Chapter 2
Finite-Element Method

2.1 Basic Idea

2.1.1 Mathematical Modeling

Developments in numerical analysis in the twentieth century generated many methods for obtaining approximate solutions to partial differential equations. Compared with the finite-difference method, spectral methods, the finite-volume method, or the singularity method, the successes of the finite-element method are unquestionable, and its supremacy is quite justifiably acclaimed.

It goes without saying that the other methods have their specific fields of application, but the finite-element method literally revolutionized our ability to handle the often complex features of partial differential equations.

It is undoubtedly this extraordinary adaptability with regard to the solution of equations - whose intrinsic complexity is partly due to the shape of the regions of integration whenever one has to deal with real problems arising in industry - that led to the spectacular development of the finite-element method in the second half of the twentieth century.

Indeed, the problem here is no longer to find an analytic solution. These are cases in which the engineer will know immediately that standard techniques are sure to fail.

Let us just summarize what is involved in numerical approximation of partial differential equations, whichever method one may choose. In practice, when it comes to studying complex systems, which may be extremely varied in nature, the common approach is to resort to what is typically referred to as a model.

In the engineering sciences, many such models lead inevitably to differential or partial differential equations.

The problem then is to find a sufficiently reliable tool for simulation, able to predict the behavior of a system under a range of different conditions, rather than to stand by and measure the degree of damage in a real trial run.

J. Chaskalovic, *Mathematical and Numerical Methods for Partial Differential Equations*, Mathematical Engineering, DOI: 10.1007/978-3-319-03563-5_2,

However, in this first stage of modeling, when we select the key mechanisms governing the life and evolution of the given system and translate them into mathematical language, some information must necessarily be left by the wayside.

By its very essence, a model cannot reproduce every aspect of reality, rather as the mirror on the bathroom wall can produce only a two-dimensional picture of what is inevitably a three-dimensional body, even though it may do so with accuracy and elegance!

But this first level of approximation may prove to be fatal in the context of a mathematical model. Indeed, analysis of the model may well lead to the conclusion that there is no solution, in which case the model must be revised, and probably enriched with one or more mechanisms neglected on the first attempt.

The question of uniqueness must also be examined with great care. If the model generates several possible solutions, that too raises a general question about the legitimacy of such a situation with regard to the behavior of a real system.

In addition, the numerical methods implemented downstream must also integrate this question of multiple approximate solutions.

It is thus important to set up a global methodology for approximate solution that incorporates all the necessary precautions if we hope to obtain final results that make sense in the real-world system under investigation.

2.1.2 Formalism and Functional Framework for Partial Differential Equations

Once a model has been set up, the next task is to select a method for solution that best takes advantage of the mathematical aspects of the problem. In particular, the mathematical model must be manipulated into a form as well suited as possible for numerical approximation.

To exemplify this aspect of things, let us consider the two-dimensional Laplace–Dirichlet problem, which will provide a good illustration for later discussion.

So, let Ω be a bounded open subset of $\mathbb{R}^2$. The problem is to find a real-valued function u defined on Ω that solves

$$\textbf{(CP)} \qquad \begin{cases} -\Delta u = f \ \text{ in } \Omega, \\ \quad u = 0 \ \text{ on } \partial\Omega, \end{cases} \tag{2.1}$$

where f is a given function.

At this stage, it is important to note that this formulation is not complete, since the regularity of the boundary $\partial\Omega$ of the integration domain Ω is not specified, and

neither is the regularity of f, while it is clear that these features will have a significant bearing on the regularity of the solution u of the continuous problem **(CP)**, and hence on the space V within which one expects to find u.

For reasons to be explained shortly, we thus assume that the domain of integration Ω has a boundary $\partial\Omega$ with C^2 regularity. That is to say, the curvature is a continuous function of the curvilinear coordinate describing the boundary $\partial\Omega$.

Furthermore, if we assume that f belongs to $C^0(\Omega)$, it makes sense to look for solutions to the continuous problem **(CP)** in $C^2(\Omega)$, since this would guarantee that the Laplacian itself was continuous. One then speaks of classical solutions.

In this case, Poisson's equation can be reinterpreted, not as a functional equation, but at each point M of Ω in the following form:

Find u in $C^2(\Omega)$ solving

$$\textbf{(CP)} \quad \begin{cases} -\Delta u(M) = f(M), \ \forall\, M \in \Omega, \\ \quad u(M) = 0, \qquad \forall\, M \in \partial\Omega. \end{cases} \tag{2.2}$$

Naturally, the function f on the right-hand side does not always satisfy the regularity requirements of the space C^0.

Consider, for instance, a case in which f belongs to $L^2(\Omega)$. The Laplacian of the solution u (equal to $-f$) must then also be an element of $L^2(\Omega)$.

For this reason, we look for the solution u of the continuous problem **(CP)** in the Sobolev space $H^2(\Omega)$, because if this is the case, the Laplacian of u is then an element of $L^2(\Omega)$.

However, we should stress once again that this is merely a reasonable choice of space in which to seek solutions u, because the only obligation here, imposed by the assumed regularity of f, is to find a solution whose Laplacian belongs to $L^2(\Omega)$.

In principle, Poisson's equation can now no longer be considered pointwise, as it could when f was C^0, but must be treated as a functional equation. In fact, it must be considered as an equality in $L^2(\Omega)$, i.e., a mean squared equality, or an "energy" balance:

$$\Delta u + f = 0 \ \text{ in } \ L^2(\Omega) \quad \Longleftrightarrow \quad \int_\Omega (\Delta u + f)^2 \, \mathrm{d}\Omega = 0. \tag{2.3}$$

To end this discussion, note that treating Poisson's equation as a functional equation in $L^2(\Omega)$ nevertheless implies that this equation can be considered at each point M of Ω, apart from a set of measure zero (see Sect. 1.5.1).

For the reader unfamiliar with the idea of a set of measure zero, a first approach would be to say that Poisson's equation is satisfied at all points M of Ω with the exception of a countably infinite number of points of Ω. But we shall nevertheless consider Poisson's equation as a global equation like (2.3) expressed in $L^2(\Omega)$, rather than a local equation like (2.2).

2.1.3 Constructing a Variational Formulation

We now discuss the basic principles underlying the finite-element method. The main idea is to consider the unknown u, not as a scalar field that associates a real number $u(M)$ (to be determined) with each point M of Ω, but rather as an element of a space of functions V in which various search paths will be explored in order to identify the solution.

Concerning the notion of approximation, the problem is no longer to determine a numerical sequence $(\tilde{u}_1, \ldots, \tilde{u}_N)$ providing an approximation, as in the finite-difference method, to the values $(u_1, \ldots, u_N)$ of the solution u to the continuous problem **(CP)** at points M_j, $(j = 1, \ldots, N)$, chosen on a suitable mesh covering the domain of integration Ω.

The idea now is rather to devise a procedure that allows us to obtain an approximate function $\tilde{u}$. Naturally, in the end, knowing the solution u, or rather its approximation $\tilde{u}$, we will be able to evaluate $\tilde{u}$ at any point of the domain Ω, and not only at a limited set of points sitting on a predetermined mesh, as happens with the finite-difference method.

A second key feature of the finite-element method is the integral formulation of the continuous problem **(CP)**, known as a variational problem **(VP)**, as discussed in Chap. 1.

To obtain this, we consider a real-valued function v on Ω called the *test function*, which is not specified *a priori*, lying in a function space V to be constructed later.

We then multiply (2.1) by the *test function* v and integrate both sides of the equation over Ω to obtain

$$-\int_{\Omega} \Delta u \cdot v \, \mathrm{d}\Omega = \int_{\Omega} f v \, \mathrm{d}\Omega, \quad \forall v \in V. \tag{2.4}$$

This transformation is motivated by the historical traditions of the finite-element method, which was originally introduced as a generalization of the principle of virtual work in continuum mechanics (see, for example, [1]).

Indeed, the partial differential Eq. (2.1) of the continuous formulation **(CP)** can be construed as the fundamental principle of statics expressing the equilibrium of an elastic membrane subjected to a density of transverse forces f generating a displacement field u perpendicular to the membrane.

On the other hand, (2.4) expresses an "energy" formulation, where even the non-specialist will recognize that the right-hand side of (2.4) can be interpreted as the work done by external forces f causing a displacement field v, which remains arbitrary at this stage.

For its part, the left-hand side of (2.4) corresponds to the work done by internal forces, intrinsic to the deformation of the elastic medium Ω.

Furthermore, the transformation of the local expression of the problem **(CP)** into a global or integral formulation **(VP)** is motivated by the need for a formalism amenable to the idea of search paths in a function space V.

But this is precisely the situation in an integral formulation, in the sense that the values of the function at specific points M of Ω no longer appear directly, since only its average is apparent through the integral.

Formally, we may then apply Green's formula (1.159) from Lemma 1.25 to rewrite (2.4) in the form

$$\int_{\Omega} \nabla u \cdot \nabla v \, \mathrm{d}\Omega - \int_{\partial\Omega} \frac{\partial u}{\partial n} \, v \, \mathrm{d}\Gamma = \int_{\Omega} f v \, \mathrm{d}\Omega, \quad \forall v \in V. \tag{2.5}$$

We are now in a position to determine the characteristics of the space V.

The first point to make concerns the total conservation of information between the formulation of the continuous problem **(CP)** and the formulation of the variational problem **(VP)**.

For example, the Dirichlet condition $u = 0$ on the boundary $\partial\Omega$ of Ω cannot be taken into account directly in the integral formulation (2.5). Given that the future solution u of the variational problem **(VP)** must be one of the functions v of V, we simply require all functions v in V to satisfy the Dirichlet condition:

$$v = 0 \ \text{ on } \ \partial\Omega. \tag{2.6}$$

Then (2.5) can be written

$$\int_{\Omega} \nabla u \cdot \nabla v \, \mathrm{d}\Omega = \int_{\Omega} f v \, \mathrm{d}\Omega, \quad \forall v \in V. \tag{2.7}$$

The second point concerns the existence of the integrals in the formulation (2.7). Indeed, it is essential to impose adequate convergence conditions on the integrals in (2.7).

Since we are thinking of sufficient conditions for convergence, several functional frameworks may meet our needs. For reasons that will become clear later on, we consider the framework provided by the Sobolev spaces, which ensure all the desired properties, even beyond the issues we are dealing with here.

Convergence of the right-hand side of (2.7) is ensured simply by the upper bound guaranteed by the Cauchy–Schwarz inequality:

$$\left| \int_{\Omega} f v \, \mathrm{d}\Omega \right| \leq \int_{\Omega} |f v| \, \mathrm{d}\Omega \leq \left(\int_{\Omega} |f|^2 \, \mathrm{d}\Omega \right)^{1/2} \left(\int_{\Omega} |v|^2 \, \mathrm{d}\Omega \right)^{1/2}. \tag{2.8}$$

Therefore, since f is a given function in $L^2(\Omega)$, it suffices also to take v in $L^2(\Omega)$, in order to ensure convergence of the right-hand side of (2.7).

Regarding the convergence of the first integral on the left-hand side of (2.7), we once again consider absolute convergence of the integral and apply the Cauchy–Schwarz inequality as before:

$$\left|\int_\Omega \nabla u\cdot\nabla v\,\mathrm{d}\Omega\right| \le \int_\Omega |\nabla u\cdot\nabla v|\,\mathrm{d}\Omega \le \left(\int_\Omega |\nabla u|^2\,\mathrm{d}\Omega\right)^{1/2}\left(\int_\Omega |\nabla v|^2\,\mathrm{d}\Omega\right)^{1/2}. \tag{2.9}$$

Convergence of the left-hand side of (2.7) is thus ensured if we require the *test functions* v in V to have gradients belonging to $L^2(\Omega)$.

To sum up, we have established that the following are sufficient conditions for convergence of the integrals in (2.7):

$$v \in L^2(\Omega) \quad \text{and} \quad \nabla v \in [L^2(\Omega)]^2.$$

This explains why we choose the variational space V to be the Sobolev space $H^1(\Omega)$ introduced in Chap. 1.

We must then also add the homogeneous Dirichlet condition (2.6). Put another way, we set the functional space V as follows:

$$V \equiv H_0^1(\Omega) \equiv \left\{v : \Omega \to \mathbb{R},\ \ v \in L^2(\Omega),\ \ \nabla v \in [L^2(\Omega)]^2,\ \ v = 0 \text{ on } \partial\Omega\right\}. \tag{2.10}$$

We can now bring together all our results and state in full the variational problem **(VP)** that we will be discussing in the rest of the book:

$$\textbf{(VP)} \qquad \begin{cases} \text{Find } u \in H_0^1(\Omega) \text{ solution of} \\ \displaystyle\int_\Omega \nabla u\cdot\nabla v\,\mathrm{d}\Omega = \int_\Omega f v\,\mathrm{d}\Omega, \quad \forall\, v \in H_0^1(\Omega). \end{cases} \tag{2.11}$$

2.2 Existence, Uniqueness, and Regularity of a Weak Solution

2.2.1 Application to the Homogeneous Laplace–Dirichlet Problem

General results regarding the existence and uniqueness of solutions of differential equations or partial differential equations, or indeed, variational equations, remain a completely open question. The complexity of such results depends on the nature and structure of the equation or system of equations.

As far as variational formulations are concerned, there is a rather general formalism, which we discussed in Sect. 1.6, that provides theorems guaranteeing the existence and uniqueness of the solution of the differential problem under certain conditions.

This is precisely what happens for the variational problem **(VP)** defined by (2.11), for which we shall now propose a first application of the Lax–Milgram theorem to establish the existence and uniqueness of the solution whenever the data f belongs to $L^2(\Omega)$.

In order to apply the Lax–Milgram theorem, one must identify the space V, the bilinear form $a(.,.)$, and the linear form $L(.)$. The variational problem **(VP)** defined by (2.11) suggests introducing the following quantities.

Let V be the space in which we seek the solution u of the variational problem:

$$V \equiv H_0^1(\Omega). \tag{2.12}$$

The space $H_0^1(\Omega)$ is equipped with the natural norm $\| \, . \, \|_{H^1(\Omega)}$ on functions belonging to $H^1(\Omega)$.

Then, $\forall\, v \in H^1(\Omega)$, we set

$$\|v\|^2_{H^1(\Omega)} \equiv \int_\Omega v^2 \, \mathrm{d}\Omega + \int_\Omega \left(\frac{\partial v}{\partial x}\right)^2 \mathrm{d}\Omega + \int_\Omega \left(\frac{\partial v}{\partial y}\right)^2 \mathrm{d}\Omega. \tag{2.13}$$

This is a Hilbert norm for the space $H^1(\Omega)$ (see Sect. 1.20), and also for $H_0^1(\Omega)$, since it is a closed vector subspace of $H^1(\Omega)$.

Let $a(.,.)$ be the bilinear form defined by

$$\begin{aligned} a : V \times V &\longrightarrow \mathbb{R} \\ (u, v) &\longmapsto a(u, v) \equiv \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}\Omega. \end{aligned} \tag{2.14}$$

Likewise, let $L(.)$ be the linear form defined by

$$\begin{aligned} L : V &\longrightarrow \mathbb{R} \\ v &\longmapsto L(v) \equiv \int_\Omega f v \, \mathrm{d}\Omega. \end{aligned} \tag{2.15}$$

Then, the variational problem **(VP)** specified by (2.11) can be written in the following form:

$$\text{Find } u \in V \text{ solution of } a(u, v) = L(v), \quad \forall\, v \in V. \tag{2.16}$$

We now check the premises of the Lax–Milgram theorem (Theorem 1.11):

1. $a(.,.)$ is a continuous bilinear form. Bilinearity is obvious. Concerning continuity, consider any two elements u and v of $H_0^1(\Omega)$. Then,

$$|a(u, v)| \leq \int_\Omega \left|\nabla u \cdot \nabla v\right| \mathrm{d}\Omega \leq \left(\int_\Omega |\nabla u|^2 \, \mathrm{d}\Omega\right)^{1/2} \left(\int_\Omega |\nabla v|^2 \, \mathrm{d}\Omega\right)^{1/2}, \tag{2.17}$$

where we have used the Cauchy–Schwarz inequality. Now,

$$\int_\Omega |\nabla u|^2 \, \mathrm{d}\Omega = \int_\Omega \left[\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 \right] \mathrm{d}\Omega = \left\| \frac{\partial u}{\partial x} \right\|^2_{L^2(\Omega)} + \left\| \frac{\partial u}{\partial y} \right\|^2_{L^2(\Omega)}, \tag{2.18}$$

where $\|.\|_{L^2(\Omega)}$ is the natural norm on $L^2(\Omega)$, defined by

$$\forall u \in L^2(\Omega), \quad \|u\|_{L^2(\Omega)} \equiv \left(\int_\Omega |u|^2 \, \mathrm{d}\Omega \right)^{1/2}. \tag{2.19}$$

This implies that

$$\int_\Omega |\nabla u|^2 \, \mathrm{d}\Omega \leq \|u\|^2_{H^1(\Omega)}. \tag{2.20}$$

The inequality (2.17) then gives

$$|a(u, v)| \leq \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}, \tag{2.21}$$

and the continuity constant of the map $a(., .)$ is clearly equal to 1.

2. $a(., .)$ is a V-elliptic form. To establish this property, we must find a lower bound for the quantity $a(v, v)$ defined in (2.14). Now, for any function $v \in H^1_0(\Omega)$, we have

$$a(v, v) = \int_\Omega |\nabla v|^2 \, \mathrm{d}\Omega = \left\| \frac{\partial v}{\partial x} \right\|^2_{L^2(\Omega)} + \left\| \frac{\partial v}{\partial y} \right\|^2_{L^2(\Omega)}. \tag{2.22}$$

To find a lower bound for $a(v, v)$ relative to the $H^1(\Omega)$ norm, recall that for every function v in $H^1_0(\Omega)$, we have the Poincaré inequality (1.130).

This tells us that there is a constant $C(\Omega) > 0$ such that

$$\int_\Omega |v|^2 \, \mathrm{d}\Omega \leq C(\Omega) \int_\Omega |\nabla v|^2 \, \mathrm{d}\Omega. \tag{2.23}$$

We now add the square of the $L^2(\Omega)$ norm of the modulus of ∇v on each side of the Poincaré inequality (2.23), which brings in the square of the $H^1(\Omega)$ norm of the function v:

$$\|v\|^2_{H^1(\Omega)} \equiv \|v\|^2_{L^2(\Omega)} + \left\| \frac{\partial v}{\partial x} \right\|^2_{L^2(\Omega)} + \left\| \frac{\partial v}{\partial y} \right\|^2_{L^2(\Omega)} \tag{2.24}$$

$$\leq \left[1 + C(\Omega) \right] \left[\left\| \frac{\partial v}{\partial x} \right\|^2_{L^2(\Omega)} + \left\| \frac{\partial v}{\partial y} \right\|^2_{L^2(\Omega)} \right] \tag{2.25}$$

$$\leq \left[1 + C(\Omega) \right] a(v, v). \tag{2.26}$$

It thus follows that

$$a(v, v) \geq C' \|v\|^2_{H^1(\Omega)}, \tag{2.27}$$

where the coercivity constant C' is given by

$$C' = \frac{1}{1 + C(\Omega)}.$$

3. $L(.)$ is a continuous linear form. Once again, the linearity of $L(.)$ is obvious. It is particularly easy to find a bound for L because the data f is a function in $L^2(\Omega)$:

$$|L(v)| \leq \int_\Omega |fv| \, \mathrm{d}\Omega \leq \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \leq \|f\|_{L^2(\Omega)} \|v\|_{H^1(\Omega)}. \tag{2.28}$$

The continuity constant of the linear form L is thus $\|f\|_{L^2(\Omega)}$.

Summary: Since $H^1_0(\Omega)$ is closed in $H^1(\Omega)$ for the norm $\|.\|_{H^1(\Omega)}$, it is a Hilbert space for this norm. According to the Lax–Milgram theorem, there thus exists one and only one function in $H^1_0(\Omega)$ that solves the variational problem **(VP)** specified by (2.16).

Notes:

- When the space $H^1_0(\Omega)$ was given the norm induced from $H^1(\Omega)$, all the conditions of the Lax–Milgram theory were satisfied. However, if we consider the norm $|\,.\,|_1$ defined by

$$|v|_1 \equiv \left(\int_\Omega |\nabla v|^2 \, \mathrm{d}\Omega \right)^{1/2}, \tag{2.29}$$

  it can also be shown that the conditions of the Lax–Milgram theorem are satisfied. To do this, we exploit the fact that the natural norm on $H^1(\Omega)$ and the one defined by (2.29) are equivalent, according to Lemma 1.23.

- When the data f is less regular than we have supposed, i.e., $L^2(\Omega)$, the tools required for functional analysis of the variational problem go beyond the scope of this book. The interested reader is referred to more-specialized literature, such as the book by Brézis [2] or the series by Dautray and Lions [3].

We now state a lemma that gives an *a priori* estimate for the solution u of the variational problem **(VP)** specified by (2.11), by establishing a result concerning the continuous dependence of the solution u on the data f.

Lemma 2.1 *If Ω is a bounded open set with sufficiently regular boundary $\partial\Omega$, and if f is a given function in $L^2(\Omega)$ and u_f solves the variational problem **(VP)** specified by (2.11), the map Δ defined by*

$$\begin{aligned} \Delta : L^2(\Omega) &\longrightarrow H_0^1(\Omega), \\ f &\longmapsto \Delta(f) \equiv u_f, \end{aligned}$$

is linear and continuous. Put another way,

$$\exists\, C > 0 \text{ such that } \forall\, f \in L^2(\Omega), \quad \|u_f\|_{H^1(\Omega)} \leq C \|f\|_{L^2(\Omega)}. \tag{2.30}$$

Proof:

1. Linearity of Δ is clear by inspection.

   - Let (f_1, f_2) be a pair of functions in $L^2(\Omega) \times L^2(\Omega)$. Let us show that $\Delta(f_1 + f_2) = \Delta(f_1) + \Delta(f_2)$, or again, that $u_{f_1+f_2} = u_{f_1} + u_{f_2}$, if u_{f_1} and u_{f_2} are solutions of

   $$a(u_{f_1}, v) = L_{f_1}(v), \quad \forall v \in H_0^1(\Omega), \tag{2.31}$$

   $$a(u_{f_2}, v) = L_{f_2}(v), \quad \forall v \in H_0^1(\Omega), \tag{2.32}$$

   where $L_f(.)$ is the linear form $L(.)$ of the variational problem **(VP)** defined by (2.11), but specifying its dependence on f:

   $$L_f(v) \equiv \int_\Omega f v \, \mathrm{d}\Omega. \tag{2.33}$$

   Then using the bilinearity of the form $a(.,.)$ and the linearity of $L_f(.)$, we have

   $$\begin{aligned} a(u_{f_1} + u_{f_2}, v) &= L_{f_1}(v) + L_{f_1}(v) = L_{f_1+f_2}(v) \\ &= a(u_{f_1+f_2}, v), \quad \forall v \in H_0^1(\Omega). \end{aligned}$$

   This, in turn, implies that $u_{f_1} + u_{f_2}$ corresponds to the solution $u_{f_1+f_2}$.

   - Likewise, for every real number λ and every function f in $L^2(\Omega)$, we have

   $$\begin{aligned} a(u_{\lambda f}, v) &= L_{\lambda f}(v) = \lambda L_f(v) \\ &= \lambda a(u_f, v) = a(\lambda u_f, v), \quad \forall v \in H_0^1(\Omega). \end{aligned}$$

   Therefore, $u_{\lambda f} = \lambda u_f$, and hence $\Delta(\lambda f) = \lambda \Delta(f)$.

2. Continuity of the map Δ is shown as follows. It results from the ellipticity of the bilinear form $a(.,.)$, whence

$$\forall v \in H_0^1(\Omega), \quad \exists\, \alpha > 0 \text{ such that } a(v, v) \geq \alpha \|v\|^2_{H^1(\Omega)}. \tag{2.34}$$

So, if u solves the variational problem **(VP)** specified by (2.11), then by choosing the generic element $v \in H_0^1(\Omega)$ as solution u_f, we obtain

$$\begin{aligned} \alpha \left\| u_f \right\|^2_{H^1(\Omega)} &\leq a(u_f, u_f) = L(u_f) = \int_\Omega f u_f \, \mathrm{d}\Omega \\ &\leq \| f \|_{L^2(\Omega)} \left\| u_f \right\|_{L^2(\Omega)} \leq \| f \|_{L^2(\Omega)} \left\| u_f \right\|_{H^1(\Omega)}, \end{aligned} \tag{2.35}$$

where we have used the Cauchy–Schwarz inequality. After dividing both sides of (2.35) by $\left\| u_f \right\|_{H^1(\Omega)}$, we obtain finally

$$\| \Delta(f) \|_{H^1(\Omega)} \equiv \left\| u_f \right\|_{H^1(\Omega)} \leq \frac{\| f \|_{L^2(\Omega)}}{\alpha}, \tag{2.36}$$

which expresses the continuity of the linear map Δ.

The inequality (2.36) is called an *a priori* estimate of the solution u for the data f.

It puts an upper bound on the "energy" of the solution u as defined by the H^1 norm relative to the "energy" of the data f as defined by the L^2 norm.

We end this section with an equivalence result for the variational problem **(VP)** of (2.11) and a suitable minimization problem **(MP)**.

Lemma 2.2 *Let Ω be a bounded open set with sufficiently regular boundary $\partial\Omega$, and f a given function in $L^2(\Omega)$. If u is the unique solution of the variational problem **(VP)** of (2.11) belonging to $H_0^1(\Omega)$, then u is also the unique solution of the minimization problem **(MP)** specified by the following:*

$$\textbf{(MP)} \quad \begin{cases} \text{Find } u \in H_0^1(\Omega) \text{ solution of} \\ J(u) \equiv \min\limits_{v \in H_0^1(\Omega)} J(v), \\ \text{with } J(v) = \dfrac{1}{2} a(v, v) - L(v), \end{cases}$$

*where the bilinear form $a(.,.)$ and the linear form $L(.)$ are those specifying the variational problem **(VP)** in (2.11).*

Proof: This is a direct application of Theorem 1.13, where we note that the bilinear form $a(.,.)$ of the variational problem **(VP)** specified by (2.11) is symmetric.

Note. The existence and uniqueness of the solution u of the minimization problem **(MP)** can be obtained directly as a consequence of Stampacchia's theorem (Theorem 1.9) by observing that the space $H_0^1(\Omega)$ is a closed convex subset of $H^1(\Omega)$.

2.2.2 Application to the Inomogeneous Laplace–Dirichlet Problem

In this section, we consider the problem of the Laplacian with an inhomogeneous Dirichlet condition. More precisely, let Ω be a bounded open subset of $\mathbb{R}^2$ with a sufficiently regular boundary $\partial\Omega$. The aim will be to find a real-valued function u on Ω that solves

$$\textbf{(CP)} \quad \begin{cases} -\Delta u = f \ \text{ in } \ \Omega, \\ \quad u = g \ \text{ on } \ \partial\Omega, \end{cases} \tag{2.37}$$

where f and g are given functions in $L^2(\Omega)$ and $L^2(\partial\Omega)$, respectively.

If we hope to implement the Lax–Milgram theorem (Theorem 1.11), we cannot adapt our analysis to the Laplacian problem with an inhomogeneous Dirichlet condition if we take as functional context the space V defined by

$$V \equiv \left\{ v : \Omega \longrightarrow \mathbb{R}, \quad v = g \ \text{ on } \ \partial\Omega \right\}. \tag{2.38}$$

Indeed, if this were the case, the definition of V would not be consistent with the first premise of the Lax–Milgram theorem, which assumes that the set V has a vector space structure.

But it is clear that the space defined by (2.38) does not have a vector space structure, since if v_1 and v_2 are two elements of the space V, the linear combination $v_1 - v_2$ is zero on the boundary $\partial\Omega$. But this means that $v_1 - v_2$ does not belong to V, which contradicts the requirement of closure under linear combination that must be satisfied by any vector space.

To get around this difficulty, we use the technique of extending a function defined on the boundary $\partial\Omega$ to the whole of Ω. Naturally, the extension will not be unique, and the whole problem now is to obtain a sufficiently regular extension to be able to carry out the ensuing differential operations.

We shall thus assume the following result:

Lemma 2.3 *If g is in $L^2(\partial\Omega)$, there is a real-valued function G on Ω that is in $H^2(\Omega)$ and satisfies $G(x) = g(x)$ (a.e.) on $\partial\Omega$. The function G is an extension of g to the whole of Ω and is not unique.*

If we take this as given, we can introduce the following real-valued function U on Ω :

$$U(x) \equiv u(x) - G(x), \tag{2.39}$$

where u solves the Laplacian problem with an inhomogeneous Dirichlet condition specified by g, and G is the extension of g given by Lemma 2.3.

Substituting the change of variable (2.39) into the Laplace–Dirichlet problem (2.37), it is easy to see that the function U solves

$$\textbf{(CP)} \qquad \begin{cases} -\Delta U = f + \Delta G \equiv F \text{ in } \Omega, \\ U = 0 \text{ on } \partial\Omega, \end{cases} \tag{2.40}$$

where F is the new right-hand side of the Laplace equation.

Since we have assumed that f belongs to $L^2(\Omega)$, and also that the extension G is an element of $H^2(\Omega)$, it follows that ΔU belongs to $L^2(\Omega)$, and as a consequence, the function F on the right-hand side is also in $L^2(\Omega)$.

Finally, U solves the Laplacian problem with a homogeneous Dirichlet condition. We may thus conclude from the discussion in Sect. 2.2.1 that $H_0^1(\Omega)$ contains one and only one solution U of the variational problem **(VP)** specified by (2.11), provided that we replace the function f on the right-hand side by the function F defined in (2.40).

Let us end this investigation with an *a priori* estimate of the solution u of the continuous problem **(CP)** specified by (2.37):

Lemma 2.4 *If u is a solution of the continuous problem **(CP)** specified by (2.37), then*

$$\exists\, C > 0, \quad \textit{such that} \quad \|u\|_{H^1(\Omega)} \leq C \left[\|f\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega)} \right]. \tag{2.41}$$

Proof: Let V be a test function in $H_0^1(\Omega)$. Multiplying both sides of the partial differential Eq. (2.40) of the continuous problem **(CP)** and integrating the resulting equation over the whole of Ω, we obtain

$$-\int_\Omega (\Delta U) V \, \mathrm{d}\Omega - \int_\Omega (\Delta G) V \, \mathrm{d}\Omega = \int_\Omega f V \, \mathrm{d}\Omega. \tag{2.42}$$

We then use Green's formula (1.159) for each integral on the left-hand side of (2.42):

$$\int_\Omega \nabla U \cdot \nabla V \, \mathrm{d}\Omega - \int_{\partial\Omega} \frac{\partial U}{\partial n} V \, \mathrm{d}\Gamma + \int_\Omega \nabla G \cdot \nabla V \, \mathrm{d}\Omega - \int_{\partial\Omega} \frac{\partial G}{\partial n} V \, \mathrm{d}\Gamma = \int_\Omega f V \, \mathrm{d}\Omega. \tag{2.43}$$

Now V is in $H_0^1(\Omega)$, which means that the integrals over $\partial\Omega$ in (2.43) both vanish. The formulation **(VP)** associated with the continuous problem **(CP)** specified by (2.40) can thus be written as follows:

$$\textbf{(VP)} \quad \begin{cases} \text{Find } U \in H_0^1(\Omega) \text{ solution of} \\ \displaystyle\int_\Omega \nabla U \cdot \nabla V \, \mathrm{d}\Omega = \int_\Omega f V \, \mathrm{d}\Omega - \int_\Omega \nabla G \cdot \nabla V \, \mathrm{d}\Omega, \quad \forall v \in H_0^1(\Omega). \end{cases} \tag{2.44}$$

Choosing $V = U$ in the variational formulation (2.44), it follows that

$$\int_\Omega |\nabla U|^2 \, \mathrm{d}\Omega = \int_\Omega f U \, \mathrm{d}\Omega - \int_\Omega \nabla G \cdot \nabla U \, \mathrm{d}\Omega. \tag{2.45}$$

However, U belongs to $H_0^1(\Omega)$, so we can use the Poincaré inequality, viz.,

$$\exists\ \alpha > 0 \ \text{ such that } \ \alpha \, \|\nabla U\|^2_{L^2(\Omega)} \geq \|U\|^2_{L^2(\Omega)}. \tag{2.46}$$

The bound on U for the H^1 norm is obtained from

$$(1 + \alpha) \, \|\nabla U\|^2_{L^2(\Omega)} \geq \|U\|^2_{H^1(\Omega)}. \tag{2.47}$$

Equation (2.45) then yields

$$\frac{1}{1+\alpha} \|U\|^2_{H^1(\Omega)} \leq \int_\Omega f U \, \mathrm{d}\Omega - \int_\Omega \nabla G \cdot \nabla U \, \mathrm{d}\Omega. \tag{2.48}$$

We now apply the Cauchy–Schwarz inequality to obtain an upper bound for the right-hand side of the inequality (2.48):

$$\begin{aligned} \frac{1}{1+\alpha} \|U\|^2_{H^1(\Omega)} &\leq \|f\|_{L^2(\Omega)} \|U\|_{L^2(\Omega)} + \|\nabla G\|_{L^2(\Omega)} \|\nabla U\|_{L^2(\Omega)} \\ &\leq \left[\|f\|_{L^2(\Omega)} + \|\nabla G\|_{L^2(\Omega)} \right] \|U\|_{H^1(\Omega)}. \end{aligned}$$

Dividing both sides by $\|U\|_{H^1(\Omega)}$, we then have

$$\begin{aligned} \|U\|_{H^1(\Omega)} &\leq (1+\alpha) \left[\|f\|_{L^2(\Omega)} + \|\nabla G\|_{L^2(\Omega)} \right] \\ &\leq (1+\alpha) \left[\|f\|_{L^2(\Omega)} + \|G\|_{H^1(\Omega)} \right]. \end{aligned}$$

We substitute the expression (2.39) for U to obtain

$$\|u - G\|_{H^1(\Omega)} \leq (1+\alpha) \left[\|f\|_{L^2(\Omega)} + \|G\|_{H^1(\Omega)} \right]. \tag{2.49}$$

However, we also have

$$\|u\|_{H^1(\Omega)} - \|G\|_{H^1(\Omega)} \leq \left| \|u\|_{H^1(\Omega)} - \|G\|_{H^1(\Omega)} \right| \leq \|u - G\|_{H^1(\Omega)}, \tag{2.50}$$

whence (2.49) implies that

$$\|u\|_{H^1(\Omega)} - \|G\|_{H^1(\Omega)} \leq (1+\alpha)\left[\|f\|_{L^2(\Omega)} + \|G\|_{H^1(\Omega)}\right]. \tag{2.51}$$

This, in turn, means that we can write

$$\|u\|_{H^1(\Omega)} \leq (2+\alpha)\left[\|f\|_{L^2(\Omega)} + \|G\|_{H^1(\Omega)}\right]. \tag{2.52}$$

To conclude here, we need to replace the H^1 norm of G by the L^2 norm of g for the measure on the boundary $\partial\Omega$. Now, since g is the trace γ_0 of G on the boundary $\partial\Omega$, the trace theorem (Theorem 1.3) tells us that

$$\exists\ \beta > 0 \ \text{ such that } \ \|G\|_{H^1(\Omega)} \leq \beta\, \|g\|_{L^2(\partial\Omega)}. \tag{2.53}$$

Using the inequality (2.53) in (2.52), we finally have

$$\|u\|_{H^1(\Omega)} \leq (2+\alpha)\left[\|f\|_{L^2(\Omega)} + \beta\, \|g\|_{L^2(\partial\Omega)}\right] \leq C\left[\|f\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega)}\right], \tag{2.54}$$

where C can be defined by $C \equiv (2+\alpha)(1+\beta)$.

2.2.3 *Application to the Laplace–Neumann–Dirichlet Problem*

When the continuous problem **(CP)** specified by (2.1) is replaced by the Laplace–Neumann–Dirichlet problem, the boundary Γ of Ω comprises two complementary parts, Γ_1 and Γ_2, on which the Dirichlet condition and the Neumann condition are defined respectively.

In this case, the continuous problem **(CP)** is formulated as follows:

$$\textbf{(CP)} \qquad \begin{cases} -\Delta u = f \ \text{ in } \ \Omega, \\ u = 0 \ \text{ on } \ \Gamma_1, \\ \dfrac{\partial u}{\partial n} = g \ \text{ on } \ \Gamma_2, \end{cases} \tag{2.55}$$

where it is assumed that f and g belong to $L^2(\Omega)$ and $L^2(\Gamma_2)$, respectively.

As a consequence, it is not difficult to show that the new associated variational problem **(VP)** can be written thus:

$$\textbf{(VP)} \qquad \begin{cases} \text{Find } u \in H^1_{\Gamma_1}(\Omega) \text{ solution of} \\ \displaystyle\int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}\Omega = \int_\Omega f v \,\mathrm{d}\Omega + \int_{\Gamma_2} g v \,\mathrm{d}\Gamma, \quad \forall\, v \in H^1_{\Gamma_1}(\Omega), \end{cases} \tag{2.56}$$

where the Sobolev space $H^1_{\Gamma_1}(\Omega)$ is defined by

$$H^1_{\Gamma_1}(\Omega) \equiv \left\{ v : \Omega \to \mathbb{R},\ \ v \in L^2(\Omega),\ \ \nabla v \in [L^2(\Omega)]^2,\ \ v = 0 \text{ on } \Gamma_1 \right\}. \quad (2.57)$$

Notes:

1. When the Dirichlet data on Γ_1 is not homogeneous, the extension technique [3], although sometimes delicate, can be used to transform the inhomogeneous problem to a homogeneous one, the same as the one presented in (2.55).

2. The Lax–Milgram theorem is applied to the variational formulation (2.56) in an analogous way to what was discussed for the variational formulation (2.11) associated with the Laplace–Dirichlet problem.

   However, quite substantial changes are required to establish continuity of the linear form $L(.)$. Indeed, in this case, the action of the form $L(.)$ on any function v belonging to $H^1_{\Gamma_1}(\Omega)$ is

   $$L(v) \equiv \int_\Omega f v \, \mathrm{d}\Omega + \int_{\Gamma_2} g v \, \mathrm{d}\Gamma, \quad \forall\, v \in H^1_{\Gamma_1}(\Omega). \quad (2.58)$$

   The bound on $L(v)$ is then obtained as follows:

   $$\begin{aligned} |L(v)| &\le \int_\Omega |f v| \, \mathrm{d}\Omega + \int_{\Gamma_2} |g v| \, \mathrm{d}\Gamma \\ &\le \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} + \|g\|_{L^2(\Gamma_2)} \|v\|_{L^2(\Gamma_2)}. \end{aligned} \quad (2.59)$$

   A new difficulty arises here because of the Neumann condition g defined on the boundary Γ_2. The problem is that the bound for $L(v)$ must be expressed solely in terms of the $H^1(\Omega)$ norm of the function v.

   For this reason, the term resulting from the Neumann condition, which involves the $L^2(\Gamma_2)$ norm of v, must be modified accordingly.

   In fact, the trace theorem (Theorem 1.3) stated in Sect. 1.5.2 can be used to obtain a bound on $L(v)$ solely in terms of the $H^1(\Omega)$ norm of v. Let C_{tr} be the continuity constant for the trace map γ_0 defined in Theorem 1.3 of Sect. 1.5.2.

   We can then modify the inequality (2.59) as follows:

   $$\begin{aligned} |L(v)| &\le \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} + C_{\mathrm{tr}} \|g\|_{L^2(\Gamma_2)} \|v\|_{H^1(\Omega)} \\ &\le C \|v\|_{H^1(\Omega)}, \quad \forall\, v \in H^1_{\Gamma_1}(\Omega), \end{aligned}$$

   where we have set $C = \|f\|_{L^2(\Omega)} + C_{\mathrm{tr}} \|g\|_{L^2(\Gamma_2)}$.

These are the main remarks that need to be made about the extension of the homogeneous Laplace–Dirichlet problem (2.1) to the Laplace–Neumann–Dirichlet problem (2.55).

Other less important modifications concern adaptations of the result in going from the context of $H_0^1(\Omega)$ to that of $H_{\Gamma_1}^1(\Omega)$, but these present no major difficulties.

For this reason, once we have dealt with the bound on the linear form $L(.)$ defined by (2.58), application of the Lax–Milgram theorem guarantees existence and uniqueness of the solution $u \in H_{\Gamma_1}^1(\Omega)$ of the variational problem **(VP)** specified by (2.56).

2.3 Equivalence of Weak and Strong Formulations

There is one further important point regarding the transformation discussed in the last section, and in particular the question of the equivalence of the two formulations, i.e., the continuous problem **(CP)** and the associated variational problem **(VP)**.

For it should not be forgotten that the process of numerical approximation will produce an approximate solution to the variational formulation **(VP)**.

But then, what can be said about the solution to the continuous problem **(CP)** if no equivalence result is established between the two formulations?

The point is that it is no easy matter to show that a solution of a variational problem **(VP)** is a solution of the associated continuous problem **(CP)**. Worse, in many cases, it is not even true.

To grasp the subtlety of this notion of equivalence between the two formulations, let us return to the example of the Laplace–Dirichlet problem **(CP)** as specified by (2.1) and the associated variational formulation **(VP)** as specified by (2.11), and assume as before that f on the right-hand side is a function in $L^2(\Omega)$.

It suffices to point out that if the solution u of the continuous problem **(CP)** can be sought in the Sobolev space $H^2(\Omega)$, then the solution of the variational problem **(VP)** is sought in $H^1(\Omega)$ and $H^2(\Omega) \subset H^1(\Omega)$.

In other words, every solution of the continuous problem **(CP)** could be a solution of the variational problem as far as its regularity is concerned, whereas there is no *a priori* reason why a solution of the variational problem **(VP)** should be a solution of the continuous problem **(CP)**.

In fact, the notion of equivalence between the two formulations depends entirely on the function spaces in which one seeks solutions to the continuous problem **(CP)** on the one hand and the variational formulation **(VP)** on the other.

So let us examine in the context of the Laplace–Dirichlet problem how one might arrange for equivalence between the variational formulation **(VP)** and the continuous formulation **(CP)**.

Clearly, by construction, any solution of the continuous problem **(CP)** belonging to $H^2(\Omega)$ will be a solution of the variational problem **(VP)**. *A priori*, i.e., regardless of whether there are solutions, the regularity properties of a solution u of the

variational problem **(VP)** depend on the regularity of the function f on the right-hand side and also on the geometric properties of the boundary $\partial\Omega$ of the integration domain Ω, and this leaves us two possibilities for establishing that a solution of the variational problem **(VP)** is a solution of the continuous problem **(CP)**.

First Method

The first method is a partial converse, which can be stated as follows: a solution of the variational problem **(VP)** with the same regularity as the solutions of the continuous problem **(CP)**, i.e., belonging to $H^2(\Omega)$ and not just having the regularity ensured by $H^1(\Omega)$, is a solution of the continuous problem **(CP)**.

We thus consider u belonging to $H^2(\Omega)$, a solution of the variational problem **(VP)**. By Green's formula, we have

$$-\int_\Omega \Delta u \cdot v \, \mathrm{d}\Omega + \int_{\partial\Omega} \frac{\partial u}{\partial n} v \, \mathrm{d}\Gamma = \int_\Omega f v \, \mathrm{d}\Omega, \quad \forall v \in H_0^1(\Omega). \tag{2.60}$$

Since we know that v belongs to $H_0^1(\Omega)$, the integral over the boundary $\partial\Omega$ vanishes in (2.60). It follows that

$$\int_\Omega (\Delta u + f) v \, \mathrm{d}\Omega = 0, \quad \forall v \in H_0^1(\Omega). \tag{2.61}$$

Note how the $H^2(\Omega)$ regularity of the solution u has already served us in the application of Green's formula by guaranteeing convergence of the integral involving the Laplacian of u.

Indeed, applying the Cauchy–Schwarz inequality, we obtain the following upper bound:

$$\left| \int_\Omega \Delta u \cdot v \, \mathrm{d}\Omega \right| \leq \int_\Omega |\Delta u \cdot v| \, \mathrm{d}\Omega \leq \left(\int_\Omega |\Delta u|^2 \, \mathrm{d}\Omega \right)^{1/2} \left(\int_\Omega |v|^2 \, \mathrm{d}\Omega \right)^{1/2}. \tag{2.62}$$

Let us now examine the problem raised by (2.61). Looking at this family of equations, and there are as many equations here as there are functions v in $H_0^1(\Omega)$, we would like to be able to conclude that

$$\Delta u + f = 0 \ \text{ in } \ \Omega. \tag{2.63}$$

Now, the obvious way to justify the step from the integral Eq. (2.61) to the partial differential Eq. (2.63) would be to choose the specific function $v^* = \Delta u + f$ among all the functions v of $H_0^1(\Omega)$, because in that case, the integral in (2.61) would become

$$\int_{\Omega} (\Delta u + f)^2 \, \mathrm{d}\Omega = 0, \tag{2.64}$$

which would, in turn, imply Poisson's equation (2.63), since the integral of a non-negative function can vanish only if its integrand is identically zero.

But the problem is that we cannot necessarily choose the function v in $H_0^1(\Omega)$ to be the specific function $v^* = \Delta u + f$, because we do not know whether this choice lies in $H_0^1(\Omega)$, only that it belongs to $L^2(\Omega)$.

There are two ways to get around this problem. The first is to use a density technique, appealing to "contamination" by proximity. To do this, the key is to have a density theorem that allows us to transfer the desired property, namely the integral Eq. (2.61), to suitable functions v in $L^2(\Omega)$.

To be precise, we will implement a density theorem to show that (2.61) is valid, not only for every v in $H_0^1(\Omega)$, but also for every function v in $L^2(\Omega)$, whence we will be able to choose the specific function v^* equal to $\Delta u + f$.

Be warned, however! This is not a trivial result, because the inclusion of the function spaces is not in the sense that would allow us to apply the following kind of reasoning: he who can do more can do less.

Indeed, given that $H_0^1(\Omega)$ is contained in $L^2(\Omega)$ and not the opposite, we certainly cannot claim directly that (2.61) holds for every v in $L^2(\Omega)$.

We thus use the density theorem on page 35 (Theorem 1.1) to assert that for every function w in $L^2(\Omega)$, there is a sequence of functions w_n in $C_0^\infty(\Omega)$ that converges in the sense of the $L^2(\Omega)$ norm to the function w:

$$\lim_{n\to\infty} \int_{\Omega} |w_n - w|^2 \, \mathrm{d}\Omega = 0. \tag{2.65}$$

Furthermore, we also have the Sobolev embedding $C_0^\infty(\Omega) \subset H_0^1(\Omega)$. We can thus write (2.61) for each function in the sequence w_n, since these do belong to $H_0^1(\Omega)$:

$$\int_{\Omega} (\Delta u + f) w_n \, \mathrm{d}\Omega = 0, \quad \forall n \in \mathbb{N}. \tag{2.66}$$

We can now establish this same property for the functions w in $L^2(\Omega)$:

$$\begin{aligned} \left| \int_{\Omega} (\Delta u + f) w \, \mathrm{d}\Omega \right| &= \left| \int_{\Omega} (\Delta u + f)(w - w_n) \, \mathrm{d}\Omega \right| \\ &\leq \left(\int_{\Omega} |\Delta u + f|^2 \, \mathrm{d}\Omega \right)^{1/2} \left(\int_{\Omega} |w_n - w|^2 \, \mathrm{d}\Omega \right)^{1/2}. \end{aligned} \tag{2.67}$$

We then let n tend to $+\infty$ in the inequality (2.67) and use the convergence property (2.65) to conclude that for every function w in $L^2(\Omega)$, we have

$$\int_{\Omega} (\Delta u + f) w \, \mathrm{d}\Omega = 0, \quad \forall w \in L^2(\Omega). \tag{2.68}$$

The desired conclusion is now immediate, as pointed out earlier, since we can now choose the specific function w^* given by $w^* \equiv \Delta u + f$ in (2.68).

It follows that Poisson's equation is satisfied in $L^2(\Omega)$ for every solution u of the variational problem **(VP)** that has the further regularity implied by its belonging also to $H^2(\Omega)$.

Note once again the implications of this last property, for it is indeed the fact that Δu is a function of $L^2(\Omega)$ that allows us to apply the density theorem (Theorem 1.1).

The second way of concluding from (2.61) is to use the theory of distributions. We first note that if (2.61) is satisfied for every function v in $H_0^1(\Omega)$, then it must in particular remain true for every function v in $C_0^\infty(\Omega)$, since $C_0^\infty(\Omega)$ is obviously included in $H_0^1(\Omega)$.

In this case, (2.61) becomes

$$\int_{\Omega} (\Delta u + f) v \, \mathrm{d}\Omega = 0, \quad \forall v \in C_0^\infty(\Omega). \tag{2.69}$$

Now (2.69) can be interpreted in the sense of distributions as follows:

$$\langle \Delta u + f, v \rangle = 0, \quad \forall v \in C_0^\infty(\Omega). \tag{2.70}$$

Put another way,

$$\Delta u + f = 0 \ \text{ in } \ \mathscr{D}'(\Omega). \tag{2.71}$$

However, f and Δu both belong to $L^2(\Omega)$, since we assume that the solution u of the variational formulation **(VP)** belongs to $H^2(\Omega)$.

Hence, (2.71) also holds in $L^2(\Omega)$, and we have

$$\Delta u + f = 0 \ \text{ in } \ L^2(\Omega). \tag{2.72}$$

It follows that

$$\Delta u + f = 0 \ \text{ (a.e.) in } \ \Omega. \tag{2.73}$$

This clinches the partial converse once again. Indeed, if we have a solution u of the variational problem **(VP)** in $H_0^1(\Omega) \cap H^2(\Omega)$, then u must vanish on the boundary $\partial\Omega$ of Ω.

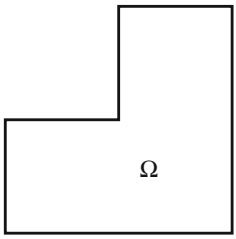

Fig. 2.1 Domain of integration Ω whose boundary has a discontinuous tangent

Second Method

The second converse regarding equivalence of the formulations **(VP)** and **(CP)** no longer considers each solution u of the variational problem **(VP)** to be in $H^2(\Omega)$, but starts from the assumption that the boundary $\partial\Omega$ of the open set Ω is C^2.

Indeed, the idea is to show that if the function f on the right-hand side is in $L^2(\Omega)$, then every solution u of the variational problem **(VP)** belonging to $H_0^1(\Omega)$ is also in $H^2(\Omega)$ for an integration domain Ω whose boundary $\partial\Omega$ has C^2 regularity (see, for example, [2]).

This is precisely what was assumed in Sect. 2.1.2. In this situation, we can, in fact, establish a full converse between the two formulations.

However, when the boundary contains corners of the kind depicted in Fig. 2.1, for example, it can be shown that the solution u of the variational problem **(VP)** no longer belongs to $H^2(\Omega)$.

Worse still, for certain boundary geometries, there are infinitely many solutions to the continuous problem **(CP)** (see, for example, [4]), while the variational problem **(VP)** has one and only one solution belonging to $H_0^1(\Omega)$.

Therefore, in situations in which the boundary $\partial\Omega$ exhibits geometric singularities, if we wish to maintain the equivalence of the two formulations, we must restrict the search for solutions of the continuous problem **(CP)** to the space $H_0^1(\Omega)$.

It can then be shown that there is one and only one function in $H_0^1(\Omega)$ that simultaneously solves the continuous problem **(CP)** and the variational problem **(VP)**.

In other words, since f belongs to $L^2(\Omega)$, Poisson's equation will be satisfied in $L^2(\Omega)$, hence almost everywhere. As a consequence, the solution u will have a Laplacian in $L^2(\Omega)$, while this does not mean that u will belong to $H^2(\Omega)$.

To end this converse, note that the homogeneous Dirichlet condition on $\partial\Omega$ is automatically satisfied for each solution u of the variational problem **(VP)**: because they lie in the space $H_0^1(\Omega)$, these functions have identically zero trace on $\partial\Omega$.

The aim of the above demonstration was to bring home to the reader, perhaps more forcefully than is usual, the importance of this kind of issue, which is sometimes

neglected for various reasons, but which, in fact, constitutes the backbone of mathematical modeling in the engineering sciences and ensures its credibility.

And make no mistake: it is only by careful treatment of these questions of methodological consistency that the numerical scientist will be able to build up a genuinely scientific approach to modeling for the purposes of decision-making and predicting the behavior of real systems.

2.4 Methodology and a Series of Approximations

We now examine the structure of the variational problem **(VP)** expressed in (2.11) and compare it with that of the continuous problem **(CP)** specified by (2.1) in order to identify and understand the mechanisms that might prevent us from applying analytic methods of solution.

For this would be a situation in which numerical approximation might prove fruitful if it were possible to eliminate some or all of the structural constraints that prevent solution, regardless of the mathematical formulation that might be adopted.

We begin with a first remark concerning the continuous problem **(CP)**. Two related mechanisms can prevent an analytic solution. The first is immediate and obvious. It lies in the complexity of the differential operations acting on the unknown function u.

Indeed, the combination of two second-order partial derivatives is undoubtedly a major problem for analytic solution, whatever the expression and complexity of the function f on the right-hand side.

Naturally, readers with experience in the standard techniques for solving partial differential equations (see, for example, [5]) might consider applying one or more miraculous transformations, such as those of Laplace, Fourier, and so on.

However, this would be to neglect the second factor entering into consideration for this type of problem, namely the shape of the integration domain, which may be fairly complex.

Indeed, when Ω has a regular shape, such as a square, circle, or ellipse, for instance, this tends to conceal the main difficulty.

For generally speaking, the problem here is to understand the formulation of the continuous problem **(CP)** from a different, perhaps unusual, angle, and in particular to realize that the continuous problem **(CP)** is a nonalgebraic system of equations, comprising an infinite number of equations in an infinite number of unknowns.

Indeed, Poisson's partial differential equation (2.1) has to hold at each point M of the domain of integration Ω, and as everyone knows, there are sure to be an infinite number of points within such an open set.

This means that we are actually considering, without explicitly acknowledging the fact, an infinite number of equations in an infinite number of unknowns that are nothing other than the values of the function u at each point M of Ω .

This is why the numerical scientist transforms the continuous problem into a finite-dimensional problem, for the human mind is not well equipped to comprehend the infinite.

The finite-element method, presented in many textbooks (see, for example, [5]), achieves this transformation by restriction, introducing a mesh, so that one need consider only the finite number of points M_i on a grid and approximations $\tilde{u}_i$ at these points that solve a system of algebraic equations obtained by approximating the partial derivatives, essentially with the help of Taylor's formula.

But in that case, does the transformation of the continuous problem **(CP)** to the variational formulation **(VP)** discussed above really generate a viable alternative? And if so, what do we gain by this new formulation, which, at first glance, seems to complicate the original continuous problem **(CP)**.

At the present stage, let us just say that the difficulty inherent in the continuous problem **(CP)** is carried over wholesale to the variational formulation **(VP)**.

Indeed, as we have just explained, it is the infinite number of unknowns and equations associated with the Poisson equation (2.1) that constitutes one of the main problems for solution.

And in the present case of the variational formulation **(VP)** as specified in (2.11), this problem of infinite dimensions is still extant, in the dimension of the search space V, here $H_0^1(\Omega)$, and as a consequence, in the infinite number of equations constituting the variational formulation (2.11).

For this reason, the approximation process adopted in the finite-element method, which is based on Galerkin's method, consists in considering a subspace $\tilde{V} \subset V$ that is in fact finite-dimensional. Let K_h be the dimension of this space $\tilde{V}$.

The transition from the variational problem **(VP)** to the approximate variational problem $\widetilde{\textbf{(VP)}}$ is made by replacing the pair of functions (u, v) in $V \times V$ by their approximations $(\tilde{u}, \tilde{v})$ in $\tilde{V} \times \tilde{V}$.

Hence, $\widetilde{\textbf{(VP)}}$ is expressed as follows:

$$\widetilde{\textbf{(VP)}} \quad \begin{cases} \text{Find } \tilde{u} \in \tilde{V} \text{ solution of} \\ \displaystyle\int_\Omega \nabla\tilde{u}\cdot\nabla\tilde{v}\,\mathrm{d}\Omega = \int_\Omega f\tilde{v}\,\mathrm{d}\Omega, \quad \forall\,\tilde{v} \in \tilde{V}. \end{cases} \tag{2.74}$$

The reader should not be beguiled by the apparent simplicity of the approximation procedure. The approximate variational formulation $\widetilde{\textbf{(VP)}}$ is not a simple rewriting of the exact formulation **(VP)**. On the contrary, it represents real progress with regard to the possibility of finding an approximate solution to the variational problem **(VP)**,

while of course it also corresponds to a loss of information that will be important to estimate later on.

In order to get a feel for the advantages brought about by this decisive step, let us introduce a basis $(\varphi_i)_{i=1,\dots,K_h}$ for the approximation space $\tilde{V}$, which we consider to have finite dimension K_h.

In this case, the unknown $\tilde{u}$ can be decomposed in the following form relative to the basis of functions φ_i:

$$\tilde{u} = \sum_{j=1}^{K_h} \tilde{u}_j \varphi_j . \tag{2.75}$$

Put another way, since (2.74) is valid for all $\tilde{v} \in \tilde{V}$, we are free to choose, among the approximate test functions $\tilde{v}$, each of the basis functions φ_i, $(i = 1, \dots, K_h)$, whereupon we set $\tilde{v} = \varphi_i$.

The approximate variational Eq. (2.74) can then be written as follows:

$$\widetilde{\textbf{(VP)}} \quad \begin{cases} \text{Find } \tilde{u}_j, \quad j = 1, \dots, K_h \text{ solution of} \\ \displaystyle\sum_{j=1}^{K_h} \left(\int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j \, \mathrm{d}\Omega \right) \tilde{u}_j = \int_\Omega f \varphi_i \, \mathrm{d}\Omega, \quad \forall\, i = 1, \dots, K_h. \end{cases} \tag{2.76}$$

We thus set

$$A_{ij} = \int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j \, \mathrm{d}\Omega, \qquad B_i = \int_\Omega f \varphi_i \, \mathrm{d}\Omega. \tag{2.77}$$

The approximate variational problem $\widetilde{\textbf{(VP)}}$ can then be expressed in the following form:

$$\widetilde{\textbf{(VP)}} \quad \begin{cases} \text{Find } \tilde{u}_j, \quad (j = 1, \dots, K_h) \text{ solution of} \\ \displaystyle\sum_{j=1}^{K_h} A_{ij} \tilde{u}_j = B_i, \quad \forall\, i = 1, \dots, K_h. \end{cases} \tag{2.78}$$

This last form clearly shows how the variational formulation **(VP)** is reduced by the approximation process to a finite-dimensional problem consisting of a system of K_h linear equations in K_h unknowns.

In order to implement the finite-element method, we must therefore specify the basis functions φ_i in (2.78), decide how to calculate the integrals in the matrix elements A_{ij} and the coefficients B_i on the right-hand side, and set up the algorithm for inverting the linear system, provided, of course, that the matrix with entries A_{ij} is in fact invertible.

From the theoretical standpoint, it is clear that one must check for invertibility of the matrix before applying the inversion algorithm to the linear system (2.78) (see, for example, [5]).

Returning now to the general problem of the approximation process described at the beginning of this section, we may summarize the successive transformations leading up to these approximations in the following diagram:

$$\begin{array}{lll} \text{Real system} & \longrightarrow & \text{Mathematical model or continuous problem } \mathbf{(CP)} \\ & \longrightarrow & \text{Variational formulation } \mathbf{(VP)} \\ & \longrightarrow & \text{Approximate variational formulation } \widetilde{\mathbf{(VP)}} \\ & \longrightarrow & \text{Method for inverting linear system } \widetilde{\widetilde{\mathbf{(VP)}}}. \end{array}$$

Viewing the process as a whole in this way should persuade the numerical scientist of the need for humility and caution when publishing final results. It is true that there are various theorems concerning error estimation in the context of the finite-element method, but as is often the case, this type of result is never all-encompassing, but concerns only a part of the above process.

In general, these theorems deal only with errors introduced when the variational problem **(VP)** is replaced by its approximate formulation $\widetilde{\mathbf{(VP)}}$ (see the Bramble–Hilbert lemma in the next section).

2.5 Variational Formulations and Approximations

Having laid down the fundamental principles underpinning the overall methodology, we now turn to the approximation of the variational formulations and the various choices available in the finite-element method.

It is this process as a whole that leads to the estimation of an approximate solution $\tilde{u}$ for both the variational problem **(VP)** and the continuous problem **(CP)** that inspired it.

As we have seen in (2.78), the variational problem **(VP)** associated with the Laplace–Dirichlet problem leads by Galerkin's method to an approximate formulation $\widetilde{\mathbf{(VP)}}$ that is nothing but a linear system to be inverted.

It is the solution of this linear system that generates the approximation $\tilde{u}$ to the solution of the variational problem **(VP)**, and hence an approximation to the solution of the continuous problem **(CP)**, whenever the two formulations **(CP)** and **(VP)** are indeed equivalent.

As a matter of fact, many mathematical models in engineering science lead to a formalism that is analogous to the one we have described for the Laplace–Dirichlet problem. This formalism can be viewed as a generic family of variational problems **(VP)** with the following abstract description:

$$\text{Find } u \text{ in } V \text{ which solves } \ a(u, v) = L(v), \quad \forall v \in V, \tag{2.79}$$

where

- V is a vector space of functions,
- $a(.,.)$ is a bilinear form on $V \times V$,
- $L(.)$ is a linear form on V.

Naturally, as already discussed in Sect. 2.3, further investigations involving suitable techniques of functional analysis will be needed if we are to obtain a variational formulation **(VP)** with one and only one solution that is, in addition, equivalent to the solution of the continuous problem **(CP)**.

However, in ensuring a good match between the continuous and variational problems, the approximation to the variational formulation (2.79) is unavoidable. This observation is intimately related to the infinite dimension of the function spaces arising in most mathematical models in the engineering sciences.

In order to get around the fact that we cannot solve formulations with the structure of (2.79), the method put forward by Galerkin considers a subspace $\widetilde{V} \subset V$ with finite dimension K_h. The abstract variational formulation (2.79) is then transformed to the following approximation $\widetilde{\textbf{(VP)}}$:

$$\text{Find } \tilde{u} \text{ in } \widetilde{V} \text{ which solves } \; a(\tilde{u}, \tilde{v}) = L(\tilde{v}), \quad \forall\, \tilde{v} \in \widetilde{V}. \tag{2.80}$$

Now that we have introduced an approximation space $\widetilde{V}$ of finite dimension K_h, it becomes possible, and indeed natural, to consider a basis of functions φ_i, $(i = 1, \ldots, K_h)$, and seek the approximation $\tilde{u}$, which replaces the solution u belonging to V, in the form

$$\tilde{u} = \sum_{j=1,\ldots,K_h} \tilde{u}_j \varphi_j. \tag{2.81}$$

Note that the decomposition (2.81) would not be possible without first making our transition from the space V, of infinite dimension, to its inner approximation $\widetilde{V}$, of finite dimension K_h.

Indeed, in the case of an infinite-dimensional function space V, apart from specific vector spaces in which every element can be decomposed relative to a basis containing a *countable* infinity of elements (as happens for separable Hilbert spaces, for instance), this kind of decomposition (2.81) remains impossible and so could not contribute to solving the variational problem **(VP)**.

But returning to the approximate variational formulation $\widetilde{\textbf{(VP)}}$ specified by (2.80), by choosing the functions $\tilde{v}$ equal to the basis functions φ_i, $(i = 1, \ldots, K_h)$, we may rewrite the formulation (2.80) in the following way:

Find $\tilde{u} = [\tilde{u}_1, \ldots, \tilde{u}_{K_h}]^{\mathrm{t}}$ in $\widetilde{V}$ that solves

$$a\left(\sum_{j=1,K_h} \tilde{u}_j \varphi_j, \varphi_i \right) = L(\varphi_i), \quad \forall\, i = 1, \ldots, K_h. \tag{2.82}$$

We then exploit the bilinearity of the form $a(.,.)$ and the linearity of the form $L(.)$. The variational formulation $\widetilde{\mathbf{(VP)}}$ can thus be expressed in the following form:

Find $\tilde{u} = [\tilde{u}_1, \ldots, \tilde{u}_{K_h}]^t$ in $\widetilde{V}$ that solves

$$\sum_{j=1,K_h} a(\varphi_j, \varphi_i)\tilde{u}_j = L(\varphi_i), \quad \forall i = 1, \ldots, K_h. \tag{2.83}$$

Finally, we introduce the quantities A_{ij} and b_i defined by

$$A_{ij} = a(\varphi_j, \varphi_i), \qquad b_i = L(\varphi_i). \tag{2.84}$$

The approximate variational formulation $\widetilde{\mathbf{(VP)}}$ then assumes the following final form:

Find $\tilde{u} = [\tilde{u}_1, \ldots, \tilde{u}_{K_h}]^t$ in $\widetilde{V}$ that solves

$$\sum_{j=1,K_h} A_{ij}\tilde{u}_j = b_i, \quad \forall i = 1, \ldots, K_h. \tag{2.85}$$

At this point, we observe that the formulation (2.85) is none other than a linear system involving a matrix A with entries A_{ij} and a right-hand side b with components b_i.

Put another way, we have just shown that every variational formulation **(VP)** that can be written in the form (2.79), and in which the forms $a(.,.)$ and $L(.)$ are respectively bilinear and linear, can be solved by an approximation with solution $\tilde{u}$ given by the linear system (2.85).

The problem at present is to select suitable parameters to proceed to an effective solution of the linear system (2.85) and thereby obtain an approximation to the variational problem **(VP)** specified by (2.79).

In order to calculate the coefficients A_{ij} and b_i, one must know the basis functions φ_i, $(i = 1, \ldots, K_h)$, of the approximation space $\widetilde{V}$. Naturally, this will depend intimately on the definition of the space $\widetilde{V}$ with finite dimension K_h.

For this reason, a first approach that allows us to fix the dimension K_h of the space $\widetilde{V}$ consists in relating this dimension K_h to a finite number of values of the functions $\tilde{v}$ belonging to $\widetilde{V}$ at preselected points or nodes M_k, $(k = 1, \ldots, K)$, of the integration domain Ω.

We now introduce an elementary geometry G_m, $(m = 1, \ldots, M)$, that generates a mesh on the domain of integration Ω, and hence also a set of nodes allowing a discretization of the problem (see Fig. 2.2). We thus arrive at the Lagrange finite-element method, defined as follows:

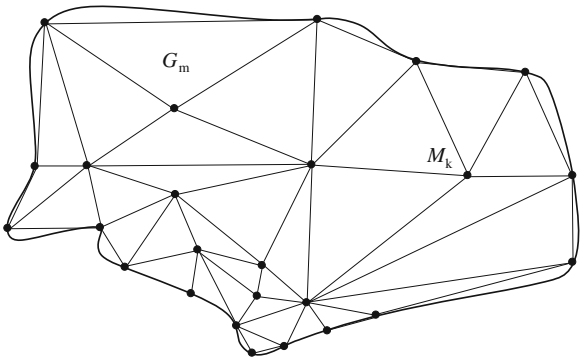

Fig. 2.2 Mesh consisting of triangular elements

Definition 2.1 The triplet $\big(G, \Sigma, P(G)\big)$ specifies a Lagrange finite-element method, where:

- G specifies the geometry of the primitive element of the mesh (segment, triangle, square, polyhedron, etc.).
- $\Sigma = (M_1, \ldots, M_{K'})$, $K' < K$, are the nodes at the vertices of, or otherwise delimiting, the primitive element G.
- $P(G)$ is the approximation space containing polynomials defined on G.

Finally, the triplet (G, Σ, P) must satisfy the property of *unisolvence*, defined as follows:

$$\forall\, \xi_1, \ldots, \xi_{K'} \in \mathbb{R}^{K'},\ \ \exists!\ p \in P(G) \text{ such that } \ p(M_k) = \xi_k,\ \ \forall k = 1, \ldots, K'. \tag{2.86}$$

In other words, there is one and only one function p in $P(G)$ that passes through the K' data points $(\xi_1, \ldots, \xi_{K'})$ at the K' nodes on the primitive element G.

Concerning unisolvence, it follows that $P(G)$ is isomorphic to $\mathbb{R}^{K'}$, and we have $\dim P(G) = K'$.

Once we have specified the functions in $P(G)$, defined on a generating element G, the approximation space $\widetilde{V}$ is constructed in the framework of the Lagrange finite-element method by setting

$$\widetilde{V} \equiv \Big\{\tilde{v} : \Omega \to \mathbb{R},\ \ \tilde{v} \in C^0(\Omega),\ \ \tilde{v}|_G \in P(G)\Big\}, \tag{2.87}$$

where we have set aside the question of the boundary conditions that may be imposed on the functions $\tilde{v}$ in $\widetilde{V}$, depending on the problem under consideration.

Consequently, if we ignore for the moment the boundary conditions, which will vary from one problem to another, the dimension of the space $\widetilde{V}$ specified by (2.87) can be deduced from the dimension K' of $P(G)$, the number of elements, and the number of nodes arising in the geometric discretization of the integration domain Ω.

The Lagrange finite-element method can be generalized. A general finite-element method is defined by the triplet $\big(G, \Sigma, P(G)\big)$, where:

- G is a primitive element of the geometric discretization of $\mathbb{R}^n$, $n = 1, 2$, or 3.
- Σ is a set of degrees of freedom σ_k, $k = 1, \ldots, K'$, consisting of linear forms on the space of functions defined on G.
- $P(G)$ is a vector space of finite dimension K'.
- Unisolvence: for a K'-tuplet of real numbers in $\mathbb{R}^{K'}$, there is a unique element p of $P(G)$ with $\sigma_k(p) = \xi_k$, $(\forall\, k = 1, \ldots, K')$.

2.6 Convergence of the Finite-Element Method

As emphasized in Sect. 2.4, the different levels of approximation arising in the series of steps that lead from the model to its numerical approximation require the user to treat the estimation of error with great care and humility.

In the present case, the finite-element method can furnish a body of theoretical results for estimating the approximation error between the solution u of a variational problem **(VP)** and its approximation $\tilde{u}$, which solves the approximate variational problem $\widetilde{\textbf{(VP)}}$.

Given the kind of mathematical objects involved here, i.e., the functions u and $\tilde{u}$, we shall describe in this section a family of results that allow us to estimate, relative to a suitable norm, the distance between the solution u and its approximation $\tilde{u}$, which we shall denote by $\|u - \tilde{u}\|$.

To illustrate the discussion, we shall refer to the family of variational problems **(VP)** with abstract form:

$$\text{Find } u \in V \text{ solution of } a(u, v) = L(v), \quad \forall\, v \in V. \tag{2.88}$$

As in the last section, $\widetilde{V} \subset V$ will denote the finite-dimensional approximation space contained in the space V, and $\tilde{v}$ the generic element of that space.

The approximation $\tilde{u}$ to the solution u will be a special case among these approximation functions $\tilde{v} \in \widetilde{V}$.

In other words, the approximate formulation $\widetilde{\textbf{(VP)}}$ of the variational problem **(VP)** can be written in the following form:

$$\text{Find } \tilde{u} \in \widetilde{V} \text{ solution of } a(\tilde{u}, \tilde{v}) = L(\tilde{v}), \quad \forall\, \tilde{v} \in \widetilde{V}. \tag{2.89}$$

Under the hypotheses of the Lax–Milgram theorem (Theorem 1.11 in Sect. 1.6), with Hilbert norm $\|\,.\,\|$, we have the following lemma:

Lemma 2.5 *The variational problem* $\widetilde{(VP)}$ *specified by (2.89) admits a unique solution* $\tilde{u}$*. Further, this solution satisfies the orthogonality relation*

$$a(u-\tilde{u},\tilde{v})=0,\quad \forall\,\tilde{v}\in\widetilde{V}. \tag{2.90}$$

Proof: The existence and uniqueness of the solution $\tilde{u}$ in $\widetilde{V}$ are immediate, precisely because $\widetilde{V}\subset V$. Indeed, we begin by noting that the finite-dimensional approximation space $\widetilde{V}$ contained in V is necessarily a closed vector subspace of V, and so inherits a Hilbert structure from V.

The fact that $\widetilde{V}$ is a subspace of V thus means that all the conditions are fulfilled for applying the Lax–Milgram theorem in $\widetilde{V}$.

Regarding the orthogonality relation (2.90), we simply rewrite the variational equation of **(VP)** with $\tilde{v}$ in the place of v, whence

$$a(u,\tilde{v})=L(\tilde{v}),\quad \forall\,\tilde{v}\in\widetilde{V}. \tag{2.91}$$

The difference between (2.91) and (2.89) leads immediately to the orthogonality relation (2.90).

The first error estimate $\|u-\tilde{u}\|$ is provided by Céa's lemma:

Lemma 2.6 *Under the hypotheses of the Lax–Milgram theorem (Theorem 1.11), and if we assume in addition that the approximation* $\tilde{u}$ *of the exact solution* u *lies in* $\widetilde{V}\subset V$*, we have the following error estimate:*

$$\|u-\tilde{u}\|\leq C\inf_{\tilde{v}\in\widetilde{V}}\|u-\tilde{v}\|\,. \tag{2.92}$$

Proof: The proof exploits the double bound on $a(u-\tilde{u},u-\tilde{u})$, obtained from the coercivity property, on the one hand, and the continuity of the bilinear form $a(.,.)$, on the other.

First of all, by the orthogonality relation (2.90) and choosing $\tilde{v}=\tilde{u}$, we have

$$a(u-\tilde{u},\tilde{u})=0. \tag{2.93}$$

We now rewrite $a(u-\tilde{u},u-\tilde{u})$ as follows:

$$\begin{aligned}\forall\,\tilde{v}\in\widetilde{V},\quad a(u-\tilde{u},u-\tilde{u})&=a(u-\tilde{u},u)-a(u-\tilde{u},\tilde{u})=a(u-\tilde{u},u)\\&=a(u-\tilde{u},u)-a(u-\tilde{u},\tilde{v})=a(u-\tilde{u},u-\tilde{v}).\end{aligned}$$

But since $a(.,.)$ is continuous and V-elliptic, $\exists\,(\alpha,\beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ such that

$$\alpha \|u - \tilde{u}\|^2 \leq a(u - \tilde{u}, u - \tilde{u}) = a(u - \tilde{u}, u - \tilde{v}) \leq \beta \|u - \tilde{u}\| \|u - \tilde{v}\|. \quad (2.94)$$

Dividing through by $\|u - \tilde{u}\|$, it then follows that

$$\|u - \tilde{u}\| \leq \frac{\beta}{\alpha} \|u - \tilde{v}\|, \quad \forall\, \tilde{v} \in \widetilde{V}, \quad (2.95)$$

whereupon the constant C in the statement of the lemma can be taken as the ratio of β and α.

Naturally, (2.95) is all the more useful as the bound on the norm $\|u - \tilde{u}\|$ can be made smaller. This is why the conclusion of Céa's lemma refers to the lower bound of the quantities $\|u - \tilde{v}\|$ for all functions $\tilde{v}$ in $\widetilde{V}$, i.e.,

$$\|u - \tilde{u}\| \leq C \inf_{\tilde{v} \in \widetilde{V}} \|u - \tilde{v}\|. \quad (2.96)$$

The next step in specifying the error estimate produced by Céa's lemma is to characterize the approximation space $\widetilde{V}$.

As mentioned in Sect. 2.5, the Lagrange finite-element method provides a simple solution for systematically producing a finite-dimensional approximation space $\widetilde{V}$. This process depends on the *unique* determination of an approximation function from the set of values it takes at a finite number of points M_k, $(k = 1, \ldots, K)$, arranged on a given mesh on the integration domain Ω.

At this point, the reader should recall the discussion of the Lagrange finite-element method in Sect. 2.5, and in particular, the fact that the dimension of the approximation space $\widetilde{V}$ corresponds to the number of nodes in the mesh on the region Ω, neglecting for the moment the boundary conditions that may have to be imposed on the approximation functions $\tilde{v}$.

More generally, we thus introduce the interpolation operator π_h defined by

$$\begin{aligned} \pi_h : C^0(\bar{\Omega}) &\longrightarrow \widetilde{V} \\ v &\longmapsto \pi_h v \equiv \sum_{k=1,\ldots,K} v(M_k)\varphi_k, \end{aligned} \quad (2.97)$$

where φ_k is the basis function in the approximation space $\widetilde{V}$ characterizing the node M_k, i.e., satisfying

$$\varphi_k(M_l) = \delta_{kl}, \quad (2.98)$$

with δ_{kl} the Kronecker symbol.

It is easy to check that the function $\pi_k v$, interpolated from v to the K nodes M_k of the mesh in the integration domain Ω, is the unique function of $\widetilde{V}$ satisfying

$$\pi_h v(M_k) = v(M_k), \quad \forall k = 1, \ldots, K. \tag{2.99}$$

We can then write the inequality of Céa's lemma in the particular case that $\tilde{v} = \pi_h u$, obtaining

$$\|u - \tilde{u}\| \leq C \|u - \tilde{v}\| = C \|u - \pi_h u\| . \tag{2.100}$$

According to the bound (2.100), the approximation error and the interpolation error will be of the same order of magnitude. For this reason, estimating the interpolation error provides an adequate method for measuring the approximation error, which depends on the nature and properties of each Lagrange finite-element method.

In order to make full use of Céa's lemma, we now discuss the Bramble–Hilbert lemma, which is based on these considerations. For present purposes, we limit the discussion to straight-edged but nonflat finite elements, and we take the variational space V to be the Sobolev space $H^1(\Omega)$, where Ω is a regular bounded open subset of $\mathbb{R}^2$.

Indeed, many problems arising in the engineering sciences correspond well to this functional framework, or are even more regular, but the reader should note that applications that cannot be formulated within this framework would require mathematical techniques of functional analysis that go well beyond the scope of the present book.

Lemma 2.7 *Let h measure the size of the primitive element of a Lagrange finite-element mesh. If the approximation space $\widetilde{V}$ contains the space P_k of polynomials of order less than or equal to k in the pair of variables (x, y), then for any sufficiently regular solution u, let us say at least in $H^1(\Omega)$, of a variational problem **(VP)** of the form (2.88), we have*

$$\|u - \pi_h u\|_{H^1(\Omega)} = O(h^k), \qquad \|u - \tilde{u}\|_{H^1(\Omega)} = O(h^k), \tag{2.101}$$

where $O(h^k)$ is Landau's notation indicating that there is a positive constant C such that

$$|O(h^k)| \leq Ch^k .$$

Naturally, the technical interest of this lemma lies in estimating the norm measuring the difference between the solution u and its interpolation $\pi_h u$.

The whole point of the preamble following Céa's lemma was to emphasize the need to estimate the latter norm if we are to draw conclusions about the approximation error in the finite-element analysis, at least in the context that we have described here.

2.7 Description of Commonly Used Finite-Element Methods

In this section we present the main finite-element geometries commonly encountered in the engineering sciences. We shall systematically present each finite-element method according to the following scheme:

1. Definition of the geometry G of the mesh element.
2. Definition of the approximation space $P(G)$, along with its dimension.
3. Definition of the set of linear forms σ_i on the function space defined on G.
4. Determination of the canonical basis functions for the space $P(G)$, i.e., the functions $p_1, \ldots, p_{\dim P(G)}$ satisfying $\sigma_i(p_j) = \delta_{ij}$, where δ_{ij} is the Kronecker symbol.

Note. The fact that there exists a collection of functions $p_1, \ldots, p_{\dim P(G)}$ in $P(G)$ satisfying the canonical property

$$\sigma_i(p_j) = \delta_{ij}, \quad \forall\, (i,\, j) \in \{1, \ldots, \dim P(G)\},$$

implies that this system of functions constitutes a basis for $P(G)$.

Indeed, let us show that the functions $p_1, \ldots, p_{\dim P(G)}$ are independent in $P(G)$. Suppose then that there are $(\alpha_1, \ldots, \alpha_{\dim P(G)}) \in \mathbf{R}^{\dim P(G)}$ such that

$$\sum_{i=1,\ldots,\dim P(G)} \alpha_i\, pi = 0. \tag{2.102}$$

Let us show that the condition (2.102) implies that the coefficients α_i are all zero. To do this, for some preselected j, we apply the j^{th} linear form σ_j to (2.102):

$$\sigma_j\left[\sum_{i=1,\ldots,\dim P(G)} \alpha_i\, pi\right] = \sigma_j(0). \tag{2.103}$$

We then use the linearity of the form σ_j and the fact that $\sigma_j(0) = 0$. Equation (2.103) can thus be written

$$\sum_{i=1,\ldots,\dim P(G)} \alpha_i \sigma_j(p_i) = \sum_{i=1,\ldots,\dim P(G)} \alpha_i \delta_{ij} = \alpha_j = 0, \ \forall\, j = 1, \ldots, \dim P(G). \tag{2.104}$$

It follows that the coefficients α_j all vanish and that the family $\{p_1, \ldots, p_{\dim P(G)}\}$ is therefore independent in this space of finite dimension $\dim P(G)$.

This same family must therefore span the approximation space $P(G)$ and thus constitutes a basis for it.

Note that we use the word "canonical" to express the fact that each function p_i in this particular basis is characteristic of a favored linear form σ_j, in the sense that the other linear forms are zero on this function p_i of the canonical basis.

In particular, in the case of Lagrange finite-element methods, the linear forms reflect a number of specific values of the functions of $P(G)$ at certain points (the discretization nodes) of the integration domain.

In this case, each function in the canonical basis corresponds to the unique function equaling 1 at a given node of the discretization and 0 at all the other nodes.

2.8 Fundamental Classes of Finite-Element Methods

2.8.1 Finite-Element Analysis in One Spatial Dimension

For the finite element methods discussed in this section, the mesh element is just the interval $G \equiv [0, 1]$.

- **P_0 Finite-Element Analysis**

1. The space $P(G) \equiv P_0$ comprises the polynomials p defined and constant on the interval $[0, 1]$. The dimension of P_0 is clearly 1.
2. We consider the linear form σ defined by

$$\sigma : p \longrightarrow \int_0^1 p(x)\,\mathrm{d}x. \tag{2.105}$$

3. The only function in the canonical basis is the constant function equal to 1 on the interval $[0, 1]$. To see this, we turn to the definition of the functions in the canonical basis discussed above, viz.,

$$\sigma(p) = 1 \iff \int_0^1 p(x)\,\mathrm{d}x = 1, \quad \text{where } p(x) = \text{constant}, \quad \forall\, x \in [0, 1]. \tag{2.106}$$

We deduce immediately that $p(x) = 1$, for all $x \in [0, 1]$.

For this first finite element, the functions $\tilde{v}$ belonging to $\widetilde{V}$ are constant functions on each mesh element. Note also that the constant on each mesh element corresponds to the average value of the function $\tilde{v}$ on the corresponding element.

- **P_1 Finite-Element Analysis**

1. The approximation space $P(G) \equiv P_1$ comprises the affine functions defined on the primitive mesh element $[0, 1]$. The dimension of the space P_1 is 2.

2. The two linear forms are

$$\sigma_1 : p \longrightarrow p(0), \qquad \sigma_2 : p \longrightarrow p(1). \tag{2.107}$$

3. To determine the functions in the canonical basis of the space P_1, we express the basic property of the two basis functions p_1, p_2:

$$\begin{aligned} &\sigma_1(p_1) = 1 \iff p_1(0) = 1, \qquad \sigma_1(p_2) = 0 \iff p_2(0) = 0, \\ &\sigma_2(p_1) = 0 \iff p_1(1) = 0, \qquad \sigma_2(p_2) = 1 \iff p_2(1) = 1. \end{aligned} \tag{2.108}$$

It is then straightforward to deduce that the basis functions p_1, p_2 solving (2.108) in the space P_1 of affine functions on the interval [0, 1] are

$$p_1(x) = 1 - x, \qquad p_2(x) = x. \tag{2.109}$$

- **P_2 Finite-Element Analysis**

1. The approximation space $P(G) \equiv P_2$ comprises the polynomials of degree less than or equal to two defined on the mesh element [0, 1]. The dimension of the space P_2 is 3.

2. Consider the three linear forms defined by

$$\sigma_1 : p \longrightarrow p(0), \qquad \sigma_2 : p \longrightarrow p(1/2), \qquad \sigma_3 : p \longrightarrow p(1). \tag{2.110}$$

3. We now express the defining property of the functions p_1, p_2, p_3 in the canonical basis for P_2 :

$$\begin{aligned} &\sigma_1(p_1) = 1 \iff p_1(0) = 1, \quad \sigma_1(p_2) = 0 \iff p_2(0) = 0, \\ &\sigma_1(p_3) = 0 \iff p_3(0) = 0, \quad \sigma_2(p_1) = 0 \iff p_1(1/2) = 0, \\ &\sigma_2(p_2) = 1 \iff p_2(1/2) = 1, \; \sigma_2(p_3) = 0 \iff p_3(1/2) = 0, \\ &\sigma_3(p_1) = 0 \iff p_1(1) = 0, \quad \sigma_3(p_2) = 0 \iff p_2(1) = 0, \\ &\sigma_3(p_3) = 1 \iff p_3(1) = 1. \end{aligned} \tag{2.111}$$

We then exploit the fact that each of the polynomials p_i, of degree less than or equal to two must have the form $ax^2 + bx + c$.

The nine relations (2.111) can be used to determine the nine coefficients of the three polynomials p_1, p_2, p_3. The result is

$$p_1(x) = (2x - 1)(x - 1), \;\; p_2(x) = 4x(1 - x), \;\; p_3(x) = x(2x - 1). \tag{2.112}$$

- **Hermite Finite-Element Analysis**

1. The approximation space $P(G) \equiv P_3$ comprises the polynomials of degree less than or equal to three, defined on the primitive mesh element [0, 1]. The dimension of the space P_3 is 4.
2. Consider the four linear forms defined by

$$\sigma_1 : p \to p(0), \quad \sigma_2 : p \to \frac{\mathrm{d}p}{\mathrm{d}x}(0), \quad \sigma_3 : p \to p(1), \quad \sigma_4 : p \to \frac{\mathrm{d}p}{\mathrm{d}x}(1). \tag{2.113}$$

3. We now determine the four functions p_1, p_2, p_3, p_4 in the canonical basis for P_3. To do this, we write down the 16 defining relations of the form $\sigma_i(p_j) = \delta_{ij}$:

$$\begin{array}{llll}
\sigma_1(p_1) = 1 \iff p_1(0) = 1, & \sigma_1(p_2) = 0 \iff p_2(0) = 0, \\
\sigma_1(p_3) = 0 \iff p_3(0) = 0, & \sigma_1(p_4) = 0 \iff p_4(0) = 0, \\
\sigma_2(p_1) = 0 \iff p_1'(0) = 0, & \sigma_2(p_2) = 1 \iff p_2'(0) = 1, \\
\sigma_2(p_3) = 0 \iff p_3'(0) = 0, & \sigma_2(p_4) = 0 \iff p_4'(0) = 0, \\
\sigma_3(p_1) = 0 \iff p_1(1) = 0, & \sigma_3(p_2) = 0 \iff p_2(1) = 0, \\
\sigma_3(p_3) = 1 \iff p_3(1) = 1, & \sigma_3(p_4) = 0 \iff p_4(1) = 0, \\
\sigma_4(p_1) = 0 \iff p_1'(1) = 0, & \sigma_4(p_2) = 0 \iff p_2'(1) = 0, \\
\sigma_4(p_3) = 0 \iff p_3'(1) = 0, & \sigma_4(p_4) = 1 \iff p_4'(1) = 1.
\end{array} \tag{2.114}$$

Once again, the 16 relations (2.114) allow us to obtain the 16 coefficients of the four polynomials p_1, p_2, p_3, p_4 in the canonical basis for P_3. The result is

$$\begin{array}{ll}
p_1(x) = (x-1)^2(2x+1), & p_2(x) = x(x-1)^2, \\
p_3(x) = x^2(3-2x), & p_4(x) = (x-1)x^2.
\end{array} \tag{2.115}$$

2.8.2 Finite-Element Methods in Two Spatial Dimensions

Triangular Meshes

In this section we discuss finite-element methods in which the mesh element G is an arbitrary triangle with vertices M_1, M_2, and M_3 in the plane $(O; x, y)$ (see Fig. 2.3).

- **P_0 Finite-Element Analysis**

1. The approximation space $P(G) \equiv P_0$ comprises the constant functions on the triangle G. The dimension of the space P_0 is 1.
2. We consider the linear form σ defined by

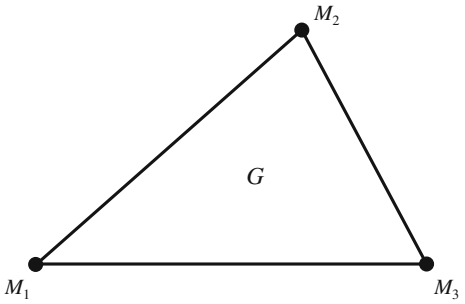

Fig. 2.3 Triangular mesh element

$$\sigma : p \longrightarrow \frac{1}{\text{area}(G)} \iint_G p(x, y)\, \mathrm{d}x\mathrm{d}y. \tag{2.116}$$

3. We now determine the basis function p of P_0 satisfying the defining property $\sigma(p) = 1$:

$$\sigma(p) = 1 \iff \frac{1}{\text{area}(G)} \iint_G p(x, y)\, \mathrm{d}x\mathrm{d}y = 1, \tag{2.117}$$

where $p(x,y) =$ constant, for all $(x,y) \in G$. This implies that the canonical basis function p is the constant function equal to 1 everywhere on the triangle G.

- ***P*$_1$ Finite-Element Analysis**

1. The approximation space P_1 comprises the polynomial functions of degree less than or equal to one for the pair of variables (x,y). In other words, every function p of P_1 can be written in the form

$$p(x,y) = ax + by + c, \tag{2.118}$$

where (a, b, c) is an arbitrary triplet in $\mathbb{R}^3$. The previous definition implies that the dimension of the space P_1 is 3.

2. We consider the three linear forms defined by

$$\sigma_1 : p \to p(M_1), \qquad \sigma_2 : p \to p(M_2) \qquad \sigma_3 : p \to p(M_3). \tag{2.119}$$

3. The three functions p_1, p_2, p_3 in the canonical basis correspond to the three barycentric functions $\lambda_1, \lambda_2, \lambda_3$, whose existence is proven in [5].

Note, however, that these polynomial functions in the pair of variables (x,y), with degree less than or equal to one, satisfy the canonical property

$$\sigma_j(\lambda_i) \equiv \lambda_i(M_j) = \delta_{ij}, \tag{2.120}$$

by definition.

- **P_2 Finite-Element Analysis**

1. The approximation space $P(G) \equiv P_2$ comprises the polynomial functions of degree less than or equal to two in the pair of variables (x, y). In other words, every function p in P_2 can be written in the form

$$p(x,y) = ax^2 + by^2 + cxy + dx + ey + f, \tag{2.121}$$

where (a, b, c, d, e, f) is an arbitrary point in $\mathbb{R}^6$. The definition (2.121) implies that the dimension of the space P_2 is 6.

2. To define the six linear forms $\sigma_i, i = 1, \ldots, 6$, we introduce three further nodes M_{12}, M_{13}, and M_{23} at the middle of each side of the triangle G, as shown in Fig. 2.4. The six linear forms are now defined as follows:

$$\sigma_1 : p \longrightarrow p(M_1), \qquad \sigma_2 : p \longrightarrow p(M_2), \tag{2.122}$$

$$\sigma_3 : p \longrightarrow p(M_3), \qquad \sigma_4 : p \longrightarrow p(M_{12}), \tag{2.123}$$

$$\sigma_5 : p \longrightarrow p(M_{13}), \qquad \sigma_6 : p \longrightarrow p(M_{23}). \tag{2.124}$$

3. The functions $p_1, p_2, p_3, p_4, p_5, p_6$ in the canonical basis are constructed as follows.

   Take the example of the function p_1. This second-degree polynomial in the pair (x, y) must vanish at the points M_2, M_3, M_{12}, M_{13}, and M_{23}.

   - On the segment M_2M_3, the polynomial p_1, whose trace is a trinomial of second degree in the variable parameterizing the segment M_2M_3, is identically zero, since it vanishes at the three points M_2, M_3, and M_{23}.

     But since the segment M_2M_3 is characterized by the equation $\lambda_1 = 0$, this means that the barycentric function λ_1 is a factor in the expression for the polynomial p_1.

   - Similarly, the polynomial p_1 is zero at the nodes M_{13} and M_{12}. Since the barycentric functions λ_i are affine in x and in y, at these two nodes λ_1 is equal to $1/2$ on the segment $M_{13}M_{12}$.

     In other words, factoring the quantity $\lambda_1 - 1/2$ from p_1, we ensure that p_1 is indeed zero at the nodes M_{13} and M_{12}.

   The polynomial structure of the function p_1 is thus

$$p_1(M) = \alpha\, \lambda_1(M) \left[\lambda_1(M) - \frac{1}{2}\right], \tag{2.125}$$

where α is a constant to be determined such that the polynomial p_1 takes the value 1 at its characteristic node, i.e., at the node M_1.

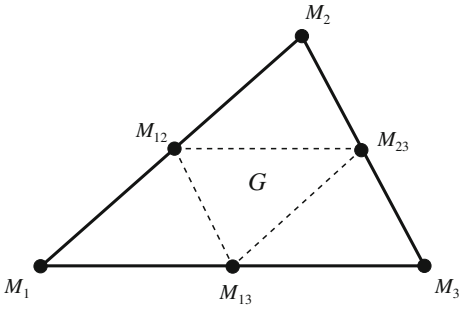

Fig. 2.4 Triangular element for the P_2 finite-element analysis

Note also that the expression (2.125) indeed gives p_1 the structure of a second-degree polynomial in the pair of variables (x, y), since the polynomial λ_1 is of first degree in the pair (x, y).

We can thus write

$$p_1(M_1) \equiv \alpha\, \lambda_1(M_1) \left[\lambda(M_1) - \frac{1}{2}\right] = \frac{1}{2}\,\alpha, \tag{2.126}$$

which ensures that $p_1(M_1) = 1$, whereupon we deduce that $\alpha = 2$.

Finally, the polynomial p_1 takes the form

$$p_1(M) = \lambda_1(M)\big[2\lambda_1(M) - 1\big]. \tag{2.127}$$

The other polynomials in the canonical basis are obtained by the same procedure.

The results are as follows:

$$p_1(M) = \lambda_1(M)(2\lambda_1(M) - 1), \qquad p_2(M) = \lambda_2(M)\big[2\lambda_2(M) - 1\big], \tag{2.128}$$

$$p_3(M) = \lambda_3(M)\big[2\lambda_3(M) - 1\big], \qquad p_{12}(M) = 4\lambda_1\lambda_2(M), \tag{2.129}$$

$$p_{13}(M) = 4\lambda_1\lambda_3(M), \qquad p_{23}(M) = 4\lambda_2\lambda_3(M). \tag{2.130}$$

- ***P*$_3$ Finite-Element Analysis**

1. The approximation space $P(G) \equiv P_3$ comprises the polynomial functions of degree less than or equal to three for the pair of variables (x, y).

   In other words, every function p in P_3 can be written in the form

$$p(x, y) = ax^3 + by^3 + cx^2y + dxy^2 + ex^2 + fy^2 + gxy + hx + iy + j, \tag{2.131}$$

   where $(a, b, c, d, e, f, g, h, i, j)$ is an arbitrary point in $\mathbb{R}^{10}$.

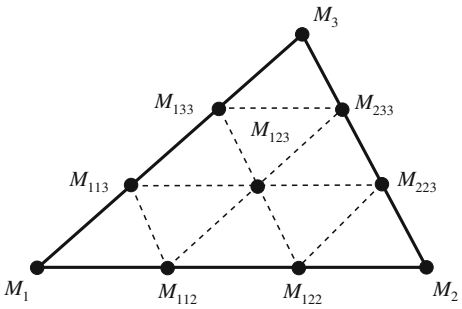

Fig. 2.5 Triangular element for the P_3 finite-element analysis

The definition (2.131) implies that the dimension of the space P_3 is 10.

2. To define the 10 linear forms σ_i, $(i = 1, \ldots, 10)$, we introduce seven further nodes M_{112}, M_{122}, M_{113}, M_{133}, M_{223}, M_{233}, and M_{123} a third of the way along each side of the triangle G, as shown in Fig. 2.5.

   The 10 linear forms are then defined as follows:

$$\sigma_1 : p \to p(M_1), \qquad \sigma_2 : p \to p(M_2), \tag{2.132}$$

$$\sigma_3 : p \to p(M_3), \qquad \sigma_4 : p \to p(M_{112}), \tag{2.133}$$

$$\sigma_5 : p \to p(M_{122}), \qquad \sigma_6 : p \to p(M_{223}), \tag{2.134}$$

$$\sigma_7 : p \to p(M_{233}), \qquad \sigma_8 : p \to p(M_{113}), \tag{2.135}$$

$$\sigma_9 : p \to p(M_{133}), \qquad \sigma_{10} : p \to p(M_{123}). \tag{2.136}$$

3. The 10 functions p_i, $(i = 1, \ldots, 10)$, of the canonical basis are determined by the same kind of arguments as were used for the P_2 triangular finite-element analysis.

   Consider again the polynomial p_1 characterizing the node M_1, i.e., satisfying $p_1(M_1) = 1$. Since it must vanish at the other nine nodes, we deduce the following factors:

   - λ_1 is a factor of p_1, because this polynomial must vanish at the nodes M_2, M_3, M_{223}, and M_{233}.
   - $(\lambda_1 - 2/3)$ must be a factor of p_1, because it vanishes at the nodes M_{112} and M_{113}.
   - $(\lambda_1 - 1/3)$ must be a factor p_1, because it vanishes at the nodes M_{122}, M_{133}, and M_{123}.

   The polynomial p_1 thus has the form

$$p_1(M) = \alpha\, \lambda_1(M) \left[\lambda_1(M) - \frac{1}{3}\right] \left[\lambda_1(M) - \frac{2}{3}\right], \tag{2.137}$$

where once again, the constant α is adjusted so that the polynomial p_1 is equal to 1 at the node M_1.

Note also that the form of the polynomial p_1 defined by (2.137) is consistent with the definition (2.131) of functions belonging to P_3, since the barycentric function λ_1 is a first-degree polynomial in the pair of variables (x, y).

It is straightforward to show that $\alpha = 9/2$, and the final form of p_1 is thus

$$p_1(M) = \frac{9}{2}\lambda_1(M)\left[\lambda_1(M) - \frac{1}{3}\right]\left[\lambda_1(M) - \frac{2}{3}\right]. \tag{2.138}$$

By the obvious symmetry, the polynomials p_2 and p_3 can be deduced immediately from the expression for p_1, giving

$$p_2(M) = \frac{9}{2}\lambda_2(M)\left[\lambda_2(M) - \frac{1}{3}\right]\left[\lambda_2(M) - \frac{2}{3}\right], \tag{2.139}$$

$$p_3(M) = \frac{9}{2}\lambda_3(M)\left[\lambda_3(M) - \frac{1}{3}\right]\left[\lambda_3(M) - \frac{2}{3}\right]. \tag{2.140}$$

Now consider the polynomial p_{112}. We can immediately extract certain factors:

- λ_1 is a factor in the expression for p_{112}, because this polynomial must vanish at the nodes M_2, M_3, M_{223}, and M_{233}.
- λ_2 must be a factor of p_{112}, because it must vanish at the nodes M_1, M_3, M_{113}, and M_{133}.
- $(\lambda_1 - 1/3)$ must be a factor of p_1, because it must vanish at the nodes M_{122}, M_{133}, and M_{123}.

Therefore, p_{112} has the structure

$$p_{112}(M) = \beta\,\lambda_1(M)\lambda_2(M)\left[\lambda_1(M) - \frac{1}{3}\right], \tag{2.141}$$

where the constant β is adjusted so that p_{112} is equal to 1 at the node M_{112}.

Since $\lambda_1 = 2/3$ and $\lambda_2 = 1/3$ at the node M_{112}, we obtain

$$\beta = \frac{27}{2}. \tag{2.142}$$

Finally, the basis function p_{112} has the form

$$p_{112}(M) = \frac{27}{2}\lambda_1(M)\lambda_2(M)\left[\lambda_1(M) - \frac{1}{3}\right]. \tag{2.143}$$

Once again, by the symmetry of the situation, the other basis functions p_{ijk}, for triplets (i, j, k) differing from $(1, 2, 3)$, are obtained immediately:

$$p_{122}(M) = \frac{27}{2}\lambda_1(M)\lambda_2(M)\left[\lambda_2(M) - \frac{1}{3}\right], \tag{2.144}$$

$$p_{113}(M) = \frac{27}{2}\lambda_1(M)\lambda_3(M)\left[\lambda_1(M) - \frac{1}{3}\right], \tag{2.145}$$

$$p_{133}(M) = \frac{27}{2}\lambda_1(M)\lambda_3(M)\left[\lambda_3(M) - \frac{1}{3}\right], \tag{2.146}$$

$$p_{223}(M) = \frac{27}{2}\lambda_2(M)\lambda_3(M)\left[\lambda_2(M) - \frac{1}{3}\right], \tag{2.147}$$

$$p_{233}(M) = \frac{27}{2}\lambda_1(M)\lambda_3(M)\left[\lambda_3(M) - \frac{1}{3}\right]. \tag{2.148}$$

We end by examining the last polynomial function in the canonical basis of P_3, viz., p_{123}.

This polynomial contains the following factors:

- λ_1 is a factor in the expression for p_{123}, because this polynomial must vanish at the nodes M_2, M_3, M_{223}, and M_{233}.
- λ_2 is a factor of p_{123}, because it must vanish at the nodes M_1, M_3, M_{113}, and M_{133}.
- λ_3 is a factor of p_{123}, because it must vanish at the nodes M_1, M_2, M_{112}, and M_{122}.

The function p_{123} thus has the following polynomial structure:

$$p_{123}(M) = \gamma\lambda_1(M)\lambda_2(M)\lambda_3(M), \tag{2.149}$$

where the constant γ is adjusted so that the polynomial p_{123} satisfies the characteristic property at the node M_{123}, namely $p_{123}(M_{123}) = 1$.

Since the barycentric functions λ_1, λ_2, and λ_3 are all equal to $1/3$ at the node M_{123}, we clearly obtain

$$\gamma = 27. \tag{2.150}$$

Finally, the polynomial p_{123} has the form

$$p_{123}(M) = 27\lambda_1(M)\lambda_2(M)\lambda_3(M). \tag{2.151}$$

Quadrilateral Meshes

In this section, we discuss finite-element methods in which the geometry G of the primitive mesh element is a square $[0, 1] \times [0, 1]$ with vertices M_1, M_2, M_3, and M_4 in the plane $(O; x, y)$, as shown in Fig. 2.6.

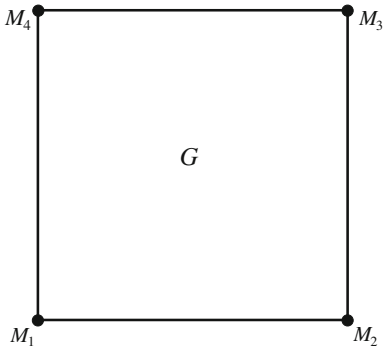

Fig. 2.6 Square primitive element for plane finite-element analysis

- **Q_1 Finite-Element Analysis**

1. The space $P(G) \equiv Q_1$ is defined as the set of polynomials of degree less than or equal to 1 in each of the two variables x and y. So every function p in Q_1 has the form

$$p(x,y) = axy + bx + cy + d, \tag{2.152}$$

where (a, b, c, d) runs over $\mathbb{R}^4$. By simple inspection of the definition (2.152), we see that Q_1 has dimension 4.

2. We introduce the four linear forms defined by

$$\sigma_i : p \longrightarrow p(M_i), \quad \forall\, i = 1, \ldots, 4. \tag{2.153}$$

3. To determine the four canonical basis functions p_i , $(i = 1, \ldots, 4)$, of the space Q_1, we recall that these functions must satisfy the definition:

$$\sigma_j(p_i) = p_i(M_j) = \delta_{ij}$$

So each of the canonical basis functions characterizes a single vertex of the square G, taking the value 1 at this vertex and 0 at all the others. We use this fact to examine the factorization properties of these functions.

Consider, for example, the polynomial p_1, which has the following properties:

- Since p_1 vanishes on the segment M_2M_3 parameterized by $x = 1$, the monomial $(x - 1)$ must be a factor in the expression for p_1.
- Since p_1 vanishes on the segment M_3M_4 parameterized by $y = 1$, the monomial $(y - 1)$ must also be a factor in the expression for p_1.

Hence the canonical basis function p_1 must have the structure

$$p_1(x,y) = \alpha(x - 1)(y - 1), \tag{2.154}$$

where the constant α is determined as usual in such a way that $p_1(M_1) = 1$. We thus find that the coefficient α is equal to 1 and the function p_1 in the canonical basis is given by

$$p_1(x,y) = (x-1)(y-1). \tag{2.155}$$

By analogous reasoning, the three other canonical basis functions of Q_1 are given by

$$p_2(x,y) = x(1-y), \qquad p_3(x,y) = xy, \qquad p_4(x,y) = y(1-x). \tag{2.156}$$

- **Q_2 Finite-Element Analysis**

1. The space $P(G) \equiv Q_2$ is the set of polynomials of degree less than or equal to two in each of the variables x and y. So every function p in Q_2 can be written in the form

$$p(x,y) = ax^2y^2 + bx^2y + cxy^2 + dx^2 + ey^2 + fxy + gx + hy + i, \tag{2.157}$$

where $(a, b, c, d, e, f, g, h, i)$ runs over $\mathbb{R}^9$. The definition (2.157) implies that Q_2 has dimension 9.

2. To define the nine linear forms σ_i, we introduce five further discretization nodes M_5, M_6, M_7, M_8, and M_9, where the first four of these lie in the middle of each side of the square G, and M_9 lies in the middle of the square, as shown in Fig. 2.7.

   The nine linear forms $\sigma_i, i = 1, \ldots, 9$ are then defined by

$$\sigma_i : p \longrightarrow p(M_i), \quad \forall\, i = 1, \ldots, 9. \tag{2.158}$$

3. We apply exactly the same logic as for the Q_1 quadrilateral finite-element analysis, obtaining the following nine canonical basis functions p_i, $(i = 1, \ldots, 9)$:

$$\begin{aligned}
p_1(x,y) &= (1-x)(1-2x)(1-y)(1-2y),\\
p_2(x,y) &= x(2x-1)(1-y)(1-2y),\\
p_3(x,y) &= xy(2x-1)(2y-1),\\
p_4(x,y) &= y(1-x)(1-2x)(2y-1),\\
p_5(x,y) &= 4x(1-x)(1-y)(1-2y),\\
p_6(x,y) &= 4xy(2x-1)(1-y),\\
p_7(x,y) &= 4xy(1-x)(2y-1),\\
p_8(x,y) &= 4y(1-x)(1-2x)(1-y),\\
p_9(x,y) &= 16xy(1-x)(1-y).
\end{aligned} \tag{2.159}$$

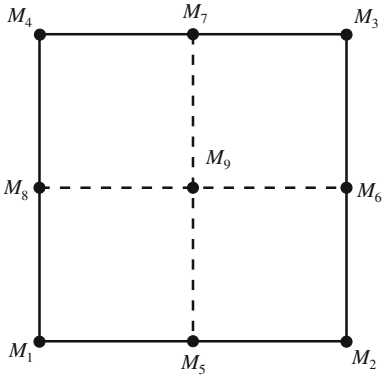

Fig. 2.7 Square mesh element for the Q_2 finite-element analysis

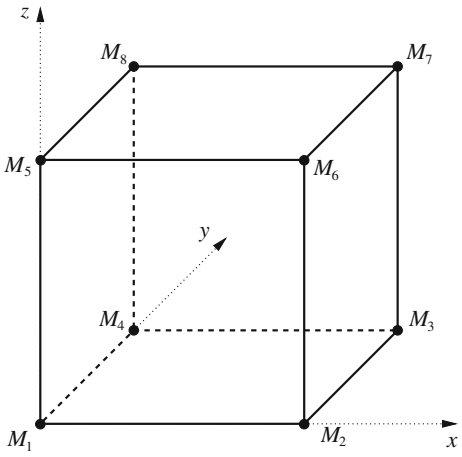

Fig. 2.8 Cubic mesh element for finite-element analysis in 3-dimensional space

2.8.3 Finite-Element Methods in Three Spatial Dimensions

Cubic Meshes

In this last section, we discuss a finite-element analysis in which the mesh element G is a cube $[0, 1] \times [0, 1] \times [0, 1]$ with vertices M_i, $(i = 1, \ldots, 8)$, in the space $(O; x, y, z)$ (see Fig. 2.8).

1. The space Q_1 is defined as the set of polynomials of degree less than or equal to 1 in each of the variables x, y, and z. So every function p in Q_1 can be expressed in the form

$$p(x,y,z) = axyz + bxy + cxz + dyz + ex + fy + gz + h, \tag{2.160}$$

where (a, b, c, d, e, f, g, h) runs over $\mathbb{R}^8$. By simple inspection of the definition (2.160), we see that Q_1 has dimension 8.

2. We introduce the eight linear forms defined by

$$\sigma_i : p \longrightarrow p(M_i), \quad \forall\, i = 1, \ldots, 8. \tag{2.161}$$

3. As usual, the eight canonical basis functions p_i, $(i = 1, \ldots, 8)$, have to satisfy the relations $\sigma_j(p_i) \equiv p_i(M_j) = \delta_{ij}$. To construct these, we identify the monomial factors in the expressions for each.

   Consider, for example, the polynomial p_1 characterizing the node M_1, where it takes the value unity, and which is zero at the seven other nodes M_i, $(i = 2, \ldots, 8)$.

   - The monomial $(1 - x)$ must be a factor in the expression for p_1 so that it vanishes at the nodes M_2, M_3, M_6, and M_7.
   - The monomial $(1 - y)$ must be a factor in the expression for p_1 so that it vanishes at the nodes M_3, M_4, M_7, and M_8.
   - The monomial $(1 - z)$ must be a factor in the expression for p_1 so that it vanishes at the nodes M_5, M_6, M_7, and M_8.

   The basic function p_1 thus has the structure

$$p_1(x,y,z) = \alpha(1 - x)(1 - y)(1 - z), \tag{2.162}$$

   where the constant α is adjusted so that p_1 takes the value unity at its characteristic node M_1.

   Now at the node M_1, $x = y = z = 0$, which implies that α is equal to 1 and the function p_1 has the final form

$$p(x,y,z) = (1 - x)(1 - y)(1 - z). \tag{2.163}$$

   By analogous reasoning, we obtain each polynomial in the canonical basis of Q_1 :

$$\begin{aligned}
&p_1(x, y, z) = (1 - x)(1 - y)(1 - z), && p_2(x, y, z) = x(1 - y)(1 - z),\\
&p_3(x, y, z) = xy(1 - z), && p_4(x, y, z) = (1 - x)y(1 - z),\\
&p_5(x, y, z) = (1 - x)(1 - y)z, && p_6(x, y, z) = x(1 - y)z,\\
&p_7(x, y, z) = xyz, && p_8(x, y, z) = (1 - x)yz.
\end{aligned} \tag{2.164}$$

References

1. G. Duvaut, *Mécanique des milieux continus* (Dunod, Paris, 1998)
2. H. Brézis, *Analyse fonctionnelle, théorie et applications* (Masson, Paris, 1983)
3. R. Dautray, J.-L. Lions, *Analyse mathématique et calcul numérique pour les sciences et les techniques* (Masson, Paris, 1987)
4. M. Moussaoui, in *Singularities and Constructive Methods for Their Treatment*, ed. by P. Grisvard, W. Wendland, J.R. Whiteman. Sur l'approximation des solutions du problème de Dirichlet dans un ouvert avec coins. Lecture Notes in Mathematics, vol. 1121 (Springer, Berlin, 1984), p. 136
5. D. Euvrard, *Résolution des équations aux dérivées partielles de la physique, de la mécanique et des sciences de l'ingénieur* (Masson, Paris, 1994)

Springer

http://www.springer.com/978-3-319-03562-8

Mathematical and Numerical Methods for Partial Differential Equations
Applications for Engineering Sciences
Chaskalovic, J.
2014, XIV, 358 p. 38 illus., Hardcover
ISBN: 978-3-319-03562-8