in the form (2.3.4). Then if $\hat{p}_i$, $1 \leq i \leq N$, are the basis functions of the finite element $\hat{K}$, the functions $p_i$, $1 \leq i \leq N$, are the basis functions of the finite element $K$. The interpolation operators $\Pi$ and $\hat{\Pi}$ are such that*

$$(\Pi\, v)\hat{} = \hat{\Pi}\, \hat{v} \tag{2.3.19}$$

*for any functions $\hat{v} \in \operatorname{dom} \hat{\Pi}$ and $v \to \operatorname{dom} \Pi$ associated in the correspondence*

$$\hat{v} \in \operatorname{dom} \hat{\Pi} \to v = \hat{v} \cdot F^{-1} \in \operatorname{dom} \pi. \tag{2.3.20}$$

**Proof.** The $P$-interpolation operator $\Pi$ is of the form (with obvious notations):

$$\Pi v = \sum_i v(a_i^0) p_i^0 + \sum_{i,k} \{Dv(a_i^1) \xi_{ik}^1\} p_{ik}^1 + \sum_{i,k,l} \{D^2 v(a_i^2)(\xi_{ik}^2, \xi_{il}^2)\} p_{ikl}^2.$$

Using the derivation of composition of functions, we obtain

$$\begin{aligned} Dv(a_i^1)\xi_{ik}^1 &= Dv(F(\hat{a}_i^1)) B \hat{\xi}_{ik}^1 = Dv(F(\hat{a}_i^1)) DF(\hat{a}_i^1) \hat{\xi}_{ik}^1 \\ &= D(v \cdot F)(\hat{a}_i^1) \hat{\xi}_{ik}^1 = D\hat{v}(\hat{a}_i^1) \hat{\xi}_{ik}^1, \end{aligned}$$

and, taking also into account that $D^2 F = 0$,

$$\begin{aligned} D^2 v(a_i^2)(\xi_{ik}^2, \xi_{il}^2) &= D^2 v(F(\hat{a}_i^2))(B\hat{\xi}_{ik}^2, B\hat{\xi}_{il}^2) \\ &= D^2 v(F(\hat{a}_i^2))(DF(\hat{a}_i^2)\hat{\xi}_{ik}^2, DF(\hat{a}_i^2)\hat{\xi}_{il}^2) \\ &= D^2(v \cdot F)(\hat{a}_i^2)(\hat{\xi}_{ik}^2, \hat{\xi}_{il}^2) = D^2 \hat{v}(\hat{a}_i^2)(\hat{\xi}_{ik}^2, \hat{\xi}_{il}^2). \end{aligned}$$

Thus we also have

$$\Pi v = \sum_i \hat{v}(\hat{a}_i^0) p_i^0 + \sum_{i,k} \{D\hat{v}(\hat{a}_i^1) \hat{\xi}_{ik}^1\} p_{ik}^1 + \sum_{i,k,l} \{D^2 \hat{v}(\hat{a}_i^2)(\hat{\xi}_{ik}^2, \hat{\xi}_{il}^2)\} p_{ikl}^2,$$

from which we deduce, using the correspondence (2.3.17),

$$(\Pi v)^{\wedge} = \sum_i \hat{v}(\hat{a}_i^0)\hat{p}_i^0 + \sum_{i,k} \{D\hat{v}(\hat{a}_i^1)\hat{\xi}_{ik}^1\}\hat{p}_{ik}^1 + \sum_{i,k,l} \{D^2\hat{v}(\hat{a}_i^2)(\hat{\xi}_{ik}^2, \hat{\xi}_{il}^2)\}\hat{p}_{ikl}^2.$$

If we apply the previous identity to a function $v \in P$, we infer that the functions $\hat{p}_i^0$, $\hat{p}_{ik}^1$, $\hat{p}_{ik}^2$ are the basis functions of the finite element $(\hat{K}, \hat{P}, \hat{\Sigma})$, by virtue of identity (2.3.8). Using this result, we conclude that the function $(\Pi v)^{\wedge}$ is equal to the function $\hat{\Pi}\hat{v}$, by definition of the $\hat{P}$-interpolation operator $\hat{\Pi}$. □

**Remark 2.3.5.** To obtain the conclusion of the previous theorem when the sets of degrees of freedom are in the general form $\hat{\Sigma} = \{\hat{\phi}_i;\ 1 \le i \le N\}$ and $\Sigma = \{\phi_i;\ 1 \le i \le N\}$, it is necessary and sufficient that the degrees of freedom be such that

$$\forall \hat{v} \in \operatorname{dom} \hat{\Pi}, \quad \hat{\phi}_i(\hat{v}) = \phi_i(v), \quad 1 \le i \le N, \tag{2.3.21}$$

and, in essence, the proof of Theorem 2.3.1 consisted in showing that the above relations do hold (as consequences of relations (2.3.13)–(2.3.14)) for the type of degrees of freedom heretofore encountered. In Section 4.2, we shall consider a new type of degrees of freedom for which the validity of relations (2.3.21) will be verified. □

A family of finite elements is called an *affine family* if all its finite elements are affine-equivalent to a single finite element, which is called the *reference finite element* of the family (the reference finite element $(\hat{K}, \hat{P}, \hat{\Sigma})$ need not belong itself to the family).

In the case of an affine family of simplicial finite elements, a customary choice for the set $\hat{K}$ is the *unit n-simplex* with vertices

$$\left.\begin{aligned} \hat{a}_1 &= (1, 0, \dots, 0), \\ \hat{a}_2 &= (0, 1, 0, \dots, 0), \\ &\ \vdots \\ \hat{a}_n &= (0, \dots, 0, 1), \\ \hat{a}_{n+1} &= (0, 0, \dots, 0), \end{aligned}\right\}$$

for which the barycentric coordinates take the simple form

$$\lambda_i = x_i, \quad 1 \leqslant i \leqslant n, \quad \text{and} \quad \lambda_{n+1} = 1 - \sum_{i=1}^{n} x_i.$$

In the case of an affine family of rectangular finite elements, a usual choice for the set $\hat{K}$ is either the unit hypercube $[0, 1]^n$ or the hypercube $[-1, +1]^n$.

*The concept of an affine family of finite elements is of importance,* basically for the following reasons:

(i) *In practical computations, most of the work involved in the computation of the coefficients of the linear system* (2.1.4) *is performed on a reference finite element, not on a generic finite element.* This point will be illustrated in Section 4.1.

(ii) For such affine families, a fairly elegant *interpolation theory* can be developed (Section 3.1), which is, in turn, the basis of most *convergence theorems.*

(iii) Even when a family of finite elements of a given type is not an affine family, it is generally associated in an obvious way with an affine family whose "intermediate" role is essential. For example, when we shall study in Section 6.1 the interpolation properties of the Argyris triangle, an important step will consist in introducing a slightly different finite element (called the "Hermite triangle of type (5)"; cf. Exercise 2.3.5) which *can* be imbedded in an affine family. In the same fashion, we shall consider (Section 4.3) the "isoparametric" families of curved finite elements essentially as perturbations of affine families.

### *Construction of finite element spaces $X_h$. Basic definitions. The $X_h$-interpolation operator*

Our next task is to give a precise *description of the construction of a finite element space from the data of finite elements* $(K, P_K, \Sigma_K)$. For the sake of simplicity however, *we shall restrict ourselves to the case where the finite elements $K$ are all polygonal,* so that the set $\bar{\Omega} = \bigcup_{K \in \mathcal{T}_h} K$ is necessarily *polygonal, and to the case where the finite elements are all of Lagrange type.* These restrictions essentially avoid difficulties (of a purely technical nature) pertaining to appropriate definitions

(i) of "faces" of non polygonal elements in general and

(ii) of compatibility conditions for the degrees of freedom of adjacent finite elements along common faces in the case of Hermite finite elements.

**Remark 2.3.6.** There are indeed polygonal finite elements which are used in actual computations by engineers and which are neither $n$-simplices nor $n$-rectangles. Of course, such finite elements are not just arbitrary polygonal domains. Rather they are adapted to special circumstances: Thus, if the domain $\bar{\Omega}$ is a cylindrical domain in $\mathbf{R}^3$ it might be interesting to use *prismatic finite elements*, an example of which is given in Fig. 2.3.2.

In this case, the space $P$ is the tensor product of the space $P_1$ in the variables $x_1$, $x_2$ by the space $P_1$ in the variable $x_3$, i.e., a function $p$ in the space $P$ is of the form

$$p(x_1, x_2, x_3) = \gamma_1 + \gamma_2 x_1 + \gamma_3 x_2 + \gamma_4 x_3 + \gamma_5 x_1 x_3 + \gamma_6 x_2 x_3. \qquad \square$$

We shall assume that each polygonal set $K$ has a non-empty interior and that the interiors of the sets $K$ are pairwise disjoint, so that requirements $(\mathcal{T}_h i)$, $1 \leq i \leq 4$, are satisfied (a polygonal domain has a Lipschitz-continuous boundary). A portion $K'$ of the boundary of a polygonal finite element $K$ is a *face* of $K$ if it is a maximal connected subset of an affine hyperplane $\mathscr{H}$ of $\mathbf{R}^n$ with a nonempty interior relatively to $\mathscr{H}$.

In order to unambiguously define the functions of the finite element

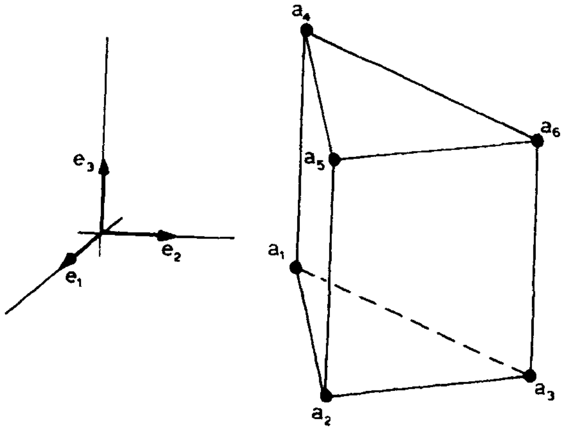


Fig. 2.3.2

space (see below), we need the following obvious extension of the condition ($\mathscr{T}_h5$) already seen for $n$-simplices and $n$-rectangles:

($\mathscr{T}_h5$) *Any face of a finite element $K_1$ is either a face of another finite element $K_2$, in which case the finite elements $K_1$ and $K_2$ are said to be adjacent, or a portion of the boundary $\Gamma$ of the set $\Omega$.*

Finally, the sets of degrees of freedom of adjacent finite elements shall be related as follows: *Whenever $(K_l, P_{K_l}, \Sigma_{K_l})$ with $\Sigma_{K_l} = \{p(a_i^l), 1 \leq i \leq N_l\}$, $l = 1, 2$, are two adjacent finite elements, then*

$$\left(\bigcup_{i=1}^{N_1}\{a_i^1\}\right)\cap K_2 = \left(\bigcup_{i=1}^{N_2}\{a_i^2\}\right)\cap K_1. \tag{2.3.22}$$

We define the set

$$\mathcal{N}_h = \bigcup_{K\in\mathscr{T}_h} \mathcal{N}_K \tag{2.3.23}$$

where, for each finite element $K \in \mathscr{T}_h$, $\mathcal{N}_K$ denotes the set of nodes. For each $b \in \mathcal{N}_h$, we let $K_\lambda$, $\lambda \in \Lambda(b)$, denote all those finite elements for which $b$ is a node. Then the associated *finite element space $X_h$* is the (generally proper) subspace of the product space $\Pi_{K\in\mathscr{C}_h} P_K$ defined by

$$X_h = \Big\{v = (v_K)_{K\in\mathscr{T}_h} \in \prod_{K\in\mathscr{T}_h} P_K;\quad \forall b \in \mathcal{N}_h,\quad \forall \lambda, \mu \in \Lambda(b),$$
$$v_{K_\lambda}(b) = v_{K_\mu}(b)\Big\}.$$

Therefore a function in the space $X_h$ is uniquely determined by the set

$$\Sigma_h = \{v(b),\quad b \in \mathcal{N}_h\}, \tag{2.3.24}$$

which is called the *set of degrees of freedom of the finite element space.*

It is thus realized that an element $v \in X_h$ is *not* in general a "function" defined over the set $\bar{\Omega}$, since it need not have a unique definition along faces common to adjacent finite elements. Nevertheless, by virtue of assumption (2.3.22), it is customary to say that the "functions" in the space $X_h$ are at least "continuous at all nodes common to adjacent finite elements" (the inclusions $P_K \subset \mathscr{C}^0(K)$, $K \in \mathscr{T}_h$, hold in practice). It is also a usual practice to consider the functions $v_K$. $K \in \mathscr{T}_h$, as being the *restrictions* to the finite elements $K$ of the function $v \in X_h$, just as if $v$ were an "ordinary" function defined over the set $\bar{\Omega}$. This is why we shall use the alternate notation $v|_K = v_K$.

If it happens, however, that for each function $v \in X_h$, the restrictions $v|_{K_1}$ and $v|_{K_2}$ coincide along the face common to any pair of adjacent finite elements $K_1$ and $K_2$, *then the function $v$ can indeed be identified with a function defined over the set $\bar{\Omega}$.*

**Remark 2.3.7.** Although this last property was true for all the examples of Section 2.2, it is by no means necessary. Following CROUZEIX & RAVIART (1973), let us consider for example the finite element space constructed with the following finite element $(K, P, \Sigma)$: The set $K$ is an $n$-simplex with vertices $a_j$, $1 \leq j \leq n+1$, the space $P$ is the space $P_1(K)$ and the set of degrees of freedom is the set $\Sigma_K = \{p(b_i),\ 1 \leq i \leq n+1\}$, where for each $i$ the point $b_i$ is the barycenter of the face which does not contain the point $a_i$, i.e.,

$$b_i = \frac{1}{n} \sum_{\substack{j=1 \\ j \neq i}}^{n+1} a_j, \quad 1 \leq i \leq n+1.$$

In Fig. 2.3.3, we have represented this finite element for $n = 2$ and $n = 3$.

To show that the set $\Sigma_K$ is $P_1(K)$-unisolvent, it suffices to observe that the points $b_i = (b_{ji})_{j=1}^n$, $1 \leq i \leq n+1$, are also the vertices of a (non-degenerate) $n$-simplex: If we let $B$ denote the $(n+1) \times (n+1)$ matrix

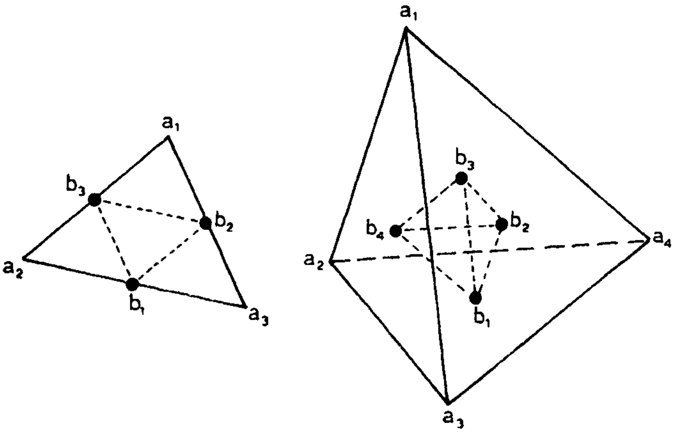


Fig. 2.3.3

defined by

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1,n+1} \\ b_{21} & b_{22} & \cdots & b_{2,n+1} \\ \vdots & & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{n,n+1} \\ 1 & 1 & \cdots & 1 \end{pmatrix},$$

it is easily verified that $\det B = (-1/n)^n \det A$, where $A$ is the matrix of (2.2.4) and thus $\det B \neq 0$. One may also notice that the functions

$$p_i = 1 - n\lambda_i, \quad 1 \leq i \leq n+1,$$

are the associated basis functions. Then it is clear that the functions of the corresponding finite element space generally have two definitions along faces common to adjacent finite elements, except at the centroids of these faces. □

All the previous considerations can be extended so as to include the case of finite element spaces constructed with Hermite finite elements, and the details are left to the reader (Exercise 2.3.8). We shall simply point out that it is often necessary to choose between various possible sets of degrees of freedom (corresponding to the same finite element) so as to unambiguously define a *set $\Sigma_h$ of degrees of freedom* of the corresponding finite element space. These considerations have been illustrated at various places in Section 2.2.

When the degrees of freedom of all finite elements are of some of the forms (2.3.4), the degrees of freedom of the finite element space are of some of the following forms:

$$\left.\begin{aligned} &v \to v(b_j^0), \\ &v \to Dv(b_j^1)\eta_{jk}^1, \\ &v \to D^2v(b_j^2)(\eta_{jk}^2, \eta_{jl}^2), \end{aligned}\right\} \tag{2.3.25}$$

where the points $b_j^r$, $r = 0, 1, 2$, called the *nodes of the finite element space*, make up a set which shall be generally denoted $\mathcal{N}_h$.

If we write the set $\Sigma_h$ as

$$\Sigma_h = \{\phi_{j,h}, \quad 1 \leq j \leq M\}, \tag{2.3.26}$$

then the *basis functions $w_j$, $1 \leq j \leq M$, of the finite element space* are defined by the relations

$$w_j \in X_h \quad\text{and}\quad \phi_{i,h}(w_j) = \delta_{ij}, \quad 1 \leq i \leq M. \tag{2.3.27}$$

We leave it to the reader to verify on each example that *the basis functions of the finite element space are derived from the basis functions of the finite elements*, as follows: Let $\phi_h \in \Sigma_h$ be of one of the form (2.3.25), let $b$ be the associated node, and let $K_\lambda$, $\lambda \in \Lambda(b)$, denote all the finite elements of $\mathcal{T}_h$ for which $b$ is a node (see Fig. 2.3.4 in the case of rectangles of type (2)).

For each $\lambda \in \Lambda(b)$, let $p_\lambda$ denote the basis function of the finite element $K_\lambda$ associated with the restriction of $\phi_h$ to $K_\lambda$. *Then the function* $w \in X_h$ *defined by*

$$w = \begin{cases} p_\lambda \text{ over } K_\lambda, & \lambda \in \Lambda(b), \\ 0 \text{ elsewhere}, \end{cases} \tag{2.3.28}$$

*is the basis function of the space* $X_h$ *associated with the degree of freedom* $\phi_h$.

As a practical consequence, *requirement* (FEM 3) (which was set up in Section 2.1) *is always satisfied in the examples*. The reader should refer to Fig. 2.2.8 where, on an example, it was shown that the basis functions constructed in this fashion have indeed "small" supports. The "worst" case concerns a basis function attached to a vertex, say $b$, of the triangulation. In this case, the corresponding support is the union of those finite elements which have $b$ as a vertex. In most commonly encountered triangulations in the plane, the number of such finite elements is very low (six or seven, for example).

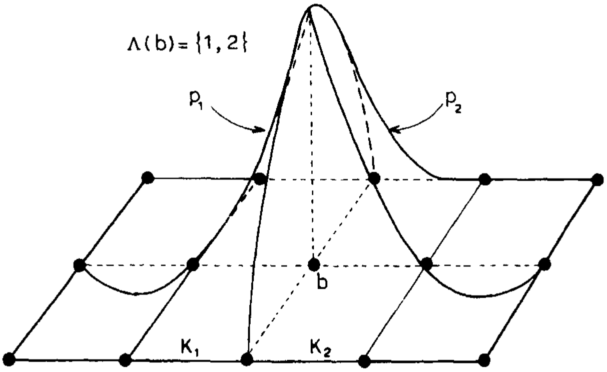


Fig. 2.3.4

Let there be given a finite element space $X_h$ with a set of degrees of freedom of the form (2.3.26). Then with any function $v: \bar{\Omega} \to \mathbf{R}$ sufficiently smooth so that the degrees of freedom $\phi_{j,h}(v)$, $1 \leq j \leq M$, are well defined, we associate the function

$$\Pi_h v = \sum_{j=1}^{M} \phi_{j,h}(v) w_j, \tag{2.3.29}$$

where the functions $w_j$ are the basis functions defined in (2.3.27). The function $\Pi_h v$, called the *$X_h$-interpolant* of the function $v$, is equivalently characterized by the conditions

$$\Pi_h v \in X_h \quad \text{and} \quad \phi_{j,h}(\Pi_h v) = \phi_{j,h}(v), \quad 1 \leq j \leq M. \tag{2.3.30}$$

If we let $s$ denote the maximal order of directional derivatives occurring in the finite elements $(K, P_k, \Sigma_K)$, $K \in \mathcal{T}_h$, we shall usually consider, in view of definition (2.3.7), that *the domain* $\operatorname{dom} \Pi_h$ *of the $X_h$-interpolation operator* $\Pi_h$ *is the space*

$$\operatorname{dom} \Pi_h = \mathscr{C}^s(\bar{\Omega}). \tag{2.3.31}$$

It might be helpful to keep in mind the following tableau (Fig. 2.3.5) where we have recapitulated the main "global" (i.e., on $\bar{\Omega}$) versus "local" (i.e., on a generic finite element $K$) notations, definitions and correspondences.

We next state a relationship of paramount importance between the

| *"Global" definitions* | *"Local" definitions* |
|---|---|
| $\bar{\Omega}$ | $K$: Finite element |
| Boundary of the set $\Omega$: $\Gamma$ | $\partial K$: Boundary of $K$ |
| | $K'$: Side, or face, of $K$ |
| Triangulation of the set $\bar{\Omega}$: $\mathcal{T}_h$ | |
| Finite element space: $X_h$ | $P$ or $P_K$ |
| Generic function of $X_h$: $v$ or $v_h$ | $p$ or $p_K$: Generic function of $P_K$ |
| Set of degrees of freedom of $X_h$: $\Sigma_h$ | $\Sigma$ or $\Sigma_K$: Set of degrees of freedom of $K$ |
| Degrees of freedom of $X_h$: $\phi_h$ or $\phi_{j,h}$, $1 \leq j \leq M$ | $\phi_i$ or $\phi_{i,K}$, $1 \leq i \leq N$: Degrees of freedom of $K$ |
| Basis functions of $X_h$: $w_j$, $1 \leq j \leq M$ | $p_i$ or $p_{i,K}$, $1 \leq i \leq N$: Basis functions of $K$ |
| Nodes of $X_h$: $b_j$ | $a_i$, $a_{ij}$, ...: Nodes of $K$ |
| $X_h$-interpolation operator: $\Pi_h$ | $\Pi$ or $\Pi_K$: $P_K$-interpolation operator |

Fig. 2.3.5

"global" interpolation operator $\Pi_h$ and the "local" interpolation operators $\Pi_K$.

**Theorem 2.3.2.** *Let $v$ be any function in the space* dom $\Pi_h$. *Then the restrictions $v|_K$ belong to the spaces* dom $\Pi_K$, *and we have*

$$\forall K \in \mathcal{T}_h, \quad (\Pi_h v)|_K = \Pi_K v|_K. \tag{2.3.32}$$

**Proof.** The above relations are direct consequences of the way in which the set $\Sigma_h$ is derived from the sets $\Sigma_K$, $K \in \mathcal{T}_h$. □

*Finite element of class $\mathscr{C}^0$ and $\mathscr{C}^1$*

It has always been assumed thus far that all the finite elements $(K, P_K, \Sigma_K)$, $K \in \mathcal{T}_h$, which are used in the definition of a finite element space are all *of the same type*: By this, we mean that, for instance, the finite elements are all $n$-simplices of type (2), or that the finite elements are all Argyris triangles, etc.... If this is the case, we shall say that any finite element $(K, P_K, \Sigma_K)$, $K \in \mathcal{T}_h$, is the *generic* finite element of the finite element space. We next state two definitions which are of particular importance, in view of Theorems 2.1.1 and 2.1.2.

We shall say that a finite element $(K, P_K, \Sigma_K)$ is *of class $\mathscr{C}^0$* if (i) the inclusion $P_K \subset \mathscr{C}^0(K)$ holds and (ii) whenever it is the generic finite element of a triangulation and $K_1$ and $K_2$ are two adjacent finite elements, the restrictions $v_h|_{K_1}$ and $v_h|_{K_2}$ coincide along the face common to $K_1$ and $K_2$ for any function $v_h$ of the corresponding finite element space. As a consequence, it is legitimate in this case to consider that the inclusion $X_h \subset \mathscr{C}^0(\bar{\Omega})$ holds.

Likewise, we shall say that a finite element $(K, P_K, \Sigma_K)$ of a given type is *of class $\mathscr{C}^1$* if (i) the inclusion $P_K \subset \mathscr{C}^1(K)$ holds and (ii) whenever it is the generic finite element of a triangulation and $K_1$ and $K_2$ are two adjacent finite elements, for any function $v_h$ in the corresponding finite element space the restrictions $v_h|_{K_1}$ and $v_h|_{K_2}$ coincide along the face $K'$ common to $K_1$ and $K_2$ and the outer normal derivatives $\partial_\nu v_h|_{K_1}$ and $\partial_\nu v_h|_{K_2}$ have a zero sum along $K'$. As a consequence, it is legitimate in this case to consider that the inclusion $X_h \subset \mathscr{C}^1(\bar{\Omega})$ holds.

Thus for instance, *all the finite elements seen in Section 2.2 are of class $\mathscr{C}^0$, and the Argyris and Bell triangles and the Bogner–Fox–Schmit rectangle are in addition of class $\mathscr{C}^1$. There are also finite* *elements which are not of class* $\mathscr{C}^0$, such as the one that was considered in Remark 2.3.7.

**Remark 2.3.8.** One may of course use finite elements of *different types* in a triangulation, provided some compatibility conditions are satisfied along faces which are common to adjacent finite elements, in such a way that a function in the space $X_h$ is still unambiguously defined on the one hand, and an inclusion such as $X_h \subset \mathscr{C}^0(\bar{\Omega})$ (for example) holds on the other hand. Thus one may combine triangles of type $(k)$ or $(k')$ with rectangles of type $(k)$ and still obtain the inclusion $X_h \subset \mathscr{C}^0(\bar{\Omega})$, etc.... Such an example is considered in Fig. 2.3.6. □

*Taking into account boundary conditions. The spaces* $X_{0h}$ *and* $X_{00h}$

The last topic we wish to examine in this section is *the way in which boundary conditions are taken into account in a finite element space.* Again, we shall essentially concentrate on examples.

Let $X_h$ be a finite element space whose generic finite element is any one of the following: $n$-simplex of type $(k)$, $k \geq 1$, or of type $(3')$, $n$-rectangle of type $(k)$, $k \geq 1$, rectangle of type $(2')$ or $(3')$. Then the inclusion $X_h \subset \mathscr{C}^0(\bar{\Omega}) \cap H^1(\Omega)$ holds (Theorems 2.2.3 and 2.2.7) and it follows that the inclusion

$$X_{0h} = \{v_h \in X_h;\quad v_{h|\Gamma} = 0\} \subset H_0^1(\Omega) \tag{2.3.33}$$

holds. In each of the above cases, it is easily verified that a sufficient (and obviously necessary) condition for a function $v_h \in X_h$ to vanish along $\Gamma$ is that *it vanishes at all the boundary nodes*, i.e., those nodes of the space $X_h$ which are on the boundary $\Gamma$. In other words, if we let $\mathcal{N}_h$ denote the set of nodes of the space $X_h$, the finite element space $X_{0h}$ of (2.3.33) is simply given by

$$X_{0h} = \{v_h \in X_h;\quad \forall b \in \mathcal{N}_h \cap \Gamma,\quad v_h(b) = 0\}. \tag{2.3.34}$$

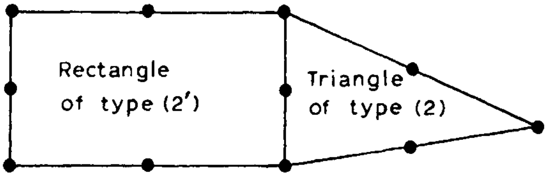


Fig. 2.3.6

When Hermite finite elements are used, the situation is less simple. Let us consider for example the case of Hermite $n$-simplices of type (3'), in which case the set of nodes of the triangulation coincides with the set of all the vertices of the triangulation. Then for each boundary node $b \in \mathcal{N}_h \cap \Gamma$, we let $\tau_\gamma(b)$, $\gamma \in \Gamma(b)$, denote a maximal set of linearly independent vectors in $\mathbf{R}^n$ with the property that the points $(b + \tau_\gamma(b))$, $\gamma \in \Gamma(b)$, belong to the boundary $\Gamma$. Following the notation introduced in Section 1.2, we recall that the differential operator $\partial_\tau$ is defined by $\partial_\tau v(a) = Dv(a)\tau$. Then the space $X_{0h}$ of (2.3.33) is given in this case by

$$\begin{aligned} X_{0h} = \{v_h \in X_h;\quad & \forall b \in \mathcal{N}_h \cap \Gamma,\quad v_h(b) = 0,\\ & \forall b \in \mathcal{N}_h \cap \Gamma,\quad \forall \gamma \in \Gamma(b),\quad \partial_{\tau_\gamma(b)} v_h(b) = 0\}. \end{aligned} \tag{2.3.35}$$

We have indicated in Fig. 2.3.7 the directional derivatives which must be set equal to zero along a specific portion of the boundary of a polygonal set in $\mathbf{R}^2$.

In particular, one should observe that at a corner such as $b^*$, the directional derivatives $\partial_{\mu_3} v(b)$ and $\partial_{\mu_4} v(b)$ must necessarily vanish.

If we next assume that the inclusion $X_h \subset \mathscr{C}^1(\bar{\Omega}) \cap H^2(\Omega)$ holds, it follows that we have the inclusions

$$X_{0h} = \{v_h \in X_h;\quad v_h|_\Gamma = 0\} \subset H^2(\Omega) \cap H_0^1(\Omega), \tag{2.3.36}$$

$$X_{00h} = \{v_h \in X_h;\quad v_h|_\Gamma = \partial_\nu v_h = 0\} \subset H_0^2(\Omega), \tag{2.3.37}$$

so that we are facing the problem of constructing such spaces $X_{0h}$ and $X_{00h}$. Again they are obtained by canceling appropriate values and directional derivatives at boundary nodes. As an example, we have indicated in Fig. 2.3.8 all the directional derivatives which must be set

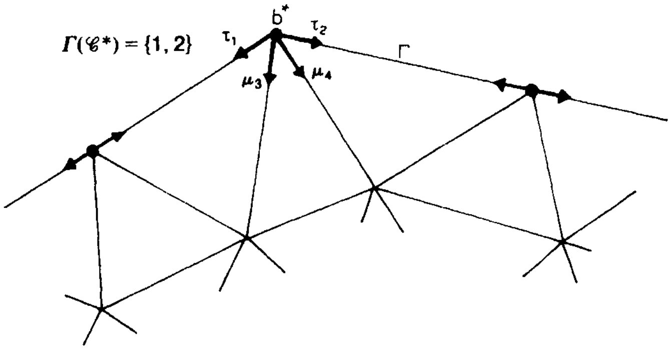


Fig. 2.3.7

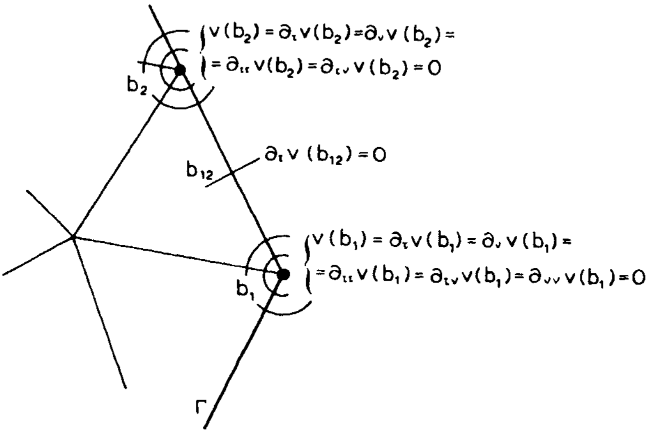

v(b₂) = ∂ₜv(b₂) = ∂ᵥv(b₂) =
= ∂ₜₜv(b₂) = ∂ₜᵥv(b₂) = 0
b₂
b₁₂
∂ₜv(b₁₂) = 0
v(b₁) = ∂ₜv(b₁) = ∂ᵥv(b₁) =
= ∂ₜₜv(b₁) = ∂ₜᵥv(b₁) = ∂ᵥᵥv(b₁) = 0
b₁
Γ

Fig. 2.3.8

equal to zero when Argyris triangles are used and the second inclusion (2.3.37) is needed.

It should be realized that at a boundary node such as $b_2$, the only "free" degree of freedom is $\partial_{\nu\nu}(b_2)$ while all degrees of freedom are zero at a corner such as $b_1$.

We shall also record for subsequent uses (particularly in the next chapter) the following crucial properties:

(i) *All finite elements of class $\mathscr{C}^0$ and of class $\mathscr{C}^1$ described in Section 2.2 have the property that*

$$v \in \operatorname{dom} \Pi_h \quad \text{and} \quad v|_\Gamma = 0 \Rightarrow \Pi_h v \in X_{0h}, \tag{2.3.38}$$

where the finite element space $X_{0h}$ is defined as in (2.2.33), or (2.3.36) for finite elements of class $\mathscr{C}^1$.

(ii) *All finite elements of class $\mathscr{C}^1$ described in Section 2.2 have the property that*

$$v \in \operatorname{dom} \Pi_h \quad \text{and} \quad v|_\Gamma = \partial_\nu v = 0 \Rightarrow \Pi_h v \in X_{00h}, \tag{2.3.39}$$

where the finite element space $X_{00h}$ is defined as in (2.3.37).

**Remark 2.3.9.** It is clearly possible to extend the previous analyses so as to include the case where boundary conditions are imposed only over a *portion* $\Gamma_0$ of the boundary $\Gamma$, provided such a portion $\Gamma_0$ is exactly the union of some faces of the finite elements found in the triangulation. □

*Final comments*

**Remark 2.3.10.** Let us briefly point out how some of the previously studied properties of finite elements and finite element spaces may be in fact derived from a *single "local" property*. For ease of exposition, we shall restrict ourselves to the case of Lagrange finite elements, leaving as a problem (Exercise 2.3.10) the case of Hermite finite elements.

Let $(K, P, \Sigma)$ be a Lagrange finite element, with $\mathcal{N}$ as its set of nodes, i.e., the set of degrees of freedom is of the form $\Sigma = \{p(a);\ a \in \mathcal{N}\}$. If $K'$ is any face of the set $K$, we let

$$\Sigma|_{K'} = \{p(a);\quad a \in \mathcal{N} \cap K'\}, \tag{2.3.40}$$

$$P(K') = \{p|_{K'}\colon K' \to \mathbf{R};\quad p \in P\}. \tag{2.3.41}$$

Then all the Lagrange finite elements heretofore described, with the exception of the finite element of Remark 2.3.7, have the property that *for each one of their faces, say $K'$, the set $\Sigma|_{K'}$ defined in* (2.3.40) *is $P(K')$-unisolvent, where the space $P(K')$ is defined in* (2.3.41). This is a crucial underlying basic property, which has the following easily established consequences:

(i) The $P_K$-interpolant of a function $v \in \operatorname{dom} \Pi_K$ which vanishes along a face $K'$ is also zero along $K'$. As a consequence, the global property (2.3.38) holds.

(ii) Let $\phi\colon p \in P \to p(a)$ be any one of the degrees of freedom of the finite element and let $p$ be the associated basis function. Then the function $p$ vanishes along any face which does not contain the node $a$. From the way the basis functions of the space $X_h$ are constructed from the basis functions of the finite elements (cf. (2.3.28)), we in turn deduce the "global" property that the basis functions of the space $X_h$ have indeed small supports (FEM 3).

(iii) Assume in addition that, for each pair $(K_1, K_2)$ of adjacent finite elements found in a triangulation, one has $P_{K_1}|_{K'} = P_{K_2}|_{K'}$ along the common face $K'$, and that the inclusions $P_K \subset \mathscr{C}^0(K)$, $K \in \mathscr{T}_h$, hold. Then the inclusion $X_h \subset \mathscr{C}^0(\bar{\Omega})$ hold. □

In the *choice* of a finite element for solving a given problem, the following considerations are usually taken into account:

(i) The finite element must be well adapted to the *geometry* of the problem. For example, assembling three-dimensional finite elements is not an easy task. This is especially true for tetrahedra, so that prismatic finite elements (Remark 2.3.6) are usually preferred whenever possible. In this respect, see the discussion in ZIENKIEWICZ (1971, Chapter 6). Geometrical considerations also justify the choice of curved finite elements instead of straight finite elements in the case of "particularly curved" domains.

(ii) The finite element must of course be appropriate for the problem to be solved. For conforming finite element methods, we have seen that this requires the use of finite elements of class $\mathscr{C}^0$ or $\mathscr{C}^1$. In addition, we shall see that a mathematical proof of convergence requires (among other things) the inclusions $P_1(K) \subset P_K$, $K \in \mathscr{T}_h$, for second-order problems and the inclusions $P_2(K) \subset P_K$, $K \in \mathscr{T}_h$, for fourth-order problems. Incidentally, the engineers were well aware of these conditions, which they discovered empirically, long before the mathematicians arrived!

(iii) Once the two previous criteria have been satisfied, it remains to obtain a linear system whose coefficients are easy to compute on the one hand and which is as easy as possible to solve on the other hand. We shall not go very far into this interesting and manifold aspect of finite element methods. However, we shall record two rules which tend to reduce certain computational difficulties:

A *first guideline* is that, if possible, the sets of degrees of freedom associated with a given node in the triangulation be all alike, so as to avoid different instructions depending on the node. This explains why Hermite $n$-simplices of type (3') may be preferred to Hermite $n$-simplices of type (3), or why Bell's triangles may be preferred to Argyris triangles, even though there is in both cases a decrease of one in the order of convergence, as we shall see (in addition, such choices slightly reduce the dimension of the resulting linear system).

A *second guideline* is that each node of the space should be common to the greatest number of finite elements. For example, the reader will easily convince himself that *for a given triangulation*, Hermite triangles of type (3) lead to a smaller linear system than triangles of type (3) (with the same order of convergence).

(iv) In addition, miscellaneous questions may be considered. For instance, one may argue in the above example that the use of Hermite triangles of type (3) introduces artificial constraints (the continuity of the first derivatives at the vertices) on the one hand, but on the other hand, this is an advantage if one needs to compute the stresses at the vertices. Likewise, one may argue that the use of Argyris triangles for solving a plate problem introduces artificial constraints (the continuity of the second derivatives at the vertices and "extra" boundary conditions as shown in Fig. 2.3.8) on the one hand, but on the other hand, this is an advantage if one needs to compute the bending moments at the vertices (such moments are obtained from the second partial derivatives of the solution), etc. . . .

To conclude this discussion, we shall simply emphasize the fact that, for all practical purposes, nothing replaces the numerical experience accumulated over the years by the engineers.

### *Exercises*

***2.3.1.*** Let the points $a_i$, $5 \leqslant i \leqslant 8$, be as in Fig. 2.2.10. Is $\{p(a_i),\ 5 \leqslant i \leqslant 8\}$ a $Q_1$-unisolvent set?

***2.3.2.*** Let there be given a triangle with vertices $a_i$, $1 \leqslant i \leqslant 3$. Among the following sets of degrees of freedom, which ones are $P_2$-unisolvent?

$$\Sigma = \{p(a_{iij}),\quad 1 \leqslant i, j \leqslant 3,\quad i \neq j\},$$

$$\Sigma' = \{p(a_i),\quad 1 \leqslant i \leqslant 3;\quad p(a_{iij}),\quad 1 \leqslant i < j \leqslant 3\},$$

$$\Sigma'' = \{p(a_i),\quad Dp(a_i)(a_{i+1} - a_i),\quad 1 \leqslant i \leqslant 3\}$$

(in $\Sigma''$, the indices are counted modulo 3).

***2.3.3.*** Let $(\hat{K}, \hat{P}, \hat{\Sigma})$ be a finite element with degrees of freedom of the form (2.3.4), and let $K$, $P$ and $\Sigma$ be defined through relations (2.3.11), (2.3.12), (2.3.13) and (2.3.14), with $F$ any invertible affine mapping.

(i) Show that the triple $(K, P, \Sigma)$ is a finite element.

(ii) Is a generalization possible so as to include more general (and smooth enough) invertible mappings $F$?

(iii) Let $(\hat{K}, \hat{P}, \hat{\Sigma})$ be an $n$-rectangle of type $(k)$. Then the above process, with $F$ affine, allows the derivation of finite elements which are parallelograms for $n = 2$, parallelepipeds for $n = 3$, etc. . . (such finite elements are seldom used in practice however). Describe the corresponding space $P$.

***2.3.4.*** Are two Bell's triangles affine-equivalent in general?

***2.3.5.*** Let $K$ be a triangle with vertices $a_i$, $1 \leqslant i \leqslant 3$.

(i) Show that the set

$$\begin{aligned}\Sigma = \{&p(a_i),\quad 1\leqslant i\leqslant 3;\quad Dp(a_i)(a_j-a_i),\quad 1\leqslant i,j\leqslant 3,\quad j\neq i;\\ &D^2p(a_i)(a_j-a_i,\quad a_k-a_i),\quad 1\leqslant i,j,k\leqslant 3,\quad j\neq i,\quad k\neq i;\\ &Dp(a_{ij})(a_k-a_{ij}),\quad 1\leqslant i<j\leqslant 3,\quad k\neq i,\quad k\neq j\}\end{aligned}$$

is $P_5(K)$-unisolvent. The corresponding finite element is called the *Hermite triangle of type* (5).

(ii) Show that this finite element differs in general from the Argyris triangle.

(iii) Is it a finite element of class $\mathscr{C}^1$?

*2.3.6.* Give a definition and a proof of the statement: "The barycentric coordinates are invariant through an invertible affine mapping". How is this fact reflected when the basis functions of finite elements, such as the $n$-simplices of type $(k)$, or the Hermite triangle of type (3′), are expressed in terms of barycentric coordinates?

*2.3.7.* Is a finite element space $X_h$ completely specified by the data of the spaces $P_K$, $K\in\mathscr{T}_h$, and of the inclusion $X_h\subset\mathscr{C}^0(\bar{\Omega})$?

*2.3.8.* Give a complete description of the construction of a finite element space made up of Hermite finite elements. In particular, extend condition (2.3.22) so as to include degrees of freedom which involve directional derivatives of the first and second order.

*2.3.9.* Let $K$ be a triangle with vertices $a_i$, $1\leqslant i\leqslant 3$. In each one of the following cases, prove the $P$-unisolvence of the set $\Sigma$ of degrees of freedom. Which finite elements are of class $\mathscr{C}^0$? (These finite elements have been considered by CROUZEIX & RAVIART (1973), who showed that they could be used for approximating the solution of the Stokes problem):

$$\text{(i)}\ P=\mathrm{V}\{\lambda_i^2,\quad 1\leqslant i\leqslant 3;\quad \lambda_i\lambda_j,\ 1\leqslant i<j\leqslant 3;\quad \lambda_1\lambda_2\lambda_3\},$$

$$\Sigma=\{p(a_i),\quad 1\leqslant i\leqslant 3;\quad p(a_{ij}),\ 1\leqslant i<j\leqslant 3;\quad p(a_{123})\}.$$

In particular show that the inclusion $P_2\subset P$ holds.

$$\begin{aligned}\text{(ii)}\ P=\mathrm{V}\{&\lambda_i^3,\quad 1\leqslant i\leqslant 3;\quad \lambda_i^2\lambda_j,\ \lambda_i\lambda_j^2,\quad 1\leqslant i<j\leqslant 3;\quad \lambda_i^2\lambda_{i+1}\lambda_{i+2},\\ &1\leqslant i\leqslant 3\ (\text{mod. } 3)\},\end{aligned}$$

$$\begin{aligned}\Sigma=\{&p(a_i),\quad 1\leqslant i\leqslant 3;\quad p(a_{iij}),\quad 1\leqslant i,j\leqslant 3,\quad i\neq j;\\ &p(a_i^*),\quad 1\leqslant i\leqslant 3\},\end{aligned}$$

where $a_i^*=\frac{3}{2}(1-\alpha)a_{123}+\frac{1}{2}(3\alpha-1)a_i$, $1\leqslant i\leqslant 3$, and $\alpha$ is any number which satisfies $0<\alpha<1$, $\alpha\neq\frac{1}{3}$. In particular show that the inclusion $P_3\subset P$ holds.

(iii) The space $P$ is the same as in (ii),

$$\Sigma=\{p(a_{ij}),\quad 1\leqslant i<j\leqslant 3;\quad p(a_{ij}^*),\quad 1\leqslant i,j\leqslant 3,\ i\neq j;\ p(b_i^*),\quad 1\leqslant i\leqslant 3\},$$

where

$$a_{ij}^*=\gamma_1a_i+\gamma_2a_j,\quad 1\leqslant i,j\leqslant 3,\quad i\neq j,$$

$$\gamma_1=\frac{1}{2}\left(1-\sqrt{\frac{3}{5}}\right),\quad \gamma_2=\frac{1}{2}\left(1+\sqrt{\frac{3}{5}}\right)$$

(notice that the points $\gamma_1$, $\frac{1}{2}$, $\gamma_2$ are the Gaussian quadrature points of the interval $[0,1]$),

$$b_i^*=\alpha a_i+\frac{1-\alpha}{2}(a_{i+1}+a_{i+2}),\quad 1\leqslant i\leqslant 3\ (\mathrm{mod.}\ 3),$$

and $\alpha$ is any number which satisfies $0<\alpha<1$, $\alpha\neq\frac{1}{3}$.

***2.3.10.*** Extend the content of Remark 2.3.10 so as (i) to cover the cases of Hermite finite elements and (ii) to obtain "local" conditions which imply the inclusion $X_h\subset\mathscr{C}^1(\bar{\Omega})$.

***2.3.11.*** The following finite element which resembles the Hermite triangle of type (3′) has sometimes been used for solving two-dimensional problems: The set $K$ is a triangle with vertices $a_i$, $1\leqslant i\leqslant 3$, the space $P$ is the space $P_3'(K)$, where

$$P_3'=\left\{p: x\in\mathbf{R}^2\rightarrow\sum_{i_1+i_2\leqslant 3}\gamma_{i_1i_2}x_1^{i_1}x_2^{i_2};\quad \gamma_{21}=\gamma_{12}\right\}$$

(obviously the inclusion $P_2(K)\subset P_3'(K)$ holds), and

$$\Sigma=\{p(a_i),\quad \partial_1p(a_i),\quad \partial_2p(a_i),\quad 1\leqslant i\leqslant 3\},$$

so that this finite element is of class $\mathscr{C}^0$.

(i) Is $\Sigma$ a $P_3'(K)$-unisolvent set?

(ii) Can this finite element be imbedded in an affine family?

## 2.4. General considerations on convergence

### *Convergent family of discrete problems*

Whereas up to now, our discussion has been concerned with *one* discrete problem, we shall now consider *families* of discrete problems.

More specifically, assume that we are approximating the solution $u$ of the variational equations

$$\forall v \in V, \quad a(u, v) = f(v), \tag{2.4.1}$$

where the space $V$, the bilinear form $a(\cdot, \cdot)$ and the linear form $f$ satisfy the assumptions of the Lax–Milgram lemma (Theorem 1.1.3). Confining ourselves to the case of *conforming finite element methods*, we consider a family $(V_h)$ of subspaces of the space $V$, where it is understood that the *parameter $h$* (which shall be given a specific meaning in Section 3.2) is the defining parameter of the family and *has limit zero.*

With each finite element space $V_h$ is associated the *discrete solution $u_h$* which satisfies

$$\forall v_h \in V_h, \quad a(u_h, v_h) = f(v_h). \tag{2.4.2}$$

Then we shall say that *the associated family of discrete problems is convergent*, or equivalently, that *convergence* holds, if, for *any* problem of the form (2.4.1) posed in the space $V$, one has

$$\lim_{h \to 0} \|u - u_h\| = 0, \tag{2.4.3}$$

where $\|\cdot\|$ denotes the norm in the space $V$.

### *Céa's lemma. First consequences. Orders of convergence*

We are therefore interested in giving sufficient conditions for convergence and, as a first result in this direction, we have the following basic *abstract error estimate.*

**Theorem 2.4.1.** (*Céa's lemma*). *There exists a constant $C$ independent upon the subspace $V_h$ such that*

$$\|u - u_h\| \leq C \inf_{v_h \in V_h} \|u - v_h\|. \tag{2.4.4}$$

*Consequently, a sufficient condition for convergence is that there exists a family $(V_h)$ of subspaces of the space $V$ such that, for each $u \in V$,*

$$\lim_{h \to 0} \inf_{v_h \in V_h} \|u - v_h\| = 0. \tag{2.4.5}$$

**Proof.** Let $w_h$ be an arbitrary element in $V_h$: It follows from (2.4.1) and (2.4.2) that $a(u - u_h, w_h) = 0$. Using the same constants $\alpha, M$ as in (1.1.3) and (1.1.19), we have, for any $v_h \in V_h$,

$$\begin{aligned}\alpha\|u - u_h\|^2 &\leq a(u - u_h, u - u_h) = a(u - u_h, u - v_h)\\ &\leq M\|u - u_h\|\,\|v - v_h\|,\end{aligned}$$

and the conclusion follows with $C = M/\alpha$. □

**Remark 2.4.1.** When the bilinear form is *symmetric*, there is a remarkable interpretation of the discrete solution: Since we have $a(u - u_h, w_h) = 0$ for all $w_h \in V_h$, it follows that *$u_h$ is the projection over $V_h$ of the exact solution $u$, with respect to the inner product $a(\cdot, \cdot)$*. Therefore, we have in this case

$$a(u - u_h, u - u_h) = \inf_{v_h \in V_h} a(u - v_h, u - v_h).$$

Using the $V$-ellipticity and the continuity of the bilinear form, we deduce

$$\|u - u_h\| \leq \sqrt{\frac{M}{\alpha}} \inf_{v_h \in V_h} \|u - v_h\|.$$

Thus we have obtained a "better" constant than in the proof of Theorem 2.4.1, since the constant $M$ is necessarily larger than the constant $\alpha$. □

The simple, yet crucial, inequality (2.4.4) shows that *the problem of estimating the error $\|u - u_h\|$ is reduced to a problem in Approximation Theory*: To evaluate the distance $d(u, V_h) = \inf_{v_h \in V_h} \|u - v_h\|$ between a function $u \in V$ and a subspace $V_h \subset V$. This explains why this problem will be a central theme of the next chapter, where we shall essentially prove results of the following type: Assuming appropriate smoothness on the function $u$, we shall show that the distance $d(u, V_h)$ is itself bounded by a constant (which usually involves norms of higher order derivatives of the function $u$) times $h^\beta$, for some exponent $\beta > 0$. As a consequence, we have the additional information that, *for a given solution $u$*, there exists a constant $C(u)$ independent of $h$ such that

$$\|u - u_h\| \leq C(u)h^\beta. \qquad (2.4.6)$$

If this is the case, we shall say that the *order of convergence* is $\beta$, or equivalently, that we have an $0(h^\beta)$ *convergence*, and we shall simply write

$$\|u - u_h\| = 0(h^\beta). \qquad (2.4.7)$$

Using more elaborated techniques, we shall also evaluate the difference $(u - u_h)$ in other norms, or semi-norms, than the norm of the space $V$ (which is either the $\|\cdot\|_{1,\Omega}$ or the $\|\cdot\|_{2,\Omega}$ norm), such as the $|\cdot|_{0,\Omega}$ and the $|\cdot|_{0,\infty,\Omega}$ norms (cf. Sections 3.2 and 3.3 respectively), and we shall also call *error* the corresponding norms $|u - u_h|_{0,\Omega}$, $|u - u_h|_{0,\infty,\Omega}$, etc....

Whereas a mathematician is generally satisfied with a sufficient condition for convergence such as that of Theorem 2.4.1, this condition rightly appears as a philosophical matter to many an engineer, who is much more concerned in getting (even rough) estimates of the error *for a given space* $V_h$: For practical problems, one chooses often one, sometimes two, seldom more, subspaces $V_h$, but certainly not an infinite family. *In other words, the parameter h never approaches zero in practice*!

Nevertheless, we found it worth examining such questions of convergence because (besides providing the subject of this book...) (i) the problem of estimating the error for a given $h$ (i.e., of getting a realistic estimate of the constant $C(u)$ which appears in inequality (2.4.6)) is at the present time not solved in a satisfactory way, and (ii) at least there is a "negative" aspect that few people contest: Presumably, a method should not be used in practice if it were impossible to mathematically prove its convergence....

## Bibliography and comments

*2.1.* The finite element method was first conceived in a paper by Courant (1943), but the importance of this contribution was ignored at that time. Then the engineers independently re-invented the method in the early fifties: The earliest references generally quoted in the engineering literature are those of Argyris (1954–1955), Turner, Clough, Martin & Topp (1956). The name of the method was proposed by Clough (1960). Historical accounts on the development of the method, from the engineering point of view, are given in Oden (1972a), and Zienkiewicz (1973).

It is only in the sixties that mathematicians, notably Mikhlin (1964, 1971), showed real interest in the analysis of the Galerkin and Ritz methods. Although they were not aware of the engineers contributions, it is interesting to notice that the approximate methods which they studied resembled more and more the finite element method, as exemplified by the basic contributions of Céa (1964), Varga (1966) (for the one-dimensional case), Birkhoff, Schultz & Varga (1968) (for the multidimensional – but still tensor-product like – case). Then the outbreak came with the paper of Zlámal (1968), which is generally regarded as the first mathematical error analysis of the "general" finite element method as we know it to-day.

***2.2 and 2.3.*** The finite elements which are described in these sections can be found in the book of Zienkiewicz (1971), where they are sometimes given different names. In this respect, the reader who wishes to look into the Engineering literature may consult the following table, which lists a few correspondences.

| *Name given in this book* | *Name given in Zienkiewicz' book* |
|---|---|
| Triangle or tetrahedron of type (1), (2), (3) | Linear, quadratic, cubic triangle or tetrahedron |
| Rectangle of type (1), (2), (3) | Linear, quadratic, cubic rectangle |
| Rectangle of type (2'), (3') | Quadratic, cubic rectangle of the serendipity family |
| 3-rectangle | right prism or rectangular prism |
| barycentric coordinates | area or volume co-ordinates |
| basis functions | shape functions |

Regarding the attribution of names to finite elements, we have tried to follow the most common usages.

In particular, Courant's triangle is named after Courant (1943). The rectangles of type (2') and (3') are also called *serendipity finite elements*, because their discovery required some ingenuity indeed! Other examples of serendipity finite elements may be found in Zienkiewicz (1971, p. 108, p. 121, p. 126), particularly for $n = 3$. We mention that Zlámal (1973d) has given an interesting alternate approach for such serendipity finite elements. The Zienkiewicz triangle is named after Bazeley, Cheung, Irons & Zienkiewicz (1965). The Argyris triangle is named after Argyris, Fried & Scharpf (1968), while Bell's triangle is named after Bell (1969). Although these last two finite elements have appeared in these and several other publications around 1968–1969 (cf. the references given in Zienkiewicz (1971, p. 209)), it was recently brought to the author's attention that they should also be attributed to Felippa (1966), where they appeared for the first time.

For the numerical handling of the Argyris triangle (derivation of the basis functions, etc...), the reader is referred to Argyris, Fried & Scharpf (1968). See also Thomasset (1974). Finally the Bogner–Fox–Schmit rectangle is named after Bogner, Fox & Schmit (1965). We also note that Theorem 3 of Zlámal (1968) yields an alternate proof of Theorem 2.2.11.

Whereas it is fairly easy to conceive finite element spaces contained in $\mathscr{C}^0(\bar{\Omega})$, the construction of finite element spaces contained in $\mathscr{C}^1(\bar{\Omega})$ is less obvious, as shown by the last three examples of Section 2.2 (and also by additional examples that will be seen in Section 6.1). In this direction, see the discussion in Zienkiewicz (1971, Section 10.3), whose heuristic considerations have been recently justified by a beautiful *result of* Ženíšek (1973, 1974), who has proved the following: Let $n = 2$, let $X_h$ be a finite element space for which all finite elements $K$ are triangles, and for which the spaces $P_K$ are spaces of polynomials, i.e., there exists some integer $l$ such that the inclusions $P_K \subset P_l(K)$ hold for all $K \in \mathcal{T}_h$ (therefore finite elements of class $\mathscr{C}^1$ using "singular functions", or of "composite" type, as described in Section 6.1 are excluded from the present analysis). Then for any integer $m \geqslant 0$, the inclusion $X_h \subset \mathscr{C}^m(\bar{\Omega})$ implies that, at each vertex $b$ of the triangulation, the linear forms $v_h \to \partial^\alpha v_h(b)$ are degrees of freedom of the space $X_h$ for all $|\alpha| \leqslant 2m$. As a corollary, the inequality $l \geqslant 4m + 1$ holds (the proof of the corollary is simple, but the proof of the first result is by no means trivial). Thus for instance the particular choice $m = 1$ shows that Bell's triangle is the optimal finite element for fourth-order problems inasmuch as the dimension of its space $P_5'(K)$ is the smallest possible, at least for conforming finite element methods using piecewise polynomial spaces.

Ženíšek (1972) has also extended his results to the case of higher dimensions, and there has been recently substantial interest in the study of the properties of finite element spaces whose functions are piecewise polynomials and which are contained in $\mathscr{C}^m(\bar{\Omega})$. In this respect, we mention Barnhill & Gregory (1975b), Délèze & Goël (1976), Morgan & Scott (1975, 1976), Scott (1974), Strang (1973, 1974a).

There is a large literature on the various aspects of the numerical implementation of the finite element method. We shall quote here only a few papers: Birkhoff & Fix (1974), Boisserie & Planchard (1971), Bossavit (1973), Bossavit & Frémond (1976), Descloux (1972a, 1972b), Fix & Larsen (1971), Fried (1971a, 1973a, 1973b), Frémond (1974), Goël (1968a, 1968b). We also mention Brauchli & Oden (1971), ODEN (1973a), ODEN & REDDY (1976a, Section 6.5) for the "conjugate basis functions" approach. The paper of FELIPPA & CLOUGH (1970) is a nice blend of mathematical analysis and practical aspects.

Finally, we mention that the definition of a finite element as a triple $(K, P, \Sigma)$ is due to CIARLET (1975).

***2.4.*** Céa's lemma (Theorem 2.4.1) appeared in CÉA (1964, Proposition 3.1) in the case of a symmetric bilinear form. It was independently rediscovered by VARGA (1966), and extended to the nonsymmetric case in BIRKHOFF, SCHULTZ & VARGA (1968, Theorem 13).