Chapter 2
Elliptic Boundary-Value Problems

In the first part of this chapter we focus on the question of well-posedness of boundary-value problems for linear partial differential equations of elliptic type. The second part is devoted to the construction and the error analysis of finite difference schemes for these problems. It will be assumed throughout that the coefficients in the equation, the boundary data and the resulting solution are real-valued functions.

2.1 Existence and Uniqueness of Solutions

Suppose that Ω is a bounded open set in $\mathbb{R}^n$, k is a positive integer and $a_{\alpha\beta}$, $0 \le |\alpha|, |\beta| \le k$, with $\alpha, \beta \in \mathbb{N}^n$, are real-valued-functions defined on Ω. We consider the linear partial differential operator $P(x, \partial)$ of order $2k$ defined by

$$P(x,\partial)u := \sum_{0\le|\alpha|,|\beta|\le k} (-1)^{|\alpha|}\partial^\alpha\big(a_{\alpha\beta}(x)\partial^\beta u\big), \quad x \in \Omega. \tag{2.1}$$

The *principal part* $P_0(x, \partial)$ of the differential operator $P(x, \partial)$ is defined by

$$P_0(x,\partial)u := \sum_{|\alpha|,|\beta|=k} (-1)^{|\alpha|}\partial^\alpha\big(a_{\alpha\beta}(x)\partial^\beta u\big), \quad x \in \Omega.$$

$P(x, \partial)$ is said to be an *elliptic operator* on Ω if, and only if,

$$\sum_{|\alpha|,|\beta|=k} a_{\alpha\beta}(x)\xi^\alpha\xi^\beta > 0 \quad \forall x \in \Omega,\ \forall \xi \in \mathbb{R}^n \setminus \{0\}.$$

$P(x, \partial)$ is called *uniformly elliptic* on Ω if, and only if, there exists a positive real number $\tilde{c}$ such that

$$\sum_{|\alpha|,|\beta|=k} a_{\alpha\beta}(x)\xi^\alpha\xi^\beta \ge \tilde{c}|\xi|^{2k} \quad \forall x \in \Omega,\ \forall \xi \in \mathbb{R}^n. \tag{2.2}$$

B.S. Jovanović, E. Süli, *Analysis of Finite Difference Schemes*,
Springer Series in Computational Mathematics 46,
DOI 10.1007/978-1-4471-5460-0_2,

Example 2.1 Consider the second-order partial differential operator, corresponding to $k = 1$ above, defined by

$$P(x, \partial)u := -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_j}\left(a_{ij}(x)\frac{\partial u}{\partial x_i}\right) + \sum_{i=1}^{n}\left[-\frac{\partial}{\partial x_i}\big(a_i(x)u\big) + b_i(x)\frac{\partial u}{\partial x_i}\right] + c(x)u, \tag{2.3}$$

with a_{ij}, $i, j = 1, \dots, n$; a_i, b_i, $i = 1, \dots, n$; and c being real-valued functions defined on an open set $\Omega \subset \mathbb{R}^n$, and such that

$$\sum_{i,j=1}^{n} a_{ij}(x)\xi_i\xi_j \ge \tilde{c}\sum_{i=1}^{n}\xi_i^2 \quad \forall x \in \Omega,\ \forall \xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n, \tag{2.4}$$

for a positive real number $\tilde{c}$, independent of x and ξ; then $P(x, \partial)$ is a second-order uniformly elliptic operator on Ω.

Example 2.2 Consider the partial differential operator $P(x, \partial)$, defined by

$$P(x, \partial)u := \partial_1^2 M_1(u) + 2\partial_1\partial_2 M_3(u) + \partial_2^2 M_2(u),$$

where $\partial_i := \partial/\partial x_i$ and $\partial_i^2 := \partial^2/\partial x_i^2$ for $i = 1, 2$,

$$\begin{aligned} M_1(u) &:= a_1(x)\partial_1^2 u + a_0(x)\partial_2^2 u, \\ M_2(u) &:= a_0(x)\partial_1^2 u + a_2(x)\partial_2^2 u, \\ M_3(u) &:= a_3(x)\partial_1\partial_2 u, \end{aligned}$$

and a_i, $i = 0, 1, 2, 3$, are four real-valued functions defined on a bounded open set $\Omega \subset \mathbb{R}^2$ such that there exist positive real numbers c_1 and c_2 for which

$$a_i(x) \ge c_1, \quad i = 1, 2, 3, \qquad a_1(x)a_2(x) - a_0^2(x) \ge c_2 \quad \forall x \in \Omega.$$

Under these hypotheses $P(x, \partial)$ is a fourth-order uniformly elliptic operator on Ω. The same is true if the above inequalities satisfied by the coefficients a_i are replaced by

$$a_i(x) \ge c_1, \quad i = 1, 2, \qquad a_1(x)a_2(x) - \big(a_0(x) + a_3(x)\big)^2 \ge c_2 \quad \forall x \in \Omega.$$

A partial differential equation on Ω is usually supplemented with *boundary conditions* on $\partial\Omega$. The differential equation in tandem with the boundary conditions imposed forms a *boundary-value problem.*

Example 2.3 For the second-order partial differential equation considered in Example 2.1 the following boundary conditions are the most common, with g denoting a given real-valued function defined on the boundary $\partial\Omega$ in each case:

❶ Dirichlet boundary condition: $u = g$ on $\partial\Omega$;
❷ Oblique derivative boundary condition:

$$\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial u}{\partial x_i}\nu_j + \sum_{i=1}^{n} a_i(x)u\nu_i + \sigma(x)u = g \quad \text{on } \partial\Omega,$$

where ν_j is the jth component of the unit outward normal vector ν to $\partial\Omega$ and σ is a given real-valued function defined on $\partial\Omega$ such that

$$\sigma + \frac{1}{2}\sum_{i=1}^{n}(a_i + b_i)\nu_i \geq 0 \quad \text{on } \partial\Omega.$$

The differential operator

$$u \mapsto \sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial u}{\partial x_i}\nu_j + \sum_{i=1}^{n} a_i(x)u\nu_i, \quad x \in \partial\Omega,$$

is called the *co-normal derivative* corresponding to the partial differential operator from Example 2.1. A particularly important special case arises when $a_{ij} = \delta_{ij}$, $i, j = 1, \ldots, n$, and $a_i = 0$, $i = 1, \ldots, n$. Then, the oblique derivative boundary condition becomes:

$$\partial_\nu u + \sigma u = g \quad \text{on } \partial\Omega,$$

and is referred to as *Robin boundary condition*. Here,

$$\partial_\nu = \frac{\partial}{\partial\nu} := \sum_{i=1}^{n}\nu_i\frac{\partial}{\partial x_i}$$

denotes the (outward) normal derivative on $\partial\Omega$; it is assumed that

$$\sigma + \frac{1}{2}\sum_{i=1}^{n} b_i\nu_i \geq 0 \quad \text{on } \partial\Omega.$$

In particular, when $\sigma = 0$ on $\partial\Omega$, the resulting boundary condition

$$\partial_\nu u = g \quad \text{on } \partial\Omega$$

is called a *Neumann boundary condition*.

In many problems that arise in applications boundary conditions of different kind are enforced on different parts of the boundary; for example, $\partial\Omega$ may be the union of two disjoint subsets $\partial\Omega_1$ and $\partial\Omega_2$, with Dirichlet boundary condition imposed on $\partial\Omega_1$ and an oblique derivative boundary condition imposed on $\partial\Omega_2$. In most of what follows we shall, for simplicity, confine ourselves to the study of elliptic boundary-value problems subject to homogeneous Dirichlet boundary conditions (corresponding, in the case of a second-order elliptic equation, to $g \equiv 0$ in Example 2.3, part ❶).

Returning to the general elliptic equation of order $2k$, we formulate the classical homogeneous Dirichlet boundary-value problem.

Definition 2.1 Let $\Omega \subset \mathbb{R}^n$ be a bounded open set and suppose that $f \in C(\Omega)$ and $a_{\alpha\beta} \in C^{|\alpha|}(\Omega)$, $|\alpha|, |\beta| \leq k$. A function

$$u \in C^{2k}(\Omega) \cap C^{k-1}(\overline{\Omega})$$

is a *classical solution* of the homogeneous Dirichlet problem if

$$P(x, \partial)u := \sum_{0 \leq |\alpha|, |\beta| \leq k} (-1)^{|\alpha|} \partial^\alpha \big(a_{\alpha\beta}(x) \partial^\beta u\big) = f(x)$$

for every x in Ω, and

$$\partial_\nu^m u = 0 \quad \text{on } \partial\Omega, \text{ for } 0 \leq m \leq k-1.$$

It is assumed here that the differential operator $P(x, \partial)$, with $x \in \Omega$, is elliptic or uniformly elliptic on Ω. Frequently, the smoothness requirements on the data stated in this definition are not satisfied. As is demonstrated by the next example, in such instances the corresponding homogeneous Dirichlet boundary-value problem has no classical solution.

Example 2.4 Let $\Omega = (-1, 1)^n \subset \mathbb{R}^n$ and consider *Poisson's equation*

$$-\Delta u := -\sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2} = f \quad \text{in } \Omega,$$

subject to the homogeneous Dirichlet boundary condition

$$u = 0 \quad \text{on } \partial\Omega.$$

Suppose further that $f(x) = \operatorname{sgn}(\frac{1}{2} - |x|)$, $x \in \Omega$.

Clearly, this problem has no classical solution, $u \in C^2(\Omega) \cap C(\overline{\Omega})$, for otherwise Δu would be a continuous function on Ω, which is impossible as $\operatorname{sgn}(\frac{1}{2} - |x|)$ is not continuous on Ω.

In order to overcome the limitations of Definition 2.1 highlighted by this example, we generalize the notion of classical solution by weakening the differentiability requirements on both the data and the corresponding solution.

Definition 2.2 Let $\Omega \subset \mathbb{R}^n$ be a bounded open set and suppose that $f \in L_2(\Omega)$ and $a_{\alpha\beta} \in M\big(W_2^{2k-|\beta|}(\Omega) \to W_2^{|\alpha|}(\Omega)\big)$, $|\alpha|, |\beta| \le k$. A function

$$u \in W_2^{2k}(\Omega) \cap \mathring{W}_2^k(\Omega)$$

is a *strong solution* of the homogeneous Dirichlet problem if

$$P(x, \partial)u := \sum_{0 \le |\alpha|, |\beta| \le k} (-1)^{|\alpha|} \partial^\alpha \big(a_{\alpha\beta}(x)\partial^\beta u\big) = f(x)$$

for almost every x in Ω.

While for classical solutions both the partial differential equation and the boundary condition are assumed to hold in the pointwise sense, for strong solutions the partial differential equation is to be understood in terms of equivalence classes consisting of functions that are equal almost everywhere on Ω; also, instead of being imposed explicitly, the boundary condition has been incorporated into the function space $W_2^{2k}(\Omega) \cap \mathring{W}_2^k(\Omega)$ in which a solution is sought. Unfortunately, it is not easy to show that the homogeneous Dirichlet problem for the partial differential equation (2.1) possesses a strong solution; in fact, as is illustrated by Example 2.5 below a strong solution will not exist unless $\partial\Omega$ and the data are sufficiently smooth. Thus we shall further relax the differentiability requirements on u and weaken the concept of solution by converting the boundary-value problem into a variational problem. The first step in this process is to create a bilinear functional associated with the differential operator $P(x, \partial)$ using integration by parts. Suppose that $u \in W_2^{2k}(\Omega)$, $f \in L_2(\Omega)$, and $v \in C_0^\infty(\Omega)$; then

$$\begin{aligned}\int_\Omega v(x) f(x)\,\mathrm{d}x &= \int_\Omega v P(x, \partial)u\,\mathrm{d}x \\ &= \sum_{0 \le |\alpha|, |\beta| \le k} (-1)^{|\alpha|} \int_\Omega v \partial^\alpha \big(a_{\alpha\beta}(x)\partial^\beta u\big)\,\mathrm{d}x \\ &= \sum_{0 \le |\alpha|, |\beta| \le k} \int_\Omega a_{\alpha\beta}(x)\partial^\beta u \partial^\alpha v\,\mathrm{d}x.\end{aligned}$$

In the transition to the last expression, by partial integration, we made use of the fact that $\operatorname{supp} v \subset\subset \Omega$. Motivated by this identity we introduce the following notation:

$$\begin{aligned}a(u, v) &:= \sum_{0 \le |\alpha|, |\beta| \le k} \int_\Omega a_{\alpha\beta}(x)\partial^\beta u \partial^\alpha v\,\mathrm{d}x, \\ (f, v) &:= \int_\Omega f(x)v(x)\,\mathrm{d}x.\end{aligned}$$

Clearly $a(\cdot,\cdot)$ is correctly defined for u that is merely in $\mathring{W}_2^k(\Omega)$ and for v in the same space; in fact, $a(\cdot,\cdot)$ is a bilinear functional on the product space $\mathring{W}_2^k(\Omega) \times \mathring{W}_2^k(\Omega)$; similarly, $v \mapsto (f, v)$ is a linear functional on $\mathring{W}_2^k(\Omega)$.

These considerations motivate the following definition.

Definition 2.3 Let $\Omega \subset \mathbb{R}^n$ be a bounded open set and suppose that $f \in W_2^{-k}(\Omega)$ and $a_{\alpha\beta} \in L_\infty(\Omega)$, $|\alpha|, |\beta| \le k$. A function

$$u \in \mathring{W}_2^k(\Omega)$$

is a *weak solution* of the homogeneous Dirichlet problem if

$$a(u, v) = \langle f, v\rangle$$

for every $v \in \mathring{W}_2^k(\Omega)$, where now $\langle\cdot,\cdot\rangle$ denotes the duality pairing between $W_2^{-k}(\Omega)$ and $\mathring{W}_2^k(\Omega)$, i.e. $\langle f, v\rangle$ signifies the value of the linear functional $f \in W_2^{-k}(\Omega) = [\mathring{W}_2^k(\Omega)]'$ at $v \in \mathring{W}_2^k(\Omega)$.

Remark 2.1 By applying the Sobolev embedding theorem, it is easily seen that the bilinear functional $a(\cdot,\cdot)$ is well defined under even weaker regularity hypotheses on the coefficients $a_{\alpha\beta}$. Indeed, it suffices to assume in Definition 2.3 that

$$a_{\alpha\beta} \in M\big(W_2^{k-|\alpha|} \to L_{p_\beta}(\Omega)\big), \quad |\alpha|, |\beta| \le k,$$

where $p_\beta = 2$ when $|\beta| = k$, $p_\beta = 2n/(n + 2(k - |\beta|))$ when $0 < k - |\beta| < n/2$; $p_\beta > 1$ (but arbitrarily close to 1) when $k - |\beta| = n/2$; and $p_\beta = 1$ when $k - |\beta| > n/2$.

Next we show that the homogeneous Dirichlet boundary-value problem has a unique weak solution. The proof is based on a simple application of the Lax–Milgram theorem (Theorem 1.13) and the following result.

Theorem 2.4 (Gårding's Inequality) *Suppose that $\Omega \subset \mathbb{R}^n$ is a Lipschitz domain. Let $P(x, \partial)$ be a linear partial differential operator of order $2k$ of the form (2.1) such that, for some $\tilde{c} > 0$, the uniform ellipticity condition (2.2) holds. Suppose also that*

$$a_{\alpha\beta} \in C(\Omega) \quad \textit{for } |\alpha| = |\beta| = k$$

and

$$a_{\alpha\beta} \in L_\infty(\Omega) \quad \textit{for } |\alpha|, |\beta| \le k.$$

Then, there exist constants $c_0 > 0$ and $\lambda_0 \ge 0$ such that

$$a(v, v) + \lambda_0 \|v\|^2_{L_2(\Omega)} \ge c_0 \|v\|^2_{W_2^k(\Omega)} \quad \textit{for all } v \in \mathring{W}_2^k(\Omega). \tag{2.5}$$

The proof of this results is long and technical, and will not be presented here; the interested reader is referred to Theorem 9.17 on p. 292 of Renardy and Rogers [155], for example.

For second-order uniformly elliptic operators of the form (2.3) the proof of Gårding's inequality is much simpler, and we shall confine ourselves to this case; in fact, as will be seen below, in the case of a second-order uniformly elliptic operator the smoothness hypotheses on the coefficients in the principal part of the operator can be slightly relaxed: they need not be continuous functions, as long as they belong to $L_\infty(\Omega)$. We note that the bilinear functional corresponding to the operator (2.3) is given by

$$a(u,v) = \sum_{i,j=1}^{n} \int_\Omega a_{ij}(x) \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} \,\mathrm{d}x + \sum_{i=1}^{n} a_i(x) u \frac{\partial v}{\partial x_i} \,\mathrm{d}x$$
$$+ \int_\Omega b_i(x) \frac{\partial u}{\partial x_i} v \,\mathrm{d}x + \int_\Omega c(x) u v \,\mathrm{d}x, \quad u, v \in \mathring{W}_2^1(\Omega).$$

Theorem 2.5 *Suppose that $\Omega \subset \mathbb{R}^n$ is a Lipschitz domain. Let $P(x,\partial)$ be the second-order linear partial differential operator defined by (2.3) where a_{ij}, a_i, $b_j \in L_\infty(\Omega)$, $i, j = 1, \ldots, n$, and $c \in L_\infty(\Omega)$ are such that, for some $\tilde{c} > 0$, the uniform ellipticity condition (2.4) holds. Then, there exist real numbers $c_0 > 0$ and $\lambda_0 \geq 0$ such that*

$$a(v,v) + \lambda_0 \|v\|^2_{L_2(\Omega)} \geq c_0 \|v\|^2_{W_2^1(\Omega)} \quad \forall v \in \mathring{W}_2^1(\Omega).$$

Proof Thanks to (2.4) and the Cauchy–Schwarz inequality we have that

$$a(v,v) = \sum_{i,j=1}^{n} \int_\Omega a_{ij}(x) \frac{\partial v}{\partial x_i} \frac{\partial v}{\partial x_j} \,\mathrm{d}x + \sum_{i=1}^{n} \int_\Omega a_i(x) v \frac{\partial v}{\partial x_i} \,\mathrm{d}x$$
$$+ \sum_{i=1}^{n} \int_\Omega b_i(x) \frac{\partial v}{\partial x_i} v \,\mathrm{d}x + \int_\Omega c(x) v^2 \,\mathrm{d}x$$
$$\geq \tilde{c} \int_\Omega |\nabla v|^2 \,\mathrm{d}x - \int_\Omega \left[2 \sum_{i=1}^{n} \left(a_i^2 + b_i^2\right) \right]^{1/2} |\nabla v||v| \,\mathrm{d}x$$
$$- \|c\|_{L_\infty(\Omega)} \int_\Omega |v|^2 \,\mathrm{d}x,$$

where, as usual $|\nabla u| = \left[(\frac{\partial u}{\partial x_1})^2 + \cdots + (\frac{\partial u}{\partial x_n})^2 \right]^{1/2}$. By applying the elementary inequality

$$ab \leq \varepsilon a^2 + \frac{1}{4\varepsilon} b^2$$

with $\varepsilon = \tilde{c}/2$, we obtain

$$a(v, v) \geq \frac{\tilde{c}}{2} \int_{\Omega} |\nabla v|^2 \, \mathrm{d}x - C\|v\|^2_{L_2(\Omega)},$$

where

$$C = \frac{1}{\tilde{c}} \left\| \sum_{i=1}^{n} \left(a_i^2 + b_i^2\right) \right\|_{L_\infty(\Omega)} + \|c\|_{L_\infty(\Omega)}.$$

Equivalently,

$$a(v, v) \geq \frac{\tilde{c}}{2} \|v\|^2_{W_2^1(\Omega)} - \left(C + \frac{\tilde{c}}{2}\right) \|v\|^2_{L_2(\Omega)},$$

which proves Gårding's inequality with $c_0 = \tilde{c}/2$ and $\lambda_0 = C + (\tilde{c}/2)$. □

Remark 2.2 We note that Theorem 2.5 can be proved under even weaker hypotheses on a_{ij}, a_i and b_i. Indeed, it suffices to assume that

$$\begin{aligned} a_{ij} &\in M\big(L_2(\Omega) \to L_2(\Omega)\big), \quad i, j = 1, \ldots, n, \\ a_i, b_i &\in M\big(W_2^1(\Omega) \to L_2(\Omega)\big), \quad i = 1, \ldots, n, \\ c &\in M\big(W_2^1(\Omega) \to L_p(\Omega)\big), \end{aligned}$$

where $p = 2n/(n+2)$ if $n > 2$; $p > 1$ (but arbitrarily close to 1) if $n = 2$; and $p = 1$ if $n = 1$.

We now state the main result of this section, which concerns the existence of a weak solution to a homogeneous Dirichlet boundary-value problem.

Theorem 2.6 *Let $P(x, \partial)$ be a linear partial differential operator of order $2k$ of the form (2.1), satisfying the conditions of Theorem 2.4 on a Lipschitz domain $\Omega \subset \mathbb{R}^n$. Then, there exists a $\lambda_0 \geq 0$ such that, for any $\lambda \geq \lambda_0$ and any $f \in W_2^{-k}(\Omega)$, the homogeneous Dirichlet problem for the operator*

$$\tilde{P}(x, \partial) = P(x, \partial) + \lambda$$

has a unique weak solution $u \in \mathring{W}_2^k(\Omega)$. Furthermore, this solution satisfies

$$\|u\|_{W_2^k(\Omega)} \leq C\|f\|_{W_2^{-k}(\Omega)}.$$

Proof According to Theorem 2.4 there exists a $\lambda_0 \geq 0$ such that the Gårding inequality (2.5) holds. For $\lambda \geq \lambda_0$ we consider the bilinear functional

$$\tilde{a}(u, v) = a(u, v) + \lambda(u, v), \quad u, v \in \mathring{W}_2^k(\Omega),$$

associated with the operator $\tilde{P}$. We shall prove that $\tilde{a}(\cdot,\cdot)$ satisfies the conditions of the Lax–Milgram theorem (Theorem 1.13) on $\mathring{W}_2^k(\Omega) \times \mathring{W}_2^k(\Omega)$. Let us take $\mathcal{U} = \mathring{W}_2^k(\Omega)$ in Theorem 1.13 and recall that $\mathring{W}_2^k(\Omega)$ is a real Hilbert space. The $\mathcal{U}$-coercivity of $\tilde{a}(\cdot,\cdot)$ is a straightforward consequence of (2.5):

$$\tilde{a}(v,v) = a(v,v) + \lambda\|v\|^2_{L_2(\Omega)} \geq c_0\|v\|^2_{W_2^k(\Omega)} \quad \forall v \in \mathring{W}_2^k(\Omega).$$

We shall now verify that $\tilde{a}(\cdot,\cdot)$ is bounded on $\mathring{W}_2^k(\Omega) \times \mathring{W}_2^k(\Omega)$. Given $v, w \in \mathring{W}_2^k(\Omega)$, using the Cauchy–Schwarz inequality repeatedly we obtain the following chain of inequalities, which ultimately lead to the conclusion that $\tilde{a}(\cdot,\cdot)$ is a bounded bilinear functional on $\mathring{W}_2^k(\Omega) \times \mathring{W}_2^k(\Omega)$:

$$\begin{aligned}
\big|\tilde{a}(v,w)\big| &\leq \big|a(v,w)\big| + \lambda\big|(v,w)\big| \\
&\leq \sum_{0\leq|\alpha|,|\beta|\leq k} \int_\Omega \big|a_{\alpha\beta}(x)\big|\big|\partial^\beta v\big|\big|\partial^\alpha w\big|\,\mathrm{d}x + \lambda\big|(v,w)\big| \\
&\leq \max_{0\leq|\alpha|,|\beta|\leq k} \|a_{\alpha\beta}\|_{L_\infty(\Omega)} \sum_{0\leq|\alpha|,|\beta|\leq k} \int_\Omega \big|\partial^\beta v\big|\big|\partial^\alpha w\big|\,\mathrm{d}x + \lambda\big|(v,w)\big| \\
&\leq c_1\|v\|_{\mathcal{U}}\|w\|_{\mathcal{U}}.
\end{aligned}$$

Thus, by the Lax–Milgram theorem (Theorem 1.13), for each $f \in W_2^{-k}(\Omega) = \mathcal{U}'$, there exists a unique weak solution $u \in \mathring{W}_2^k(\Omega)$ to the homogeneous Dirichlet problem. □

In the case of second-order elliptic equations we have an analogous result.

Theorem 2.7 *Let $P(x,\partial)$ be a linear second-order partial differential operator of the form (2.3), satisfying the conditions of Theorem 2.5 on a Lipschitz domain $\Omega \subset \mathbb{R}^n$. Then, there exists a $\lambda_0 \geq 0$ such that, for any $\lambda \geq \lambda_0$ and any $f \in W_2^{-1}(\Omega)$, the homogeneous Dirichlet problem for the operator*

$$\tilde{P}(x,\partial) = P(x,\partial) + \lambda$$

has a unique weak solution $u \in \mathring{W}_2^1(\Omega)$, and this solution satisfies

$$\|u\|_{W_2^1(\Omega)} \leq C\|f\|_{W_2^{-1}(\Omega)}.$$

Furthermore, if $a_i, b_i \in W_p^1(\Omega)$, $i = 1,\ldots,n$, where $p = n/2$ when $n > 2$; $p > 1$ is arbitrary when $n = 2$; and $p = 1$ when $n = 1$; and

$$c(x) - \frac{1}{2}\sum_{i=1}^n \frac{\partial}{\partial x_i}\big(a_i(x) + b_i(x)\big) \geq 0$$

for almost every $x \in \Omega$, then $\lambda_0 = 0$. In other words, the homogeneous Dirichlet problem corresponding to the operator $P(x, \partial)$ has a unique weak solution $u \in \mathring{W}_2^1(\Omega)$ under these hypotheses.

Proof The first part of the theorem is proved in exactly the same way as the corresponding statement in Theorem 2.6. In order to prove the second part let us observe that, by the divergence theorem,

$$\int_\Omega \big[a_i(x) + b_i(x)\big] \frac{\partial v}{\partial x_i} v \,\mathrm{d}x = -\frac{1}{2} \int_\Omega \frac{\partial}{\partial x_i} \big(a_i(x) + b_i(x)\big) v^2 \,\mathrm{d}x \quad \forall v \in \mathring{W}_2^1(\Omega);$$

we note that because $a_i, b_i \in W_p^1(\Omega)$, $i = 1, \ldots, n$, where p is as assumed, Hölder's inequality, followed by the application of Sobolev's embedding theorem, implies that the function appearing as the integrand on the right-hand side is an element of $L_1(\Omega)$. Therefore the right-hand side of this equality is meaningful.

Consequently,

$$a(v, v) \geq \tilde{c} \sum_{i=1}^n \int_\Omega \left| \frac{\partial v}{\partial x_i} \right|^2 \mathrm{d}x. \tag{2.6}$$

By applying the Friedrichs inequality (1.23) with $s = 1$ and $p = 2$, the right-hand side of (2.6) can be further bounded below to obtain

$$a(v, v) \geq c_0 \|v\|^2_{W_2^1(\Omega)}, \tag{2.7}$$

where $c_0 = \tilde{c}/c_\star$, and hence the $\mathring{W}_2^1(\Omega)$-coercivity of the bilinear functional $a(\cdot, \cdot)$. The boundedness of $a(\cdot, \cdot)$ on the space $\mathring{W}_2^1(\Omega) \times \mathring{W}_2^1(\Omega)$ follows from the boundedness of $\tilde{a}(\cdot, \cdot) = a(\cdot, \cdot) + \lambda(\cdot, \cdot)$ by setting $\lambda = 0$. The required result is now obtained from the Lax–Milgram theorem (Theorem 1.13). □

Remark 2.3 We note that Theorem 2.7 continues to hold when the regularity hypotheses of Theorem 2.5 are replaced by the weaker ones from Remark 2.2.

Having developed relatively straightforward sufficient conditions for the existence of a unique weak solution to an elliptic boundary-value problem, the question that we now need to address is whether a weak solution might possess additional regularity to qualify as a strong solution. The answer to this question very much depends on additional regularity of the data (i.e. the coefficients, the right-hand side of the partial differential equation, and the boundary $\partial\Omega$). Since a general discussion of regularity properties of weak solutions to elliptic boundary-value problems is beyond the scope of this book, we shall confine ourselves to Poisson's equation subject to a homogeneous Dirichlet boundary condition, which is sufficiently illustrative of the key ideas. We begin with a simple example, which shows that a weak solution to an elliptic boundary-value problem need not be a strong solution to the problem, and that a strong solution may not even exist.

Example 2.5 Suppose that $\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < \mathrm{e}^{-1}\}$ and let $f(x, y) := -\Delta(\log|\log(x^2 + y^2)|)$, with $\log := \log_{\mathrm{e}}$ and the differential operator Δ understood in the sense of distributions on Ω. It is easily seen by changing from Cartesian co-ordinates to polar co-ordinates that the function $u : (x, y) \mapsto \log|\log(x^2 + y^2)|$ belongs to $\mathring{W}_2^1(\Omega)$ and that, therefore, $f \in W_2^{-1}(\Omega)$. Thus, u is the unique weak solution to the boundary-value problem: $-\Delta u = f$ on Ω (with the equality understood as being between two elements of $W_2^{-1}(\Omega)$), subject to the boundary condition $u = 0$ on $\partial\Omega$. However, the function u is *not* a strong solution and, as a matter of fact, the boundary-value problem has no strong solution, since $f \notin L_2(\Omega)$.

In fact, even if f belongs to $W_2^{s-2}(\Omega)$, $s \geq 2$, it does not automatically follow that the weak solution to Poisson's equation $-\Delta u = f$, with a homogeneous Dirichlet boundary condition on $\partial\Omega$, belongs to $W_2^s(\Omega) \cap \mathring{W}_2^1(\Omega)$. Whether or not this is the case depends on the smoothness of $\partial\Omega$. In particular if Ω is a bounded polygonal domain in $\mathbb{R}^2$, the regularity of the solution is ultimately limited by the size of the maximum internal angle of Ω; the next theorem is a special case of a more general result, due to Grisvard [61].

Theorem 2.8 *Suppose that $f \in W_2^{s-2}(\Omega)$, $1 \leq s < 3$, $s \neq 3/2, 5/2$, with $\Omega = (0, 1)^2$, and consider the homogeneous Dirichlet boundary-value problem for Poisson's equation*:

$$\begin{aligned} -\Delta u &= f \quad \text{on } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

Then, the unique weak solution u in $\mathring{W}_2^1(\Omega)$ belongs to $W_2^s(\Omega) \cap \mathring{W}_2^1(\Omega)$.

The limitation $s < 3$ on the Sobolev exponent in Theorem 2.8 is sharp in the sense that the stated regularity result is invalid for $s \geq 3$ unless f satisfies certain *compatibility conditions* at the four corners of the square. More precisely, u belongs to the space $W_2^s(\Omega) \cap \mathring{W}_2^1(\Omega)$ for $s \in \mathbb{N}$, $s \geq 3$, provided that $f \in W_2^{s-2}(\Omega)$ and the following conditions hold *at the four corners*:

$$\begin{aligned} &f = 0, \\ &\partial_1^2 f - \partial_2^2 f = 0, \\ &\dots\dots\dots\dots\dots\dots\dots \\ &\partial_1^{2k} f - \partial_1^{2k-2}\partial_2^2 f + \cdots + (-1)^k \partial_2^{2k} f = 0, \quad \text{with } k = \left[\frac{s-2}{2}\right]. \end{aligned} \tag{2.8}$$

The proof proceeds similarly to the one in Volkov [193], where an analogous regularity result was shown for classical solutions. For details we refer to the work of Hell [70].

Next we formulate a result that concerns the existence of weak solutions to the homogeneous Dirichlet problem for the fourth-order uniformly elliptic equation considered in Example 2.2.

Theorem 2.9 *Let $\Omega \subset \mathbb{R}^2$ be a Lipschitz domain. Consider the partial differential operator $P(x, \partial)$, defined by*

$$P(x,\partial)u := \partial_1^2 M_1(u) + 2\partial_1\partial_2 M_3(u) + \partial_2^2 M_2(u),$$

where

$$\begin{aligned} M_1(u) &:= a_1(x)\partial_1^2(u) + a_0(x)\partial_2^2 u, \\ M_2(u) &:= a_0(x)\partial_1^2 u + a_2(x)\partial_2^2 u, \\ M_3(u) &:= a_3(x)\partial_1\partial_2 u, \end{aligned}$$

and $a_i \in L_\infty(\Omega)$, $i = 0, 1, 2, 3$, are such that there exist positive constants c_1 and c_2 for which

$$a_i(x) \geq c_1, \quad i = 1, 2, 3, \qquad a_1(x)a_2(x) - a_0^2(x) \geq c_2, \quad x \in \Omega.$$

Then, for any $f \in W_2^{-2}(\Omega)$, the homogeneous Dirichlet boundary-value problem for $P(x, \partial)$ has a unique weak solution u in $\mathring{W}_2^2(\Omega)$.

Proof The proof is, again, based on the Lax–Milgram theorem (Theorem 1.13); its nontrivial part is to verify that the bilinear functional

$$a(u,v) = \big(M_1(u), \partial_1^2 v\big) + 2\big(M_3(u), \partial_1\partial_2 v\big) + \big(M_2(u), \partial_2^2 v\big), \quad u, v \in \mathring{W}_2^2(\Omega),$$

is $\mathring{W}_2^2(\Omega)$-coercive. Clearly,

$$\begin{aligned} a(v,v) = \int_\Omega \Big[a_1(x)\big|\partial_1^2 v\big|^2 + 2a_3(x)|\partial_1\partial_2 v|^2 \\ + a_2(x)\big|\partial_2^2 v\big|^2 + 2a_0(x)\partial_1^2 v\,\partial_2^2 v\Big]\,\mathrm{d}x \quad \forall v \in \mathring{W}_2^2(\Omega). \end{aligned}$$

As v is real-valued (by the convention stated at the beginning of the chapter), we have the following identity:

$$\begin{aligned} a(v,v) = \frac{1}{2}\int_\Omega a_1(x)\left(\partial_1^2 v + \frac{a_0(x)}{a_1(x)}\partial_2^2 v\right)^2 \mathrm{d}x \\ + \frac{1}{2}\int_\Omega a_2(x)\left(\partial_2^2 v + \frac{a_0(x)}{a_2(x)}\partial_1^2 v\right)^2 \mathrm{d}x \\ + \frac{1}{2}\int_\Omega \left(a_1(x) - \frac{a_0^2(x)}{a_2(x)}\right)\big|\partial_1^2 v\big|^2\,\mathrm{d}x \\ + \frac{1}{2}\int_\Omega \left(a_2(x) - \frac{a_0^2(x)}{a_1(x)}\right)\big|\partial_2^2 v\big|^2\,\mathrm{d}x \end{aligned}$$

$$+ 2\int_{\Omega} a_3(x)|\partial_1\partial_2 v|^2 \, \mathrm{d}x \quad \forall v \in \mathring{W}_2^2(\Omega).$$

Therefore,

$$\begin{aligned} a(v,v) \geq \frac{1}{2}\int_{\Omega}\left(a_1(x) - \frac{a_0^2(x)}{a_2(x)}\right)\left|\partial_1^2 v\right|^2 \mathrm{d}x \\ + \frac{1}{2}\int_{\Omega}\left(a_2(x) - \frac{a_0^2(x)}{a_1(x)}\right)\left|\partial_2^2 v\right|^2 \mathrm{d}x \\ + 2\int_{\Omega} a_3(x)|\partial_1\partial_2 v|^2 \, \mathrm{d}x \quad \forall v \in \mathring{W}_2^2(\Omega). \end{aligned}$$

By noting the assumptions on the coefficients a_i, $i = 0, 1, 2, 3$, it follows that there exists a positive constant $\tilde{c}$ such that

$$a(v,v) \geq \tilde{c}|v|^2_{W_2^2(\Omega)} \quad \forall v \in \mathring{W}_2^2(\Omega).$$

Finally, by the Friedrichs inequality (1.23) with $s = p = n = 2$,

$$\|v\|^2_{W_2^2(\Omega)} \leq c_\star |v|^2_{W_2^2(\Omega)} \quad \forall v \in \mathring{W}_2^2(\Omega),$$

and hence

$$a(v,v) \geq c_0\|v\|^2_{W_2^2(\Omega)} \quad \forall v \in \mathring{W}_2^2(\Omega),$$

where $c_0 = \tilde{c}/c_\star$. □

Remark 2.4 Suppose that the homogeneous Dirichlet boundary condition

$$\partial_\nu^m u = 0 \quad \text{on } \partial\Omega \ \text{ for } \ m = 0, 1,$$

for the partial differential operator $P(x, \partial)$ defined in Theorem 2.9 has been replaced by the following set of boundary conditions:

$$u = 0, \quad M_1(u)\nu_1 + M_3(u)\nu_2 = 0, \qquad M_3(u)\nu_1 + M_2(u)\nu_2 = 0 \quad \text{on } \partial\Omega.$$

The weak formulation of the corresponding boundary-value problem is: find $u \in W_2^2(\Omega) \cap \mathring{W}_2^1(\Omega)$ such that

$$a(u,v) = \langle f, v\rangle$$

for every $v \in W_2^2(\Omega) \cap \mathring{W}_2^1(\Omega)$. Again, by using the Lax–Milgram theorem (Theorem 1.13), it is easy to prove that, under the same conditions on a_i, $i = 0, 1, 2, 3$, as in Theorem 2.9, this problem too has a unique weak solution, now in the function space $W_2^2(\Omega) \cap \mathring{W}_2^1(\Omega)$.

Finally, we return to the boundary-value problem considered in Example 2.4, which has been shown to have no classical solution. By applying Theorem 2.7 with $a_{ij}(x) \equiv 1$, $i = j$, $a_{ij}(x) \equiv 0$, $i \neq j$, $1 \leq i, j \leq n$, $b_i(x) \equiv 0$, $c(x) \equiv 0$, $f(x) = \operatorname{sgn}(\frac{1}{2} - |x|)$, and $\Omega = (-1, 1)^n$, we see that there is a unique weak solution $u \in \mathring{W}_2^1(\Omega)$ to this problem. In fact, it can be shown that this weak solution belongs to $W_2^2(\Omega) \cap \mathring{W}_2^1(\Omega)$ and it is, therefore, a strong solution to the boundary-value problem (see Grisvard [62, 63]).

Remark 2.5 The existence and uniqueness of a weak solution to a Neumann, Robin, or oblique derivative boundary-value problem for a second-order uniformly elliptic equation can be established in a similar fashion, using the Lax–Milgram theorem (Theorem 1.13).

Remark 2.6 Theorems 2.6 and 2.7 imply that the weak formulation of the Dirichlet boundary-value problem for the operator $\tilde{P}(x, \partial) = P(x, \partial) + \lambda$, $\lambda \geq \lambda_0 \geq 0$, is *well-posed in the sense of Hadamard*; that is, for each $f \in W_2^{-k}(\Omega)$, there exists a unique (weak) solution $u \in \mathring{W}_2^k(\Omega)$; moreover, "small" changes in f give rise to "small" changes in the corresponding solution u. The latter property follows by noting that if u_1 and u_2 are weak solutions in $\mathring{W}_2^k(\Omega)$ of the homogeneous Dirichlet problem for $\tilde{P}(x, \partial)$ corresponding to right-hand sides f_1 and f_2 in $W_2^{-k}(\Omega)$, respectively, then $u_1 - u_2$ is the unique weak solution in $\mathring{W}_2^k(\Omega)$ of the homogeneous Dirichlet boundary-value problem for the operator $\tilde{P}(x, \partial)$ corresponding to the right-hand side $f_1 - f_2$ in $W_2^{-k}(\Omega)$. It thus follows from Theorems 2.6 and 2.7 that

$$\|u_1 - u_2\|_{W_2^k(\Omega)} \leq C\|f_1 - f_2\|_{W_2^{-k}(\Omega)},$$

where C is a positive constant, independent of u_1, u_2, f_1 and f_2; this implies the continuous dependence of the solution to the homogeneous Dirichlet boundary-value problem on the right-hand side of the equation.

2.2 Approximation of Elliptic Problems

We begin this section by outlining the general approach to the construction of finite difference schemes for elliptic boundary-value problems; we then introduce basic results from the theory of finite difference schemes and present some classical tools for the error analysis of finite difference schemes for partial differential equations with smooth solutions. The limitations of the classical theory will lead us to consider finite difference schemes with mollified data, and we shall develop a theoretical framework for the error analysis of such nonstandard schemes. We conclude by considering finite difference approximations of second- and fourth-order elliptic equations with variable coefficients, and derive sharp error bounds in various mesh-dependent (discrete) norms, under minimal smoothness requirements on the data and the associated solution.

2.2.1 Introduction to the Theory of Finite Difference Schemes

Assuming that Ω is a bounded open set in $\mathbb{R}^n$, we consider a boundary-value problem on Ω of the general form

$$\mathcal{L}u = f \qquad \text{in } \Omega, \tag{2.9}$$

$$lu = g \qquad \text{on } \Gamma = \partial\Omega, \tag{2.10}$$

where $\mathcal{L}$ is a linear partial differential operator, and l is a linear operator that specifies the boundary condition. For example, we may have

$$\mathcal{L}u := -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_j}\left(a_{ij}(x)\frac{\partial u}{\partial x_i}\right) + \sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i} + c(x)u,$$

where the $a_{ij}(x)$, $i, j = 1, \ldots, n$, satisfy (2.4), with one of the following choices of the boundary operator l (Dirichlet, Neumann or oblique derivative):

$$lu := u,$$

or

$$lu := \frac{\partial u}{\partial \nu},$$

or

$$lu := \sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial u}{\partial x_i}\nu_j + \sigma(x)u,$$

where ν is the unit outward normal vector to Γ, ν_j is the jth component of ν, $j = 1, \ldots, n$, and σ is a bounded, nonnegative function defined on Γ.

The construction of a finite difference scheme for the boundary-value problem (2.9), (2.10) consists of two basic steps: first, the domain $\overline{\Omega}$ is replaced by a finite set of points, called the *mesh* or *grid*, and second, the derivatives in the differential equation and in the boundary condition are replaced by divided differences. To describe the first of these two steps more precisely, suppose that we have approximated $\overline{\Omega} = \Omega \cup \Gamma$ by the mesh

$$\overline{\Omega}^h := \Omega^h \cup \Gamma^h,$$

where $\Omega^h \subset \Omega$ is the set of *interior mesh-points*, and $\Gamma^h \subset \Gamma$ is the set of *boundary mesh-points*. Typically the mesh consist of a finite set of points obtained by considering the intersections of n families of parallel hyperplanes, each element of each family being perpendicular to one of the co-ordinate axes. If the domain Ω is not axiparallel, adjustments may need to be made to the mesh near the boundary $\partial\Omega$, which may be curved. The parameter $h = (h_1, \ldots, h_n)$ measures the spacing of the mesh; in particular, $h_i > 0$ denotes the mesh-size in the ith co-ordinate direction. Once the mesh has been constructed, we proceed by replacing the derivatives featuring in $\mathcal{L}$ by divided differences, and approximate the boundary condition in a similar fashion. This yields a *finite difference scheme* of the form

$$\mathcal{L}_h U(x) = f_h(x), \quad x \in \Omega^h, \tag{2.11}$$

$$l_h U(x) = g_h(x), \quad x \in \Gamma^h, \tag{2.12}$$

where $\mathcal{L}_h$ and l_h are linear difference operators, representing discrete counterparts of $\mathcal{L}$ and l, while f_h and g_h are suitable approximations of f and g, respectively. In algebraic terms, (2.11), (2.12) is a system of linear equations involving the values of the approximate solution U at the mesh-points.

Assuming that (2.11), (2.12) has a unique solution U, when the mesh spacing is small the sequence of values of the approximate solution at the mesh-points, $\{U(x) : x \in \overline{\Omega}^h\}$, is expected to resemble $\{u(x) : x \in \overline{\Omega}^h\}$, the set of values of the exact solution u at the mesh-points. However the closeness of $U(x)$ to $u(x)$ at $x \in \overline{\Omega}^h$ is by no means obvious, and the proof of such approximation results represents the central theme of this book. We shall consider a range of problems of the form (2.9), (2.10), and derive sharp bounds on the error between the analytical solution u (typically a weak solution) and its finite difference approximation U in terms of positive powers of the discretization parameter h. Bounds of this kind imply, in particular, that the error between the analytical solution u and its finite difference approximation U converges to zero with a certain rate, in a certain norm, as $h \to 0$.

2.2.2 Finite Difference Approximation in One Space Dimension

In this section we shall focus on the finite difference approximation of a two-point boundary-value problem. We begin by developing some basic results about mesh-functions (i.e. functions that are defined on the finite difference mesh), finite difference operators and mesh-dependent (discrete) norms.

2.2.2.1 Meshes, Mesh-Functions and Mesh-Dependent Norms

Meshes Suppose that N is a positive integer, $N \geq 2$, let $h := 1/N$, and consider the *uniform mesh* on the unit interval $(0, 1)$ of the real line, defined by

$$\Omega^h := \{x_i : x_i = ih, i = 1, \ldots, N-1\}.$$

We further define

$$\overline{\Omega}^h := \Omega^h \cup \{0, 1\}.$$

Let S^h denote the linear space of real-valued functions defined on the mesh $\overline{\Omega}^h$, and let S_0^h be the linear space of all real-valued functions defined on the mesh $\overline{\Omega}^h$ that are equal to zero on $\Gamma^h := \overline{\Omega}^h \setminus \Omega^h$. Any element of the set S^h (or of S_0^h) will be referred to as a *mesh-function*.

For a mesh-function $V \in S^h$ we define $V_i := V(x_i) = V(ih)$. We equip the linear space S_0^h with the inner product

$$(V, W)_h = (V, W)_{L_2(\Omega^h)} := \sum_{x \in \Omega^h} hV(x)W(x) = \sum_{i=1}^{N-1} hV_i W_i, \tag{2.13}$$

which closely resembles the inner product

$$(v, w) = \int_0^1 v(x)w(x)\,\mathrm{d}x,$$

of the Hilbert space $L_2(\Omega)$. The inner product $(\cdot,\cdot)_h$ induces the norm $\|\cdot\|_h$ on S_0^h defined by

$$\|V\|_h = \|V\|_{L_2(\Omega^h)} := (V, V)_h^{1/2}. \tag{2.14}$$

Analogously, we equip the linear space S^h with the inner product

$$[V, W]_h = (V, W)_{L_2(\overline{\Omega}^h)} := \frac{h}{2}\big[V(0)W(0) + V(1)W(1)\big] + (V, W)_h$$

and the induced norm

$$[\![V]\!]_h = \|V\|_{L_2(\overline{\Omega}^h)} := [V, V]_h^{1/2}.$$

We shall also need the meshes

$$\Omega_-^h := \Omega^h \cup \{0\}, \qquad \Omega_+^h := \Omega^h \cup \{1\}.$$

On the linear space of real-valued functions defined on the mesh Ω_-^h, we consider the inner product

$$[V, W)_h = (V, W)_{L_2(\Omega_-^h)} := \sum_{x \in \Omega_-^h} hV(x)W(x) = \sum_{i=0}^{N-1} hV_i W_i$$

and the associated norm

$$[\![V\|_h = \|V\|_{L_2(\Omega_-^h)} := [V, V)_h^{1/2},$$

with an analogous definition of the inner product $(V, W]_h = (V, W)_{L_2(\Omega_+^h)}$ and the corresponding norm $\|V]\!]_h = \|V\|_{L_2(\Omega_+^h)}$ on the linear space of real-valued mesh-functions defined on Ω_+^h.

Finite Difference Operators The forward, backward and central divided difference operators D_x^+, D_x^- and D_x^0 on the mesh $\overline{\Omega}_h$ are defined, respectively, by

$$D_x^+ V := \frac{V^+ - V}{h}, \qquad D_x^- V := \frac{V - V^-}{h}, \qquad D_x^0 V := \frac{1}{2}\big(D_x^+ V + D_x^- V\big),$$

where we have used the notation

$$V^\pm = V^\pm(x) := V(x \pm h).$$

With these definitions, we have the following discrete Leibniz formulae:

$$D_x^+(VW) = \big(D_x^+ V\big)W^+ + V\big(D_x^+ W\big) = \big(D_x^+ V\big)W + V^+\big(D_x^+ W\big),$$
$$D_x^-(VW) = \big(D_x^- V\big)W^- + V\big(D_x^- W\big) = \big(D_x^- V\big)W + V^-\big(D_x^- W\big),$$

and the summation-by-parts formula:

$$\big[D_x^+ V, W)_h = -(V, D_x^- W\big]_h + V(1)W(1) - V(0)W(0), \tag{2.15}$$

which immediately yields the following result.

Lemma 2.10 *Suppose that $V \in S_0^h$; then,*

$$\big(-D_x^+ D_x^- V, V\big)_h = \sum_{i=1}^{N} h\big|D_x^- V_i\big|^2 = \sum_{i=0}^{N-1} h\big|D_x^+ V_i\big|^2. \tag{2.16}$$

Proof Let us write $U_i = D_x^- V_i$, $i = 1, \dots, N$, and note that

$$\big(-D_x^+ D_x^- V, V\big)_h = -\big(D_x^+ U, V\big)_h = -\big[D_x^+ U, V\big)_h = \big(U, D_x^- V\big]_h = \big\| D_x^- V\big]\big|_h^2,$$

thanks to our assumption that $V \in S_0^h$, which implies that $V_0 = V(0) = 0$ and $V_N = V(1) = 0$, and using the identity (2.15). The second equality in (2.16) follows simply by noting that $D_x^- V_i = D_x^+ V_{i-1}$, $i = 1, \dots, N$, and shifting the index i in the summation. □

The Discrete Laplace Operator on S_0^h On the set S_0^h, we define the linear operator $\Lambda : S_0^h \to S_0^h$ by

$$(\Lambda V)(x) := \begin{cases} -D_x^+ D_x^- V(x) & \text{if } x \in \Omega^h, \\ 0 & \text{if } x \in \Gamma^h = \overline{\Omega}^h \setminus \Omega^h. \end{cases}$$

Since

$$\begin{aligned}(\Lambda V, W)_h &= -\big(D_x^+ D_x^- V, W\big)_h = \big(D_x^- V, D_x^- W\big]_h = \big(D_x^- W, D_x^- V\big]_h \\ &= \big[D_x^+ V, D_x^+ W\big)_h = \big[D_x^+ W, D_x^+ V\big)_h = (\Lambda W, V)_h,\end{aligned}$$

Λ is a symmetric linear operator on S_0^h. Moreover, thanks to (2.16),

$$(\Lambda V, V) = \| D_x^- V]|_h^2 = [| D_x^+ V \|_h^2 > 0 \quad \text{for all } V \in S_0^h \setminus \{0\},$$

and therefore Λ is positive definite on S_0^h. Thus Λ has $N-1$ distinct positive eigenvalues, which are easily shown to be (see Samarskiĭ [159], Sect. 2.4.2)

$$\lambda_k = \frac{4}{h^2} \sin^2 \frac{k\pi h}{2}, \quad k = 1, 2, \ldots, N-1; \tag{2.17}$$

these eigenvalues satisfy the inequalities

$$8 < \lambda_k < \frac{4}{h^2}, \quad k = 1, 2, \ldots, N-1. \tag{2.18}$$

The corresponding $N-1$ eigenfunctions V^k, $k = 1, \ldots, N-1$, satisfying $\Lambda V^k = \lambda_k V^k$, are

$$V^k(x) = \sin k\pi x, \quad x \in \overline{\Omega}^h, \ k = 1, 2, \ldots, N-1.$$

The set of eigenfunctions $\{V^1, \ldots, V^{N-1}\}$ is an orthogonal system in S_0^h with respect to the inner product $(\cdot, \cdot)_h$; that is,

$$\left(V^k, V^l\right)_h = \frac{1}{2}\delta_{kl}, \quad k, l = 1, 2, \ldots, N-1, \tag{2.19}$$

where δ_{kl} is the Kronecker delta; in fact, $\{V^1, \ldots, V^{N-1}\}$ forms a basis of the linear space S_0^h. Consequently an arbitrary mesh-function $V \in S_0^h$ can be expressed as a linear combination of these eigenfunctions:

$$V(x) = \sum_{k=1}^{N-1} b_k \sin k\pi x, \quad x \in \overline{\Omega}^h, \tag{2.20}$$

where

$$b_k = 2\left(V, V^k\right)_h.$$

By noting the orthogonality of the eigenfunctions we deduce the following *discrete Parseval identity*:

$$\|V\|_h^2 = \frac{1}{2} \sum_{k=1}^{N-1} b_k^2. \tag{2.21}$$

Analogously,

$$\| D_x^- V]|_h^2 = [| D_x^+ V \|_h^2 = (\Lambda V, V)_h = \frac{1}{2} \sum_{k=1}^{N-1} \lambda_k b_k^2, \tag{2.22}$$

$$\left\| D_x^+ D_x^- V \right\|_h^2 = (\Lambda V, \Lambda V)_h = \frac{1}{2} \sum_{k=1}^{N-1} \lambda_k^2 b_k^2. \tag{2.23}$$

It follows from (2.18) and (2.21)–(2.23) that

$$\left\| D_x^+ D_x^- V \right\|_h \geq 2\sqrt{2} \left\| D_x^- V \right]\!\!]_h = 2\sqrt{2} [\![D_x^+ V \right\|_h \geq 8 \| V \|_h \tag{2.24}$$

for each $V \in S_0^h$.

Discrete Sobolev Norms on S_0^h The discrete analogues of Sobolev seminorms and norms are defined similarly to their 'continuous' counterparts introduced in Chap. 1. In particular, we define

$$\begin{aligned} |V|_{1,h} = |V|_{W_2^1(\Omega^h)} &:= \left\| D_x^- V \right]\!\!]_h = [\![D_x^+ V \right\|_h, \\ |V|_{2,h} = |V|_{W_2^2(\Omega^h)} &:= \left\| D_x^+ D_x^- V \right\|_h, \\ \|V\|_{k,h} = \|V\|_{W_2^k(\Omega^h)} &:= \left(\|V\|^2_{W_2^{k-1}(\Omega^h)} + |V|^2_{W_2^k(\Omega^h)} \right)^{1/2}, \end{aligned} \tag{2.25}$$

where $k = 1, 2$, with the convention that $W_2^0(\Omega^h) = L_2(\Omega^h)$. The inequalities (2.24) imply that the seminorms $|\cdot|_{W_2^1(\Omega^h)}$ and $|\cdot|_{W_2^2(\Omega^h)}$ are equivalent to the norms $\|\cdot\|_{W_2^1(\Omega^h)}$ and $\|\cdot\|_{W_2^2(\Omega^h)}$, respectively, on S_0^h.

Lemma 2.11 (Discrete Friedrichs Inequality) *There exists a positive constant $c_\star$ such that*

$$\|V\|^2_{W_2^1(\Omega^h)} \leq c_\star \left\| D_x^- V \right\|^2_{L_2(\Omega_+^h)} \tag{2.26}$$

for all $V \in S_0^h$.

Proof The last inequality in (2.24) implies (2.26) with $c_\star = 9/8$. □

Lemma 2.12 (Discrete Sobolev Embedding) *For all $V \in S_0^h$ the following inequality holds*

$$\|V\|_{\infty,h} := \max_{x \in \overline{\Omega}^h} |V(x)| \leq \frac{1}{2} \left\| D_x^- V \right\|_{L_2(\Omega_+^h)} \tag{2.27}$$

Proof Using the Cauchy–Schwarz inequality, we obtain from the identity

$$|V_i|^2 = (1 - ih)|V_i|^2 + ih|V_i|^2 = (1 - ih) \left| \sum_{j=1}^{i} \left(D_x^- V_j \right) h \right|^2 + ih \left| \sum_{j=i+1}^{N} \left(D_x^- V_j \right) h \right|^2$$

that

$$
\begin{aligned}
|V_i|^2 &\le (1-ih)\left(\sum_{j=1}^{i} h\right)\sum_{j=1}^{i}\left(D_x^- V_j\right)^2 h + ih\left(\sum_{j=i+1}^{N} h\right)\sum_{j=i+1}^{N}\left(D_x^- V_j\right)^2 h \\
&= ih(1-ih)\sum_{j=1}^{N}\left(D_x^- V_j\right)^2 h.
\end{aligned}
$$

The required inequality then follows by taking the maximum over the index $i \in \{0, 1, \dots, N\}$ and noting that, for all such i, $0 \le ih(1-ih) \le 1/4$. □

The Discrete Laplace Operator on S^h We define the linear operator $\overline{\Lambda} : S^h \to S^h$ by

$$
(\overline{\Lambda} V)(x) := \begin{cases} -\frac{2}{h} D_x^+ V(0) & \text{if } x = 0, \\ -D_x^+ D_x^- V(x) & \text{if } x \in \Omega^h, \\ \frac{2}{h} D_x^- V(1) & \text{if } x = 1. \end{cases}
$$

Assuming that each $V \in S^h$ is extended outside $\overline{\Omega}^h$ as an even function, we have that

$$
(\overline{\Lambda} V)(x) = \left(-D_x^+ D_x^- V\right)(x) \quad \text{for } x \in \overline{\Omega}^h.
$$

The linear operator $\overline{\Lambda}$ is symmetric with respect to the inner product $[\cdot, \cdot]_h$. The eigenvalues of $\overline{\Lambda}$ are given by the formula (2.17), but now for $k = 0, 1, 2, \dots, N$. In fact, since $\lambda_0 = 0$ is an eigenvalue, $\overline{\Lambda} : S^h \to S^h$ is only nonnegative (positive semidefinite) rather than positive definite; that is,

$$
[\overline{\Lambda} V, V]_h \ge 0 \quad \text{for all } V \in S^h \setminus \{0\}.
$$

The eigenfunctions of $\overline{\Lambda}$ corresponding to the eigenvalues λ_k, $k = 0, \dots, N$, are:

$$
W^0(x) = 1, \qquad W^k(x) = \cos k\pi x, \quad k = 1, 2, \dots, N;
$$

these form an orthogonal system in the sense that

$$
\left[W^k, W^l\right]_h = \begin{cases} 1 & \text{if } k = l = 0, N, \\ \frac{1}{2} & \text{if } k = l = 1, 2, \dots, N-1, \\ 0 & \text{if } k \ne l, \end{cases}
$$

and they span the linear space S^h; hence each mesh-function $V \in S^h$ can be expressed as

$$
V(x) = \frac{1}{2}a_0 + \sum_{k=1}^{N-1} a_k \cos k\pi x + \frac{1}{2} a_N \cos N\pi x, \tag{2.28}
$$

where

$$a_k = 2[V, \cos k\pi x]_h, \quad \text{for } k = 0, 1, \ldots, N.$$

When $V \in S_0^h$, the expansions (2.20) and (2.28) coincide at all points of the mesh $\overline{\Omega}^h$.

By noting the orthogonality of the eigenfunctions W^k, $k = 0, \ldots, N$, it is easily seen that for any mesh-function V contained in S^h the following identities hold:

$$[\![V]\!]_h^2 = \frac{1}{4}a_0^2 + \frac{1}{2}\sum_{k=1}^{N-1} a_k^2 + \frac{1}{4}a_N^2,$$

$$\big\| D_x^- V\big]\!\big]_h^2 = [\overline{\Lambda} V, V]_h = \frac{1}{2}\sum_{k=1}^{N-1} \lambda_k a_k^2 + \frac{1}{4}\lambda_N a_N^2,$$

$$[\![\overline{\Lambda} V]\!]_h^2 = \frac{1}{2}\sum_{k=1}^{N-1} \lambda_k^2 a_k^2 + \frac{1}{4}\lambda_N^2 a_N^2.$$

Next, we introduce analogous discrete Sobolev norms on the linear space S^h, consisting of all real-valued functions defined on the mesh $\overline{\Omega}^h$.

Discrete Sobolev Norms on S^h Similarly as on S_0^h, we introduce on S^h the following discrete analogues of the Sobolev norms $\|\cdot\|_{W_2^k(\Omega)}$, $k = 1, 2$:

$$[\![V]\!]_{1,h} = \|V\|_{W_2^1(\overline{\Omega}^h)} := \left([\![V]\!]_h^2 + \big\|D_x^- V\big]\!\big]_h^2\right)^{1/2},$$

$$[\![V]\!]_{2,h} = \|V\|_{W_2^2(\overline{\Omega}^h)} := \left([\![V]\!]_h^2 + \big\|D_x^- V\big]\!\big]_h^2 + [\![\overline{\Lambda} V]\!]_h^2\right)^{1/2}.$$

Fractional-Order Discrete Sobolev Norms Next we shall define fractional-order Sobolev norms on S_0^h and derive an interpolation inequality that relates these to the integer-order discrete Sobolev norms defined earlier. We shall limit ourselves to the case when the Sobolev index r is in the range $(0, 1) \cup (1, 2)$. We define the seminorm $|\cdot|_{W_2^r(\Omega^h)}$ by

$$|V|_{W_2^r(\Omega^h)} := \begin{cases} \left(h^2 \sum_{x,y\in\overline{\Omega}^h, x\neq y} \frac{[V(x)-V(y)]^2}{|x-y|^{1+2r}}\right)^{1/2} & \text{if } 0 < r < 1, \\ \left(h^2 \sum_{x,y\in\Omega_-^h, x\neq y} \frac{[D_x^+ V(x)-D_x^+ V(y)]^2}{|x-y|^{1+2r}}\right)^{1/2} & \text{if } 1 < r < 2, \end{cases}$$

and we introduce the corresponding fractional-order discrete Sobolev norm

$$\|V\|_{W_2^r(\Omega^h)} := \left(\|V\|^2_{W_2^{[r]}(\Omega^h)} + |V|^2_{W_2^r(\Omega^h)}\right)^{1/2}, \quad 0 < r < 2,\ r \neq 1.$$

Higher order fractional-order discrete Sobolev norms can be defined similarly.

Next we state an interpolation inequality that establishes a relationship between fractional-order discrete Sobolev norms and the integer-order norms defined earlier.

Lemma 2.13 *Suppose that* $r \in (0, 1)$. *Then, there exists a positive real number* $C(r)$ *such that, for each mesh-function* $V \in S_0^h$,

$$\|V\|_{W_2^r(\Omega^h)} \le C(r)\|V\|_{L_2(\Omega^h)}^{1-r}\|V\|_{W_2^1(\Omega^h)}^r, \quad 0 < r < 1.$$

Proof Given a mesh-function $V \in S_0^h$, we decompose it as a finite linear combination of sine functions, as in (2.20), and define the norm $B_r(\cdot)$ on S_0^h in terms of the corresponding expansion coefficients b_k, $k = 1, \dots, N-1$, by

$$B_r(V) := \left(\frac{1}{2}\sum_{k=1}^{N-1} k^{2r} b_k^2\right)^{1/2}.$$

It is left to the reader to verify that $B_r(\cdot)$ is indeed a norm on S_0^h. By noting (2.17), the elementary inequality

$$\sin x \ge \frac{2}{\pi}x, \quad 0 \le x \le \pi/2,$$

and Hölder's inequality with exponents $p := 1/(1-r)$ and $p' := 1/r$ we obtain

$$\begin{aligned} B_r(V) &\le \left[\frac{1}{2}\sum_{k=1}^{N-1}\left(\frac{\lambda_k}{4}\right)^r b_k^2\right]^{1/2} = 2^{-r}\left[\frac{1}{2}\sum_{k=1}^{N-1} b_k^{2(1-r)}\left(\lambda_k b_k^2\right)^r\right]^{1/2} \\ &\le 2^{-r}\left(\frac{1}{2}\sum_{k=1}^{N-1} b_k^2\right)^{(1-r)/2}\left(\frac{1}{2}\sum_{k=1}^{N-1}\lambda_k b_k^2\right)^{r/2}, \end{aligned}$$

and hence, by the discrete Parseval identities (2.21) and (2.22),

$$B_r(V) \le 2^{-r}\|V\|_{L_2(\Omega^h)}^{1-r}|V|_{W_2^1(\Omega^h)}^r. \tag{2.29}$$

The rest of the proof is devoted to showing that the norm $B_r(\cdot)$ is equivalent to $\|\cdot\|_{W_2^r(\Omega^h)}$. For this purpose, we extend the function $V \in S_0^h$ from

$$\overline{\Omega}^h = \{kh : k = 0, \dots, N\}$$

to the mesh

$$\{kh : k = 0, \pm 1, \pm 2, \dots, \pm N\}$$

as an odd function; that is, $V(-x) := -V(x)$ for each x in $\overline{\Omega}^h$. The resulting function is then further extended to the infinite lattice

$$h\mathbb{Z} = \{kh : k = 0, \pm 1, \pm 2, \dots\}$$

as a 2-periodic function; as before, $h := 1/N$ and $N \geq 2$. Let $\omega^h := (-1, 1) \cap h\mathbb{Z}$ and $\overline{\omega}^h := [-1, 1] \cap h\mathbb{Z}$. For mesh-functions V defined on $\overline{\omega}^h$ we consider

$$N_r(V) := \left\{ h^2 \sum_{x\in\overline{\omega}^h}^{\star} \sum_{t\in\overline{\omega}^h, t\neq 0}^{\star} \frac{[V(x) - V(x-t)]^2}{|t|^{1+2r}} \right\}^{1/2},$$

where

$$h \sum_{x\in\overline{\omega}^h}^{\star} W(x) := \frac{1}{2} h\big[W(-1) + W(1)\big] + h \sum_{x\in\omega^h} W(x) = [W, 1]_{L_2(\overline{\omega}^h)}.$$

By noting the periodicity of the extended function (still denoted by V) and the expansion (2.20), we obtain

$$\begin{aligned}
N_r(V)^2 &= h^2 \sum_{x\in\overline{\omega}^h}^{\star} \sum_{t\in\overline{\omega}^h, t\neq 0}^{\star} |t|^{-1-2r} V(x)\big[-V(x-t) + 2V(x) - V(x+t)\big] \\
&= h^2 \sum_{x\in\overline{\omega}^h}^{\star} \sum_{t\in\overline{\omega}^h, t\neq 0}^{\star} |t|^{-1-2r} \sum_{l=1}^{N-1} b_l \sin l\pi x \sum_{k=1}^{N-1} 4b_k \sin^2 \frac{k\pi t}{2} \sin k\pi x \\
&= \sum_{l=1}^{N-1} \sum_{k=1}^{N-1} b_l b_k h \sum_{x\in\overline{\omega}^h}^{\star} \sin l\pi x \sin k\pi x h \sum_{t\in\overline{\omega}^h, t\neq 0}^{\star} |t|^{-1-2r} 4 \sin^2 \frac{k\pi t}{2} \\
&= 8 \sum_{k=1}^{N-1} b_k^2 h \sum_{t\in\Omega_+^h}^{\star\star} t^{-1-2r} \sin^2 \frac{k\pi t}{2}.
\end{aligned}$$

Here we have used the notation

$$h \sum_{t\in\Omega_+^h}^{\star\star} W(t) := h \sum_{t\in\Omega^h} W(t) + \frac{1}{2} h W(1) = (W, 1)_h + \frac{1}{2} h W(1).$$

After further transformation, we obtain

$$N_r(V)^2 = 16 \left(\frac{\pi}{2}\right)^{2r} \frac{1}{2} \sum_{k=1}^{N-1} k^{2r} b_k^2 C(k, r),$$

where

$$C(k, r) := \frac{k\pi h}{2} \sum_{t\in\Omega_+^h}^{\star\star} \left(\frac{k\pi t}{2}\right)^{-1-2r} \sin^2 \frac{k\pi t}{2}.$$

It is easily seen that $C(k,r)$ is the Riemann sum for the integral

$$\int_0^{k\pi/2} x^{-1-2r} \sin^2 x \, \mathrm{d}x$$

and can be therefore bounded from below and above as follows:

$$\frac{1}{8}\left(\frac{2}{\pi}\right)^{2r} \le C(k,r) \le \pi^{2-2r}\left(1+\frac{1}{2-2r}\right)+\frac{1}{2r}\left(\frac{2}{\pi}\right)^{2r}.$$

Thus we deduce that $N_r(\cdot)$ and $B_r(\cdot)$ are equivalent norms on S_0^h.

By noting inequality (2.29), the equivalence of the seminorm $|\cdot|_{W_2^1(\Omega^h)}$ and the norm $\|\cdot\|_{W_2^1(\Omega^h)}$ on the linear space S_0^h, in conjunction with the obvious inequality $|V|_{W_2^r(\Omega^h)} \le N_r(V)$, we then arrive at the desired inequality. That completes the proof. □

Remark 2.7 The lemma can also be proved by using the cosine expansion (2.28) and the norm

$$A_r(V) := \left(\frac{1}{2}\sum_{k=1}^{N-1} k^{2r} a_k^2 + \frac{1}{4} N^{2r} a_N^2\right)^{1/2}.$$

It can be shown that this norm is equivalent to $N_r(\cdot)$, provided that V has been extended periodically outside $\overline{\Omega}^h$ as an even function.

Remark 2.8 A similar argument shows, for $r \in (1,2)$, that there exists a positive real number $C_1(r)$ such that

$$\|V\|_{W_2^r(\Omega^h)} \le C_1(r)\|V\|_{W_2^1(\Omega^h)}^{2-r}\|V\|_{W_2^2(\Omega^h)}^{r-1}, \quad 1<r<2.$$

Remark 2.9 Finally we note that, similarly as on S_0^h, one can define a fractional-order discrete Sobolev norm on S^h as follows:

$$\|V\|^2_{W_2^r(\overline{\Omega}^h)} := \left(\|V\|^2_{W_2^{[r]}(\overline{\Omega}^h)} + |V|^2_{W_2^r(\Omega^h)}\right)^{1/2}, \quad 0<r<2,\ r \ne 1.$$

After this brief summary of notational conventions in one dimension, we consider a simple one-dimensional model problem, construct its finite difference approximation and derive bounds on the error, in the discrete norms defined above, between the analytical solution and its finite difference approximation.

2.2.3 *Finite Difference Scheme for a Univariate Problem*

We give a simple illustration of the general framework of finite difference approximation by considering the following two-point boundary-value problem for a second-order linear (ordinary) differential equation:

$$-u'' + c(x)u = f(x), \quad x \in (0, 1), \tag{2.30}$$

$$u(0) = 0, \qquad u(1) = 0. \tag{2.31}$$

We shall assume that $c \geq 0$ almost everywhere on $(0, 1)$, $c \in L_\infty(0, 1)$ and $f \in W_2^{-1}(0, 1)$.

The first step in the construction of a finite difference scheme for this boundary-value problem is to define the mesh. Let N be an integer, $N \geq 2$, and let $h := 1/N$ be the mesh-size; the mesh-points are $x_i := ih$, $i = 0, \dots, N$. We then define

$$\Omega^h := \{x_i : i = 1, \dots, N-1\},$$
$$\Gamma^h := \{x_0, x_N\} \quad \text{and} \quad \overline{\Omega}^h := \Omega^h \cup \Gamma^h.$$

Let us suppose that the unique weak solution $u \in \mathring{W}_2^1(0, 1)$ to this boundary-value problem is sufficiently smooth (e.g. $u \in C^4([0, 1])$). Then, by Taylor series expansion of u about the mesh-point x_i, $1 \leq i \leq N-1$, we deduce that, as $h \to 0$,

$$\begin{aligned} u(x_{i\pm1}) &= u(x_i \pm h) \\ &= u(x_i) \pm hu'(x_i) + \frac{h^2}{2}u''(x_i) \pm \frac{h^3}{6}u'''(x_i) + \mathcal{O}\big(h^4\big), \end{aligned}$$

so that

$$D_x^+ u(x_i) := \frac{u(x_{i+1}) - u(x_i)}{h} = u'(x_i) + \mathcal{O}(h),$$

$$D_x^- u(x_i) := \frac{u(x_i) - u(x_{i-1})}{h} = u'(x_i) + \mathcal{O}(h),$$

$$D_x^0 u(x_i) := \frac{u(x_{i+1}) - u(x_{i-1})}{2h} = u'(x_i) + \mathcal{O}\big(h^2\big)$$

and

$$\begin{aligned} D_x^+ D_x^- u(x_i) &= D_x^- D_x^+ u(x_i) \\ &= \frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1})}{h^2} \\ &= u''(x_i) + \mathcal{O}\big(h^2\big). \end{aligned}$$

Recall that D_x^+ and D_x^- are called the *forward* and *backward* divided difference operator, respectively, D_x^0 is referred to as the *central-difference operator*, while $D_x^+ D_x^-$ is the (symmetric) *second divided difference* operator. It follows from these Taylor series expansions that, for a sufficiently smooth function u (e.g. for $u \in C^2([0, 1])$), $D_x^+ u(x_i)$ and $D_x^+ u(x_i)$ approximate $u'(x_i)$ to $\mathcal{O}(h)$ for $i = 0, \dots, N-1$

and $i = 1, \dots, N$, respectively, while the central difference approximation $D_x^0 u(x_i)$ is more accurate: it approximates $u'(x_i)$ to $\mathcal{O}(h^2)$ for $i = 1, \dots, N-1$ (provided that $u \in C^3([0, 1])$. Similarly, the *second divided difference* $D_x^+ D_x^- u(x_i)$ is an $\mathcal{O}(h^2)$ approximation to $u''(x_i)$, $i = 1, \dots, N-1$, (as long as $u \in C^4([0, 1])$. Thus we replace the second derivative u'' in (2.30) by the second divided difference to obtain

$$-D_x^+ D_x^- u(x_i) + c(x_i)u(x_i) \approx f(x_i), \quad i = 1, \dots, N-1, \tag{2.32}$$

$$u(x_0) = 0, \qquad u(x_N) = 0. \tag{2.33}$$

Here we have implicitly assumed that both c and f are continuous functions on the interval $(0, 1)$; thus, $c(x_i)$ and $f(x_i)$ are correctly defined for all $i = 1, \dots, N-1$. We shall also suppose that

$$c(x) \geq 0 \quad \forall x \in (0, 1). \tag{2.34}$$

Now (2.32) and (2.33) indicate that we should seek our approximation U to u by solving the system of difference equations:

$$-D_x^+ D_x^- U_i + c(x_i)U_i = f(x_i), \quad i = 1, \dots, N-1, \tag{2.35}$$

$$U_0 = 0, \qquad U_N = 0. \tag{2.36}$$

Using matrix notation, this can be written as

$$\mathcal{A}U = F,$$

where

$$\mathcal{A} := \begin{bmatrix} \frac{2}{h^2} + c(x_1) & -\frac{1}{h^2} & & & 0 \\ -\frac{1}{h^2} & \frac{2}{h^2} + c(x_2) & -\frac{1}{h^2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{h^2} & \frac{2}{h^2} + c(x_{N-2}) & -\frac{1}{h^2} \\ 0 & & & -\frac{1}{h^2} & \frac{2}{h^2} + c(x_{N-1}) \end{bmatrix},$$

$$U := (U_1, U_2, \dots, U_{N-1})^{\mathrm{T}}$$

and

$$F := \bigl(f(x_1), f(x_2), \dots, f(x_{N-1})\bigr)^{\mathrm{T}}.$$

Thus $\mathcal{A}$ is a symmetric tridiagonal $(N-1) \times (N-1)$ matrix, and U and F are column vectors of size $N-1$.

We begin the analysis of the finite difference scheme (2.35), (2.36) by showing that it has a unique solution; this will be achieved by proving that the matrix $\mathcal{A}$ is nonsingular. For this purpose, we introduce the inner product (2.13). Let S_0^h denote the set of all real-valued functions V defined at the mesh-points x_i, $i = 0, \dots, N$, such that $V_0 = V_N = 0$.

We define the linear operator $A : S_0^h \to S_0^h$ by

$$(AV)_i := -D_x^+ D_x^- V_i + c(x_i)V_i, \quad i = 1, \ldots, N-1,$$
$$(AV)_0 = (AV)_N := 0.$$

Returning to the finite difference scheme (2.35), (2.36) and using Lemma 2.10 and (2.34), we see that, for $V \in S_0^h$,

$$\begin{aligned}(AV, V)_h &= \left(-D_x^+ D_x^- V + cV, V\right)_h \\ &= \left(-D_x^+ D_x^- V, V\right)_h + (cV, V)_h \\ &\geq \sum_{i=1}^{N} h\left|D_x^- V_i\right|^2 = \left\|D_x^- V\right\|^2_{L_2(\Omega_+^h)},\end{aligned} \tag{2.37}$$

where the norm $\|\cdot\|_{L_2(\Omega_+^h)}$ has been defined in the previous section. Thus, if $AV = 0$ for some V, then $D_x^- V_i = 0$, $i = 1, \ldots, N$; because $V_0 = V_N = 0$, this implies that $V_i = 0$, $i = 0, \ldots, N$. Hence $AV = 0$ if, and only if, $V = 0$. We deduce that $A : S_0^h \to S_0^h$ is invertible and, consequently, $\mathcal{A}$ is a nonsingular matrix; thus (2.35), (2.36) has a unique solution, $U = \mathcal{A}^{-1}F$. We summarize our findings in the next theorem.

Theorem 2.14 *Suppose that c and f are continuous functions on the interval* $(0, 1)$, *and* $c(x) \geq 0$ *for* $x \in (0, 1)$; *then, the finite difference scheme* (2.35), (2.36) *possesses a unique solution U in S_0^h.*

We note that by Theorem 2.7, for $c \in C([0, 1])$ satisfying (2.34) and $f \in C([0, 1])$, the boundary-value problem (2.30), (2.31) has a unique weak solution $u \in \mathring{W}_2^1(0, 1)$; in fact, by Sobolev's embedding theorem u belongs to $C([0, 1])$ and therefore $u'' = f - cu \in C([0, 1])$. However to derive an error bound between u and its finite difference approximation U we shall have to assume that u is even more regular (the precise regularity hypothesis required in the analysis will be stated below). A key ingredient in our error analysis will be the fact that the scheme (2.35), (2.36) is stable (or discretely well-posed) in the sense that "small" perturbations in the data result in "small" perturbations in the corresponding finite difference solution. Actually, we shall prove the discrete version of the inequality appearing in Remark 2.6. For this purpose, we shall consider the *discrete L_2 norm* (2.14) and the *discrete Sobolev norm* (2.25). From (2.37) and the discrete Friedrichs inequality (2.26) we deduce, with $c_0 = 1/c_\star = 8/9$, that

$$(AV, V)_h \geq c_0 \|V\|^2_{W_2^1(\Omega^h)}. \tag{2.38}$$

Now the stability of the finite difference scheme (2.35), (2.36) easily follows.

Theorem 2.15 *The scheme* (2.35), (2.36) *is stable in the sense that*

$$\|U\|_{W_2^1(\Omega^h)} \le \frac{1}{c_0}\|f\|_{L_2(\Omega^h)}, \tag{2.39}$$

where $c_0 = 8/9$.

Proof From (2.38) and (2.35) we have that

$$\begin{aligned} c_0\|U\|^2_{W_2^1(\Omega^h)} &\le (AU, U)_h = (f, U)_h \\ &\le \|f\|_{L_2(\Omega^h)}\|U\|_{L_2(\Omega^h)} \le \|f\|_{L_2(\Omega^h)}\|U\|_{W_2^1(\Omega^h)}, \end{aligned}$$

and hence we deduce (2.39). □

Theorem 2.15 implies that if U_1 and U_2 are solutions of the problem (2.35), (2.36) corresponding to right-hand sides f_1 and f_2, respectively, then

$$\|U_1 - U_2\|_{W_2^1(\Omega^h)} \le \frac{1}{c_0}\|f_1 - f_2\|_{L_2(\Omega^h)}.$$

Therefore, in analogy with the boundary-value problem (2.30), (2.31), the difference scheme (2.35), (2.36) is well-posed in the sense of Remark 2.6. It is important to note that the 'stability constant' $1/c_0$ is independent of the discretization parameter h: the spacing of the finite difference mesh.

By exploiting this stability result it is easy to derive a bound on the error between the analytical solution u, and its finite difference approximation U. We define the *global error*, e, by

$$e_i := u(x_i) - U_i, \quad i = 0, \dots, N.$$

Obviously $e_0 = 0$, $e_N = 0$, and

$$Ae_i = \varphi_i, \quad i = 1, \dots, N-1, \tag{2.40}$$

where the mesh-function φ, defined by

$$\varphi_i := Au(x_i) - f(x_i), \quad i = 1, \dots, N-1,$$

is called the *truncation error* of the finite difference scheme. A simple calculation using (2.30) reveals that

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i), \quad i = 1, \dots, N-1.$$

Since the global error satisfies (2.40), we can apply (2.39) to deduce that

$$\|u - U\|_{W_2^1(\Omega^h)} = \|e\|_{W_2^1(\Omega^h)} \le \frac{1}{c_0}\|\varphi\|_{L_2(\Omega^h)}. \tag{2.41}$$

It remains to bound $\|\varphi\|_{L_2(\Omega^h)}$.

Assuming now that $u \in C^4([0,1])$, the Taylor series expansions stated at the beginning of this section imply that

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i) = \mathcal{O}(h^2);$$

thus, there exists a positive constant C, independent of h, such that

$$|\varphi_i| \leq Ch^2.$$

Consequently,

$$\|\varphi\|_{L_2(\Omega^h)} = \left(\sum_{i=1}^{N-1} h|\varphi_i|^2\right)^{1/2} \leq Ch^2. \tag{2.42}$$

Combining (2.41) and (2.42), it follows that

$$\|u - U\|_{W_2^1(\Omega^h)} \leq \frac{C}{c_0} h^2. \tag{2.43}$$

In fact, a more careful treatment of the remainder term in the Taylor series expansion of u reveals that, for $i = 1, \ldots, N-1$,

$$\varphi_i = u''(x_i) - D_x^+ D_x^- u(x_i) = -\frac{1}{12} h^2 u''''(\xi_i), \quad \xi_i \in (x_{i-1}, x_{i+1}).$$

Thus

$$|\varphi_i| \leq \frac{1}{12} h^2 \max_{x \in [0,1]} \left|u''''(x)\right|, \tag{2.44}$$

and hence

$$C = \frac{1}{12} \max_{x \in [0,1]} \left|u''''(x)\right|$$

in (2.42). As $c_0 = 1/c_\star$ and $c_\star = 9/8$, we deduce that $c_0 = 8/9$. Substituting the values of the constants C and c_0 into (2.43), it follows that

$$\|u - U\|_{W_2^1(\Omega^h)} \leq \frac{3}{32} h^2 \|u''''\|_{C([0,1])}.$$

Thus we have proved the following result.

Theorem 2.16 *Let* $f \in C([0,1])$, $c \in C([0,1])$, *with* $c(x) \geq 0$ *for all* $x \in [0,1]$, *and suppose that the corresponding solution of the boundary-value problem* (2.30), (2.31) *belongs to* $C^4([0,1])$; *then,*

$$\|u - U\|_{W_2^1(\Omega^h)} \leq \frac{3}{32} h^2 \|u''''\|_{C([0,1])}. \tag{2.45}$$

We note that by the argument following Theorem 2.14 the hypotheses $f \in C([0,1])$, $c \in C([0,1])$, $c \geq 0$ imply that the unique weak solution of the boundary-value problem (2.30), (2.31) belongs to $C^2([0,1])$, and it is therefore a classical solution. Thus, the word *solution* in this theorem means *classical solution*.

It follows from (2.37) with $V = e$, (2.40), the Cauchy–Schwarz inequality, the last inequality in (2.24), (2.27) and (2.44) that

$$\|u - U\|_{\infty,h} \leq \frac{1}{48\sqrt{2}} h^2 \|u''''\|_{C([0,1])}. \tag{2.46}$$

We thus deduce the following result.

Theorem 2.17 *Suppose that the assumptions of Theorem 2.16 are satisfied; then, the error bound (2.46) holds.*

This simple stability and error analysis of the finite difference scheme (2.35), (2.36) already contains the key ingredients of a general error analysis of finite difference approximations, and it is instructive to highlight them here.

(1) The first step is to prove the stability of the scheme in an appropriate mesh-dependent norm (cf. (2.39), for example). A typical stability result for the abstract finite difference scheme (2.11), (2.12) considered at the beginning of the section is of the form

$$c_0 |||U|||_{\Omega^h} \leq \|f_h\|_{\Omega^h} + \|g_h\|_{\Gamma^h}, \tag{2.47}$$

where $|||\cdot|||_{\Omega^h}$, $\|\cdot\|_{\Omega^h}$ and $\|\cdot\|_{\Gamma^h}$ are mesh-dependent norms involving mesh-points of Ω^h (or $\overline{\Omega}^h$) and Γ^h, respectively, and c_0 is a positive constant, independent of h.

(2) The second step is to estimate the size of the *truncation error*,

$$\varphi_{\Omega^h} := L_h u - f_h \quad \text{in } \Omega^h,$$
$$\varphi_{\Gamma^h} := l_h u - g_h \quad \text{on } \Gamma^h.$$

In the case of the finite difference scheme (2.11), (2.12), $\varphi_{\Gamma^h} = 0$, and therefore φ_{Γ^h} did not appear explicitly in our error analysis. If

$$\|\varphi_{\Omega^h}\|_{\Omega^h} + \|\varphi_{\Gamma^h}\|_{\Gamma^h} \to 0 \quad \text{as } h \to 0,$$

for a sufficiently smooth solution u of (2.9), (2.10), we say that the scheme (2.11), (2.12) is *consistent*. If p is the largest positive real number such that

$$\|\varphi_{\Omega^h}\|_{\Omega^h} + \|\varphi_{\Gamma^h}\|_{\Gamma^h} \leq Ch^p \quad \text{as } h \to 0,$$

(where C is a positive constant independent of h) for all sufficiently smooth u, then the scheme is said to have *order of accuracy p*.

The finite difference scheme (2.11), (2.12) is said to *converge* to (2.9), (2.10) (and U is said to converge to u) in the norm $|||\cdot|||_{\Omega^h}$, if

$$|||u - U|||_{\Omega^h} \to 0 \quad \text{as } h \to 0.$$

If q is the largest positive real number such that, for all u sufficiently smooth,

$$|||u - U|||_{\Omega^h} \le Ch^q \quad \text{as } h \to 0$$

(where C is a positive constant independent of h), then the scheme is said to have *order of convergence* q.

From these definitions we deduce the following fundamental theorem.

Theorem 2.18 *Suppose that the finite difference scheme* (2.11), (2.12) *for problem* (2.9), (2.10) *is stable* (*i.e.* (2.47) *holds for all* f_h *and* g_h *and corresponding solution* U, *with* c_0 *independent of* h) *and that the scheme is consistent*; *then* (2.11), (2.12) *is a convergent approximation of* (2.9), (2.10) *and the order of convergence is not less than the order of accuracy.*

Proof We define the *global error* $e := u - U$; then,

$$L_h e = L_h(u - U) = L_h u - L_h U = L_h u - f_h.$$

Thus,

$$L_h e = \varphi_{\Omega^h}$$

and similarly

$$l_h e = \varphi_{\Gamma^h}.$$

By stability,

$$c_0 |||u - U|||_{\Omega^h} = c_0 |||e|||_{\Omega^h} \le \|\varphi_{\Omega^h}\|_{\Omega^h} + \|\varphi_{\Gamma^h}\|_{\Gamma^h},$$

and hence we arrive at the stated result. □

Paraphrasing Theorem 2.18, *stability* and *consistency* of the scheme imply its *convergence*. This abstract result is at the heart of the error analysis of finite difference approximations of differential equations.

2.2.4 The Multi-dimensional Case

Since the two-dimensional case is sufficiently representative, for the sake of notational simplicity we shall confine our attention to elliptic boundary-value problems in the plane.

Meshes and Divided Difference Operators Assuming that N is an integer, $N \geq 2$, we shall use a uniform square mesh Ω^h with mesh-size $h := 1/N$ over the unit square $\Omega := (0, 1)^2$, defined by

$$\Omega^h := \big\{x = (x_1, x_2) = (ih, jh) : i, j = 1, \ldots, N-1\big\},$$

and the square mesh

$$\overline{\Omega}^h := \big\{(ih, jh) : i, j = 0, \ldots, N\big\}.$$

Let $\Gamma := \partial\Omega$ be the boundary of Ω and define

$$\Gamma^h := h\mathbb{Z}^2 \cap \Gamma = \overline{\Omega}^h \setminus \Omega^h.$$

Analogously, let

$$\Gamma_{ik} := \{x \in \Gamma : x_i = k,\ 0 < x_{3-i} < 1\}, \quad i = 1, 2,\ k = 0, 1,$$

and define

$$\Gamma_{ik}^h := \Gamma_{ik} \cap h\mathbb{Z}^2, \qquad \overline{\Gamma}_{ik}^h := \overline{\Gamma}_{ik} \cap h\mathbb{Z}^2, \qquad \Gamma_*^h := \Gamma^h \setminus \big(\cup_{i,k}\Gamma_{ik}^h\big).$$

Let us also introduce

$$\Omega_i^h := \Omega^h \cup \Gamma_{i0}^h, \qquad \Omega_{i+2}^h := \Omega^h \cup \Gamma_{i1}^h, \quad i = 1, 2,$$
$$\Omega_{kl}^h := \Omega^h \cup \Gamma_{1k}^h \cup \Gamma_{2l}^h \cup \big\{(k, l)\big\}, \quad k, l = 0, 1.$$

Let S^h be the set of all real-valued functions defined on the mesh $\overline{\Omega}^h$. We shall use the notation $V_{ij} := V(ih, jh)$. By S_0^h we denote the set of all real-valued functions defined on the mesh $\overline{\Omega}^h$ that vanish at all points of Γ^h. The set S_0^h is equipped with the inner product

$$(V, W)_h = (V, W)_{L_2(\Omega^h)} := h^2 \sum_{x \in \Omega^h} V(x)W(x) = h^2 \sum_{i,j=1}^{N-1} V_{ij}W_{ij}, \tag{2.48}$$

and the norm

$$\|V\|_h = \|V\|_{L_2(\Omega^h)} := (V, V)_h^{1/2}.$$

The norms $\|\cdot\|_{L_2(\Omega_i^h)}$ and $\|\cdot\|_{L_2(\Omega_{kl}^h)}$ are defined analogously to $\|\cdot\|_{L_2(\Omega^h)}$.

The forward, backward and central divided difference operators on the mesh $\overline{\Omega}_h$ are defined analogously as in the one-dimensional case:

$$D_{x_i}^+ V := \frac{V^{+i} - V}{h}, \qquad D_{x_i}^- V := \frac{V - V^{-i}}{h}, \qquad D_{x_i}^0 V := \frac{1}{2}\big(D_{x_i}^+ V + D_{x_i}^- V\big),$$

where

$$V^{\pm i} := V^{\pm i}(x) = V(x \pm h\mathrm{e}_i), \quad \mathrm{e}_i := (\delta_{i1}, \delta_{i2}), \quad i = 1, 2,$$

and δ_{ik} is the Kronecker delta.

Discrete Sobolev Norms Analogously as in the one-dimensional case, we define the following discrete Sobolev seminorms on S^h:

$$\begin{aligned} |V|_{W_2^1(\Omega^h)} &:= \left(\|D_{x_1}^+ V\|^2_{L_2(\Omega_1^h)} + \|D_{x_2}^+ V\|^2_{L_2(\Omega_2^h)}\right)^{1/2} \\ &= \left(\|D_{x_1}^- V\|^2_{L_2(\Omega_3^h)} + \|D_{x_2}^- V\|^2_{L_2(\Omega_4^h)}\right)^{1/2}, \\ |V|_{W_2^2(\Omega^h)} &:= \left(\|D_{x_1}^+ D_{x_1}^- V\|^2_{L_2(\Omega^h)} + \|D_{x_1}^+ D_{x_2}^+ V\|^2_{L_2(\Omega_{00}^h)}\right. \\ &\quad \left. + \|D_{x_2}^+ D_{x_2}^- V\|^2_{L_2(\Omega^h)}\right)^{1/2} \end{aligned} \tag{2.49}$$

and the corresponding discrete Sobolev norms

$$\|V\|_{W_2^k(\Omega^h)} := \left(\|V\|^2_{W_2^{k-1}(\Omega^h)} + |V|^2_{W_2^k(\Omega^h)}\right)^{1/2}, \quad k = 1, 2, \tag{2.50}$$

with the notational convention $W_2^0(\Omega^h) := L_2(\Omega^h)$.

Let us also introduce the following inner products

$$[V, W]_h := h^2 \sum_{x \in \Omega^h} V(x)W(x) + \frac{h^2}{2} \sum_{x \in \Gamma^h \setminus \Gamma_*^h} V(x)W(x) + \frac{h^2}{4} \sum_{x \in \Gamma_*^h} V(x)W(x),$$

$$[V, W]_{i,h} := h^2 \sum_{x \in \Omega_i^h} V(x)W(x) + \frac{h^2}{2} \sum_{x \in \Gamma^h \setminus (\Gamma_{i0}^h \cup \overline{\Gamma}_{i1}^h)} V(x)W(x), \quad i = 1, 2,$$

and the associated norms

$$\begin{aligned} [\![V]\!]_h &= [\![V]\!]_{L_2(\Omega^h)} := [V, V]_h^{1/2}, \\ [\![V]\!]_i &= [\![V]\!]_{i,h} := [V, V]_{i,h}^{1/2}. \end{aligned}$$

In analogy with the one-dimensional case, we define the following discrete Sobolev seminorms and norms on S^h:

$$\begin{aligned} [V]_{W_2^1(\Omega^h)} &:= \left([\![D_{x_1}^+ V]\!]_1^2 + [\![D_{x_2}^+ V]\!]_2^2\right)^{1/2}, \\ [V]_{W_2^2(\Omega^h)} &:= \left([\![\overline{\Lambda}_1 V]\!]^2_{L_2(\Omega^h)} + \|D_{x_1}^+ D_{x_2}^+ V\|^2_{L_2(\Omega_{00}^h)} + [\![\overline{\Lambda}_2 V]\!]^2_{L_2(\Omega^h)}\right)^{1/2}, \\ [\![V]\!]_{W_2^k(\Omega^h)} &:= \left([\![V]\!]^2_{W_2^{k-1}(\Omega^h)} + [V]^2_{W_2^k(\Omega^h)}\right)^{1/2}, \quad k = 1, 2, \end{aligned}$$

where

$$(\overline{\Lambda}_i V)(x) := \begin{cases} -\frac{2}{h} D^+_{x_i} V & \text{if } x \in \overline{\Gamma}^h_{i0}, \\ -D^+_{x_i} D^-_{x_i} V(x) & \text{if } x \in \Omega^h \cup \Gamma^h_{3-i,0} \cup \Gamma^h_{3-i,1}, \\ \frac{2}{h} D^-_{x_i} V & \text{if } x \in \overline{\Gamma}^h_{i1}. \end{cases}$$

The Discrete Laplace Operator on S^h_0 We consider the discrete analogue of the Laplace operator in two space dimensions, defined on $h\mathbb{Z}^2$ by

$$\Delta_h V := D^+_{x_1} D^-_{x_1} V + D^+_{x_2} D^-_{x_2} V.$$

The mapping $\Lambda : S^h_0 \to S^h_0$ defined, for $V \in S^h_0$, by

$$(\Lambda V)(x) = \begin{cases} -(\Delta_h V)(x) & \text{if } x \in \Omega^h, \\ 0 & \text{if } x \in \Gamma^h, \end{cases}$$

positive definite operator with respect to the inner product $(\cdot, \cdot)_h$. In particular, for $V \in S^h_0$ we have that

$$(\Lambda V, V)_h = (-\Delta_h V, V)_h = |V|^2_{W^1_2(\Omega^h)}. \tag{2.51}$$

Furthermore,

$$\|\Delta_h V\|^2_h = \left\| D^+_{x_1} D^-_{x_1} V \right\|^2_{L_2(\Omega^h)} + 2 \left\| D^+_{x_1} D^+_{x_2} V \right\|^2_{L_2(\Omega^h_{00})} + \left\| D^+_{x_2} D^-_{x_2} V \right\|^2_{L_2(\Omega^h)},$$

and therefore,

$$\|\Delta_h V\|^2_h \ge |V|^2_{W^2_2(\Omega^h)}.$$

Similarly,

$$\|\Delta_h V\|^2_h \ge 16(-\Delta_h V, V)_h \ge 16^2 \|V\|^2_h = 16^2 \|V\|^2_{L_2(\Omega^h)}$$

and

$$|V|_{W^2_2(\Omega^h)} \ge 2\sqrt{2} |V|_{W^1_2(\Omega^h)} \ge 8\sqrt{2} \|V\|_{L_2(\Omega^h)}, \quad V \in S^h_0. \tag{2.52}$$

Consequently, on the linear space S^h_0 the seminorms $|\cdot|_{W^1_2(\Omega^h)}$ and $|\cdot|_{W^2_2(\Omega^h)}$ are equivalent to the norms $\|\cdot\|_{W^1_2(\Omega^h)}$ and $\|\cdot\|_{W^2_2(\Omega^h)}$, respectively.

Lemma 2.19 (Discrete Friedrichs Inequality) *There exists a positive real number $c_\star$, independent of h, such that*

$$\|V\|^2_{W^1_2(\Omega^h)} \le c_\star \left(\left\| D^+_{x_1} V \right\|^2_{L_2(\Omega^h_1)} + \left\| D^+_{x_2} V \right\|^2_{L_2(\Omega^h_2)} \right) \tag{2.53}$$

for all V in S^h_0.

Proof Inequality (2.53) with $c_\star = 17/16$ follows directly from the definition (2.49) of the seminorm $|\cdot|_{W_2^1(\Omega^h)}$ and the second inequality in (2.52). □

Fractional-Order Discrete Sobolev Norms We define the fractional-order discrete Sobolev seminorm $|\cdot|_{W_2^r(\Omega^h)}$ by

$$|V|^2_{W_2^r(\Omega^h)} := \sum_{i=1}^{2} h^3 \sum_{\substack{x_i,t_i=0\\ x_i\neq t_i}}^{Nh} \sum_{x_{3-i}=h}^{(N-1)h} \frac{[V(x)-V(t_i\mathrm{e}_i + x_{3-i}\mathrm{e}_{3-i})]^2}{|x_i-t_i|^{1+2r}}$$

if $0 < r < 1$, and by

$$|V|^2_{W_2^r(\Omega^h)} := \sum_{i=1}^{2} h^3 \sum_{\substack{x_i,t_i=0\\ x_i\neq t_i}}^{(N-1)h} \sum_{x_{3-i}=0}^{(N-1)h} \frac{[D^+_{x_i}V(x)-D^+_{x_i}V(t_i\mathrm{e}_i + x_{3-i}\mathrm{e}_{3-i})]^2}{|x_i-t_i|^{1+2(r-1)}}$$
$$+ \sum_{i=1}^{2} h^3 \sum_{\substack{x_{3-i},t_{3-i}=0\\ x_{3-i}\neq t_{3-i}}}^{Nh} \sum_{x_i=0}^{(N-1)h} \frac{[D^+_{x_i}V(x)-D^+_{x_i}V(x_i\mathrm{e}_i + t_{3-i}\mathrm{e}_{3-i})]^2}{|x_{3-i}-t_{3-i}|^{1+2(r-1)}}$$

if $1 < r < 2$. We also introduce the associated fractional-order discrete Sobolev norm by

$$\|V\|_{W_2^r(\Omega^h)} := \left(\|V\|^2_{W_2^{[r]}(\Omega^h)} + |V|^2_{W_2^r(\Omega^h)}\right)^{1/2}, \quad 0 < r < 2,\ r \neq 1.$$

Similarly as in one dimension, we have the interpolation inequalities

$$\begin{aligned} \|V\|_{W_2^r(\Omega^h)} &\le C(r)\|V\|^{1-r}_{L_2(\Omega^h)}\|V\|^r_{W_2^1(\Omega^h)}, \quad 0<r<1,\\ \|V\|_{W_2^r(\Omega^h)} &\le C(r)\|V\|^{2-r}_{W_2^1(\Omega^h)}\|V\|^{r-1}_{W_2^2(\Omega^h)}, \quad 1<r<2, \end{aligned} \tag{2.54}$$

which follow directly from their one-dimensional counterparts.

2.2.5 Approximation of a Generalized Poisson Problem

In Sect. 2.2.3 we presented a detailed error analysis for a finite difference approximation of a simple two-point boundary-value problem. Here we shall undertake a similar study for the generalized Poisson equation in two space dimensions subject to a homogeneous Dirichlet boundary condition:

$$-\Delta u + c(x,y)u = f(x,y) \quad \text{in } \Omega, \tag{2.55}$$

$$u = 0 \quad \text{on } \Gamma = \partial\Omega, \tag{2.56}$$

where $\Omega := (0,1)\times(0,1)$, c is a continuous function on $\overline{\Omega}$ and $c(x,y)\geq 0$. For the sake of notational simplicity we have denoted the two independent variables by x and y, instead of x_1 and x_2. As far as the smoothness of the function f is concerned, we shall consider two distinct cases:

(a) First we shall assume that f is continuous on $\overline{\Omega}$. In this case, the error analysis proceeds along the same lines as in Sect. 2.2.3.
(b) We shall then consider the case when f is in $L_2(\Omega)$ only; then the boundary-value problem (2.55), (2.56) does not necessarily have a classical solution; nevertheless, a weak solution still exists. This lack of smoothness gives rise to some technical difficulties both in the formulation of an adequate finite difference scheme and its error analysis. Since the point values of f need not be meaningful at the mesh-points (after all, f can be changed on a subset of Ω of zero Lebesgue measure without altering it as an element of $L_2(\Omega)$), instead of sampling the function f at the mesh-points we shall sample a mollified right-hand side $T_h f$. Also, since the analytical solution may not have a Taylor expansion with the required number of terms, we shall apply a different technique, based on integral representation theorems, to estimate the size of the truncation error.

We begin by considering the first of these two cases.

(a) ($f\in C(\overline{\Omega})$) The first step in the construction of the finite difference approximation to (2.55), (2.56) is to define the mesh. Let N be an integer, $N\geq 2$, and let $h:=1/N$; the mesh-points are (x_i,y_j), $i,j=0,\ldots,N$, where $x_i:=ih$, $y_j:=jh$. These mesh-points form the mesh

$$\overline{\Omega}^h := \big\{(x_i,y_j) : i,j=0,\ldots,N\big\}.$$

Similarly as in Sect. 2.2.2, we consider the set of interior mesh-points

$$\Omega^h := \big\{(x_i,y_j) : i,j=1,\ldots,N-1\big\}$$

and the set of boundary mesh-points

$$\Gamma^h := \overline{\Omega}^h \setminus \Omega^h.$$

In analogy with (2.35), (2.36), the finite difference approximation of (2.55), (2.56) is:

$$-\big(D_x^+D_x^-U_{ij} + D_y^+D_y^-U_{ij}\big) + c(x_i,y_j)U_{ij} = f(x_i,y_j), \quad (x_i,y_j)\in\Omega^h, \tag{2.57}$$

$$U = 0 \quad \text{on } \Gamma^h. \tag{2.58}$$

In expanded form, this can be written as follows:

$$\begin{aligned} -\bigg(\frac{U_{i+1,j}-2U_{ij}+U_{i-1,j}}{h^2} + \frac{U_{i,j+1}-2U_{ij}+U_{i,j-1}}{h^2}\bigg) + c(x_i,y_j)U_{ij} \\ = f(x_i,y_j) \quad \text{if } (x_i,y_j)\in\Omega^h, \end{aligned} \tag{2.59}$$

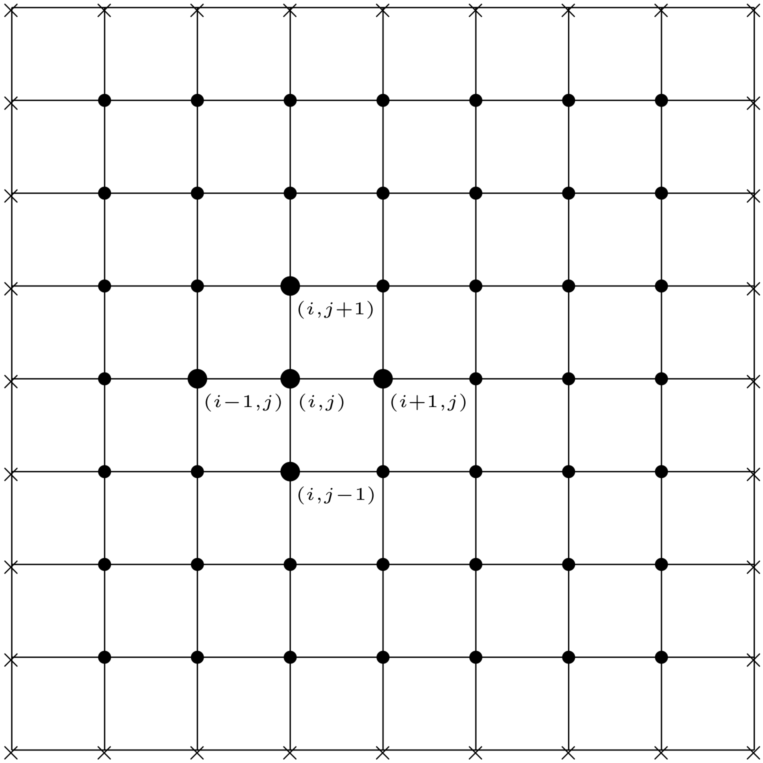

Fig. 2.1 The set of interior mesh-points Ω^h, denoted by $\bullet$, the set of boundary mesh-points Γ^h, denoted by $\times$, and a typical five-point difference stencil

$$U_{ij} = 0 \quad \text{if } (x_i, y_j) \in \Gamma^h, \tag{2.60}$$

where the divided difference operators $D_x^\pm = D_{x_1}^\pm$ and $D_y^\pm = D_{x_2}^\pm$ have been defined in Sect. 2.2.2.

For each i and j, $1 \le i, j \le N-1$, the finite difference equation (2.59) involves five values of the approximate solution U: $U_{i,j}$, $U_{i-1,j}$, $U_{i+1,j}$, $U_{i,j-1}$, $U_{i,j+1}$, as indicated in Fig. 2.1; hence its name: *five-point difference scheme*. It is possible to write (2.59), (2.60) as a system of linear equations

$$\mathcal{A}U = F, \tag{2.61}$$

where

$$\begin{aligned}
U &:= (U_{11}, U_{12}, \dots, U_{1,N-1}, U_{21}, U_{22}, \dots, U_{2,N-1}, \dots,\\
&\qquad \dots, U_{i1}, U_{i2}, \dots, U_{i,N-1}, \dots, U_{N-1,1}, U_{N-1,2}, \dots, U_{N-1,N-1})^{\mathrm{T}},\\
F &:= (F_{11}, F_{12}, \dots, F_{1,N-1}, F_{21}, F_{22}, \dots, F_{2,N-1}, \dots,\\
&\qquad \dots, F_{i1}, F_{i2}, \dots, F_{i,N-1}, \dots, F_{N-1,1}, F_{N-1,2}, \dots, F_{N-1,N-1})^{\mathrm{T}},
\end{aligned}$$

$$\mathcal{A} = \begin{pmatrix} K & J & O & O & \cdots & O & O \\ J & K & J & O & \cdots & O & O \\ O & J & K & J & \cdots & O & O \\ O & O & J & K & \cdots & O & O \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ O & O & O & O & \cdots & K & J \\ O & O & O & O & \cdots & J & K \end{pmatrix}$$

Fig. 2.2 The sparsity structure of the banded matrix $\mathcal{A}$: K is an $(N-1)\times(N-1)$ symmetric tridiagonal matrix, $J=(-1/h^2)I$, I is the $(N-1)\times(N-1)$ identity matrix, and O is the $(N-1)\times(N-1)$ zero matrix

and $\mathcal{A}$ is an $(N-1)^2\times(N-1)^2$ sparse, banded matrix.

A typical row of the matrix contains five nonzero entries, corresponding to the five values of U in the finite difference stencil shown in Fig. 2.1, while the sparsity structure of $\mathcal{A}$ is indicated in Fig. 2.2.

Next we show that (2.57), (2.58) has a unique solution. We proceed in the same way as in the previous section for the finite difference approximation of the two-point boundary-value problem. For two functions, V and W, defined on Ω^h, we introduce the discrete L_2-inner product (2.48):

$$(V,W)_h := \sum_{i=1}^{N-1}\sum_{j=1}^{N-1} h^2 V_{ij} W_{ij}.$$

Again, let S_0^h denote the set of functions V defined on $\overline{\Omega}^h$ such that $V=0$ on Γ^h.

We define the linear operator

$$A : S_0^h \to S_0^h$$

at mesh-points of Ω^h and Γ^h, respectively, as follows:

$$\begin{aligned} (AV)_{ij} &:= -\big(D_x^+D_x^-V_{ij} + D_y^+D_y^-V_{ij}\big) + c(x_i)V_i, \quad i,j=1,\dots,N-1,\\ (AV)_{i0} &= (AV)_{iN} = (AV)_{0j} = (AV)_{Nj} := 0, \quad i,j=0,\dots,N. \end{aligned}$$

Returning to the analysis of the finite difference scheme (2.57), (2.58), we note that, since $c(x,y)\ge 0$ on $\overline{\Omega}$, by (2.51) and (2.49) we have that

$$\begin{aligned} (AV,V)_h &= \big(-D_x^+D_x^-V - D_y^+D_y^-V + cV, V\big)_h \\ &= \big(-D_x^+D_x^-V, V\big)_h + \big(-D_y^+D_y^-V, V\big)_h + (cV,V)_h \\ &\ge \sum_{i=1}^{N}\sum_{j=1}^{N-1} h^2\big|D_x^-V_{ij}\big|^2 + \sum_{i=1}^{N-1}\sum_{j=1}^{N} h^2\big|D_y^-V_{ij}\big|^2, \end{aligned} \tag{2.62}$$

for any V in S_0^h. This implies, just as in the one-dimensional analysis presented in the previous section, that $\mathcal{A}$ is a nonsingular matrix. Indeed if $AV=0$, then (2.62) yields:

$$D_x^- V_{ij} = \frac{V_{ij} - V_{i-1,j}}{h} = 0, \qquad \begin{array}{l} i = 1, \dots, N, \\ j = 1, \dots, N-1; \end{array}$$

$$D_y^- V_{ij} = \frac{V_{ij} - V_{i,j-1}}{h} = 0, \qquad \begin{array}{l} i = 1, \dots, N-1, \\ j = 1, \dots, N. \end{array}$$

Since $V = 0$ on Γ^h, these imply that $V = 0$ on $\overline{\Omega}^h$. Thus $AV = 0$ if, and only if, $V = 0$. Hence $\mathcal{A}$ is nonsingular, and $U = \mathcal{A}^{-1}F$ is the unique solution of (2.57), (2.59); the solution may be found by solving the system of linear equations (2.61).

In order to prove the stability of the finite difference scheme (2.57), (2.58), we consider, similarly as in the one-dimensional case, the discrete L_2 norm

$$\|V\|_{L_2(\Omega^h)} := (V, V)_h^{1/2},$$

and the discrete W_2^1 norm (see (2.50))

$$\|V\|_{W_2^1(\Omega^h)} := \left(\|V\|^2_{L_2(\Omega^h)} + \left\|D_x^- V\right\|^2_{L_2(\Omega^h_x)} + \left\|D_y^- V\right\|^2_{L_2(\Omega^h_y)}\right)^{1/2},$$

where

$$\Omega^h_x := \Omega^h_3 = \left\{(x_i, y_j) : i = 1, \dots, N,\ j = 1, \dots, N-1\right\},$$

$$\Omega^h_y := \Omega^h_4 = \left\{(x_i, y_j) : i = 1, \dots, N-1,\ j = 1, \dots, N\right\}.$$

The norm $\|\cdot\|_{W_2^1(\Omega^h)}$ is the discrete analogue of the Sobolev norm $\|\cdot\|_{W_2^1(\Omega)}$ defined by

$$\|u\|_{W_2^1(\Omega)} := \left(\|u\|^2_{L_2(\Omega)} + \left\|\frac{\partial u}{\partial x}\right\|^2_{L_2(\Omega)} + \left\|\frac{\partial u}{\partial y}\right\|^2_{L_2(\Omega)}\right)^{1/2}.$$

In terms of this notation the inequality (2.62) has the following form:

$$(AV, V)_h \geq \left\|D_x^- V\right\|^2_{L_2(\Omega^h_x)} + \left\|D_y^- V\right\|^2_{L_2(\Omega^h_y)}. \tag{2.63}$$

The discrete Friedrichs inequality (2.53) and inequality (2.63) imply that

$$(AV, V)_h \geq c_0 \|V\|^2_{W_2^1(\Omega^h)}, \tag{2.64}$$

where $c_0 = 1/c_\star = 16/17$.

Theorem 2.20 *The scheme* (2.57), (2.58) *is stable in the sense that*

$$\|U\|_{W_2^1(\Omega^h)} \leq \frac{1}{c_0}\|f\|_{L_2(\Omega^h)}, \tag{2.65}$$

where $c_0 = 16/17$.

Proof The proof of this stability result is completely analogous to that of its one-dimensional counterpart (2.39), now using (2.64) and the Cauchy–Schwarz inequality. □

Having established the stability of the difference scheme (2.57), (2.58), we turn to the question of its accuracy. We define the *global error* e by

$$e_{ij} := u(x_i, y_j) - U_{ij}, \quad i, j = 0, \dots, N,$$

and the *truncation error* φ by

$$\varphi_{ij} := Au(x_i, y_j) - f(x_i, y_j), \quad i, j = 1, \dots, N-1.$$

Then,

$$\begin{aligned} Ae_{ij} &= \varphi_{ij}, \quad i, j = 1, \dots, N-1, \\ e &= 0 \quad \text{on } \Gamma^h. \end{aligned}$$

By noting (2.65) we have

$$\begin{aligned} \|u - U\|_{W_2^1(\Omega^h)} &= \|e\|_{W_2^1(\Omega^h)} \\ &\le \frac{1}{c_0}\|\varphi\|_{L_2(\Omega^h)}. \end{aligned} \tag{2.66}$$

Thus, in order to obtain a bound on the global error, it suffices to estimate the size of the truncation error in the $\|\cdot\|_{L_2(\Omega^h)}$ norm. To do so, let us assume that $u \in C^4(\overline{\Omega})$; then, by expanding each term in φ in a Taylor series about the point (x_i, y_j), we obtain

$$\begin{aligned} \varphi_{ij} &= \Delta u(x_i, y_j) - \big(D_x^+ D_x^- u(x_i, y_j) + D_y^+ D_y^- u(x_i, y_j)\big) \\ &= \left[\frac{\partial^2 u}{\partial x^2}(x_i, y_j) - D_x^+ D_x^- u(x_i, y_j)\right] + \left[\frac{\partial^2 u}{\partial y^2}(x_i, y_j) - D_y^+ D_y^- u(x_i, y_j)\right] \\ &= -\frac{h^2}{12}\left(\frac{\partial^4 u}{\partial x^4}(\xi_i, y_j) + \frac{\partial^4 u}{\partial y^4}(x_i, \eta_j)\right), \quad i, j = 1, \dots, N-1, \end{aligned}$$

where $\xi_i \in (x_{i-1}, x_{i+1})$, $\eta_j \in (y_{j-1}, y_{j+1})$.

Thus,

$$|\varphi_{ij}| \le \frac{h^2}{12}\left(\left\|\frac{\partial^4 u}{\partial x^4}\right\|_{C(\overline{\Omega})} + \left\|\frac{\partial^4 u}{\partial y^4}\right\|_{C(\overline{\Omega})}\right),$$

and we deduce that the truncation error φ satisfies the bound

$$\|\varphi\|_{L_2(\Omega^h)} \le \frac{h^2}{12}\left(\left\|\frac{\partial^4 u}{\partial x^4}\right\|_{C(\overline{\Omega})} + \left\|\frac{\partial^4 u}{\partial y^4}\right\|_{C(\overline{\Omega})}\right). \tag{2.67}$$

Finally (2.66) and (2.67) yield the following error bound.

Theorem 2.21 *Let* $f \in C(\overline{\Omega})$, $c \in C(\overline{\Omega})$, *with* $c(x,y) \geq 0$, $(x,y) \in \overline{\Omega}$, *and suppose that the corresponding weak solution of the boundary-value problem* (2.55), (2.56) *belongs to* $C^4(\overline{\Omega})$; *then*

$$\|u - U\|_{W_2^1(\Omega^h)} \leq \frac{17h^2}{192}\left(\left\|\frac{\partial^4 u}{\partial x^4}\right\|_{C(\overline{\Omega})} + \left\|\frac{\partial^4 u}{\partial y^4}\right\|_{C(\overline{\Omega})}\right). \tag{2.68}$$

Proof Recall that $1/c_0 = c_\star = 17/16$, and combine (2.66) and (2.67). □

According to this result, the five-point difference scheme (2.57), (2.58) for the boundary-value problem (2.55), (2.56) is second-order convergent, provided that u is sufficiently smooth; i.e. $u \in C^4(\overline{\Omega})$.

Elliptic regularity theory tells us (see, for example, Ladyzhenskaya and Ural'tseva [118], Gilbarg and Trudinger [53] or Renardy and Rogers [155]) that if the right-hand side and the coefficients are "sufficiently smooth", then the associated classical solution of the elliptic problem is "as smooth as one would expect" in the *interior* of the domain on which the problem is posed; e.g. in the case of a second-order elliptic boundary-value problem, if $f \in C^{k,\alpha}(\Omega)$, $k \geq 0$, $0 < \alpha < 1$, then $u \in C^{k+2,\alpha}(\Omega)$. Unfortunately, in general, the solution will not be smooth up to the boundary if the boundary is not of class $C^{k+2,\alpha}$, as is the case when Ω is a square. For a simple illustration, we refer to Example 9.52 on p. 325 of Renardy and Rogers [155]; a more detailed account of regularity theory for elliptic equations in domains with nonsmooth boundaries is given in Grisvard [62, 63] and Dauge [28]. Thus, in general, the solution of our simple model problem (2.55), (2.56), will not belong to $C^4(\overline{\Omega})$ even if f and c are smooth functions, because the boundary $\Gamma = \partial\Omega$ is only of class $C^{0,1}$. Consequently, the hypothesis $u \in C^4(\overline{\Omega})$ that was made in the statement of Theorem 2.21 is unrealistic (unless f satisfies suitable compatibility conditions at the four corners of Ω (cf. (2.8))).

Our analysis has another limitation: it was performed under the assumption that $f \in C(\overline{\Omega})$, which was necessary in order to ensure that the values of f are meaningfully defined at the mesh-points. However, in applications one often encounters differential equations where f is a lot less smooth (e.g. f is piecewise continuous, or $f \in L_2(\Omega)$, or f is a Borel measure). When $f \in L_2(\Omega)$, for example, we know that the homogeneous Dirichlet boundary-value problem for the partial differential equation $-\Delta u + cu = f$, with c bounded and nonnegative, still has a unique weak solution in $H_0^1(\Omega)$, so it is natural to ask whether one can construct a second-order accurate finite difference approximation of the weak solution. This brings us to case (b), formulated at the beginning of the section.

(b) ($f \in L_2(\Omega)$). We shall use the same finite difference mesh as in case (a), but we shall modify the difference scheme (2.57), (2.58) to cater for the fact that f is not continuous on $\overline{\Omega}$. The idea is to replace $f(x_i, y_j)$ in (2.57) by a cell-average

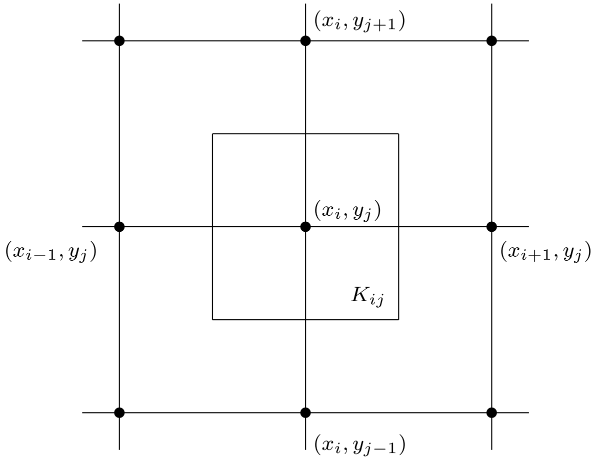

Fig. 2.3 The cell K_{ij}

of f:

$$\left(T_h^{11} f\right)_{ij} := \frac{1}{h^2} \int_{K_{ij}} f(x, y) \,\mathrm{d}x \,\mathrm{d}y,$$

where the 'cell' K_{ij} is defined by

$$K_{ij} := \left(x_i - \frac{h}{2}, x_i + \frac{h}{2}\right) \times \left(y_j - \frac{h}{2}, y_j + \frac{h}{2}\right),$$

with $i, j = 1, \ldots, N-1$.

This seemingly ad hoc approach has the following justification. Integrating the partial differential equation $-\Delta u + cu = f$ over the cell K_{ij} and using the divergence theorem we have that

$$-\int_{\partial K_{ij}} \frac{\partial u}{\partial \nu} \,\mathrm{d}s + \int_{K_{ij}} cu \,\mathrm{d}x \,\mathrm{d}y = \int_{K_{ij}} f \,\mathrm{d}x \,\mathrm{d}y, \tag{2.69}$$

where ∂K_{ij} is the boundary of K_{ij}, and ν is the unit outward normal to ∂K_{ij}.

The normal vectors to ∂K_{ij} point in the co-ordinate directions, so the normal derivative $\partial u/\partial \nu$ can be approximated by divided differences using the values of u at the five mesh-points marked by $\bullet$ in Fig. 2.3, in conjunction with a midpoint quadrature rule along each edge of K_{ij} to approximate the contour integral featuring in the first term of (2.69) (cf. Examples 2.6 and 2.7).

Approximating the second integral on the left by a midpoint quadrature rule, now in two dimensions, on K_{ij}, and dividing both sides by $\mathrm{meas}(K_{ij}) = h^2$, we obtain

$$\begin{aligned}
&-\left(D_x^+ D_x^- u(x_i, y_j) + D_y^+ D_y^- u(x_i, y_j)\right) + c(x_i, y_j) u(x_i, y_j) \\
&\quad \approx \frac{1}{h^2} \int_{K_{ij}} f(x, y) \,\mathrm{d}x \,\mathrm{d}y.
\end{aligned}$$

We note here that $(T_h^{11} f)_{ij}$ is correctly defined for $f \in L_2(\Omega)$; indeed,

$$\begin{aligned}\left|\left(T_h^{11} f\right)_{ij}\right| &= \frac{1}{h^2}\left|\int_{K_{ij}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y\right| \\ &\le \frac{1}{h^2}\left(\int_{K_{ij}} 1^2\,\mathrm{d}x\,\mathrm{d}y\right)^{1/2}\left(\int_{K_{ij}} \left|f(x,y)\right|^2\,\mathrm{d}x\,\mathrm{d}y\right)^{1/2} \\ &= h^{-1}\|f\|_{L_2(K_{ij})}(<\infty). \end{aligned} \tag{2.70}$$

Thus we define our finite difference scheme for (2.55), (2.56) by

$$-\left(D_x^+ D_x^- + D_y^+ D_y^-\right)U_{ij} + c(x_i, y_j)U_{ij} = \left(T_h^{11} f\right)_{ij}, \quad (x_i, y_j) \in \Omega^h, \tag{2.71}$$

$$U = 0 \quad \text{on } \Gamma^h. \tag{2.72}$$

Remark 2.10 Finite difference schemes that arise from integral formulations of a differential equation, such as (2.69), are called *finite volume methods*.

Since we have not changed the difference operator on the left-hand side, the argument presented in (a) concerning the existence and uniqueness of a solution to the difference scheme (2.57), (2.58) still applies to (2.71), (2.72); therefore, (2.71), (2.72) has a unique solution U in S_0^h. Moreover, we have the following stability result.

Theorem 2.22 *The scheme* (2.71), (2.72) *is stable in the sense that*

$$\|U\|_{W_2^1(\Omega^h)} \le \frac{1}{c_0}\|f\|_{L_2(\Omega)}, \tag{2.73}$$

where $c_0 = 16/17$.

Proof From (2.64) and (2.70) we have

$$\begin{aligned} c_0\|U\|^2_{W_2^1(\Omega^h)} &\le (AU, U)_h = \left(T_h^{11} f, U\right)_h \\ &\le \left\|T_h^{11} f\right\|_{L_2(\Omega^h)}\|U\|_{L_2(\Omega^h)} \le \left\|T_h^{11} f\right\|_{L_2(\Omega^h)}\|U\|_{W_2^1(\Omega^h)} \\ &\le \|f\|_{L_2(\Omega)}\|U\|_{W_2^1(\Omega^h)}, \end{aligned}$$

and hence (2.73). □

Having established the stability of the scheme (2.71), (2.72) we consider the question of its accuracy. Let us define the global error, e, as before:

$$e_{ij} := u(x_i, y_j) - U_{ij}, \quad i, j = 0, \ldots, N.$$

Clearly, for $i, j = 1, \ldots, N-1$ we have

$$\begin{aligned}
Ae_{ij} &= Au(x_i, y_j) - AU_{ij} \\
&= Au(x_i, y_j) - \left(T_h^{11} f\right)_{ij} \\
&= -\left(D_x^+ D_x^- u(x_i, y_j) + D_y^+ D_y^- u(x_i, y_j)\right) + c(x_i, y_j)u(x_i, y_j) \\
&\quad + \left[T_h^{11}\left(\frac{\partial^2 u}{\partial x^2}\right)(x_i, y_j) + T_h^{11}\left(\frac{\partial^2 u}{\partial y^2}\right)(x_i, y_j) - T_h^{11}(cu)(x_i, y_j)\right].
\end{aligned} \tag{2.74}$$

By noting that

$$\begin{aligned}
T_h^{11}\left(\frac{\partial^2 u}{\partial x^2}\right)(x_i, y_j) &= \frac{1}{h}\int_{y_j-h/2}^{y_j+h/2} \frac{\frac{\partial u}{\partial x}(x_i + h/2, y) - \frac{\partial u}{\partial x}(x_i - h/2, y)}{h}\, \mathrm{d}y \\
&= \frac{1}{h}\int_{y_j-h/2}^{y_j+h/2} D_x^+ \frac{\partial u}{\partial x}(x_i - h/2, y)\, \mathrm{d}y \\
&= D_x^+\left[\frac{1}{h}\int_{y_j-h/2}^{y_j+h/2} \frac{\partial u}{\partial x}(x_i - h/2, y)\, \mathrm{d}y\right],
\end{aligned}$$

and that, similarly,

$$T_h^{11}\left(\frac{\partial^2 u}{\partial y^2}\right)(x_i, y_j) = D_y^+\left[\frac{1}{h}\int_{x_i-h/2}^{x_i+h/2} \frac{\partial u}{\partial y}(x, y_j - h/2)\, \mathrm{d}x\right],$$

equality (2.74) can be rewritten as

$$Ae = D_x^+ \varphi_1 + D_y^+ \varphi_2 + \psi,$$

where

$$\begin{aligned}
\varphi_1(x_i, y_j) &:= \frac{1}{h}\int_{y_j-h/2}^{y_j+h/2} \frac{\partial u}{\partial x}(x_i - h/2, y)\, \mathrm{d}y - D_x^- u(x_i, y_j), \\
\varphi_2(x_i, y_j) &:= \frac{1}{h}\int_{x_i-h/2}^{x_i+h/2} \frac{\partial u}{\partial y}(x, y_j - h/2)\, \mathrm{d}x - D_y^- u(x_i, y_j), \\
\psi(x_i, y_j) &:= (cu)(x_i, y_j) - T_h^{11}(cu)(x_i, y_j).
\end{aligned}$$

Thus,

$$Ae = D_x^+ \varphi_1 + D_y^+ \varphi_2 + \psi \quad \text{in } \Omega^h, \tag{2.75}$$

$$e = 0 \quad \text{on } \Gamma^h. \tag{2.76}$$

As the stability result (2.73) implies only the crude bound

$$\|e\|_{W_2^1(\Omega^h)} \leq \frac{1}{c_0} \left\| D_x^+ \varphi_1 + D_y^+ \varphi_2 + \psi \right\|_{L_2(\Omega^h)},$$

which does not exploit the special form of the truncation error,

$$\varphi := D_x^+ \varphi_1 + D_y^+ \varphi_2 + \psi,$$

we shall proceed in a different way. The idea is to sharpen (2.73) by proving a discrete analogue of the well-posedness result from Theorem 2.7; we recall that this states that the following bound holds for the boundary-value problem (2.55), (2.56):

$$\|u\|_{\mathring{W}_2^1(\Omega)} \leq \frac{1}{c_0} \|f\|_{W_2^{-1}(\Omega)}.$$

In order to obtain a discrete counterpart of this inequality, we consider the discrete negative Sobolev norm $\|\cdot\|_{W_2^{-1}(\Omega^h)}$, defined by

$$\|V\|_{W_2^{-1}(\Omega^h)} := \sup_{V \in S_0^h \setminus \{0\}} \frac{|(V, W)_h|}{\|W\|_{W_2^1(\Omega^h)}}.$$

Theorem 2.23 *The scheme* (2.71), (2.72) *is stable in the sense that*

$$\|U\|_{W_2^1(\Omega^h)} \leq \frac{1}{c_0} \left\| T_h^{11} f \right\|_{W_2^{-1}(\Omega^h)}, \tag{2.77}$$

where $c_0 = 16/17$.

Proof From (2.64), by noting the definition of the $\|\cdot\|_{W_2^{-1}(\Omega^h)}$ norm, we have that

$$\begin{aligned} c_0 \|U\|^2_{W_2^1(\Omega^h)} &\leq (AU, U)_h = \left(T_h^{11} f, U\right)_h \\ &\leq \left\| T_h^{11} f \right\|_{W_2^{-1}(\Omega^h)} \|U\|_{W_2^1(\Omega^h)}, \end{aligned}$$

and hence (2.77). □

Now we apply Theorem 2.23 to (2.75), (2.76) to deduce that

$$\|e\|_{W_2^1(\Omega^h)} \leq \frac{1}{c_0} \left\| D_x^+ \varphi_1 + D_y^+ \varphi_2 + \psi \right\|_{W_2^{-1}(\Omega^h)}. \tag{2.78}$$

In order to bound the right-hand side of (2.78) let us consider the expression

$$\left(D_x^+ \varphi_1 + D_y^+ \varphi_2 + \psi, W\right)_h$$

for $W \in S_0^h \setminus \{0\}$. Using summation by parts, we shall pass the difference operators D_x^+ and D_y^+ from φ_1 and φ_2, respectively, onto W. As $W = 0$ on the set Γ^h, we have that

$$\begin{aligned}
\left(D_x^+ \varphi_1, W\right)_h &= \sum_{j=1}^{N-1} h \left(\sum_{i=1}^{N-1} h \frac{\varphi_1(x_{i+1}, y_j) - \varphi_1(x_i, y_j)}{h} W_{ij} \right) \\
&= - \sum_{j=1}^{N-1} h \left(\sum_{i=1}^{N} h \varphi_1(x_i, y_j) \frac{W_{ij} - W_{i-1,j}}{h} \right) \\
&= - \sum_{j=1}^{N-1} h \left(\sum_{i=1}^{N} h \varphi_1(x_i, y_j) D_x^- W_{ij} \right) \\
&= - \sum_{i=1}^{N} \sum_{j=1}^{N-1} h^2 \varphi_1(x_i, y_j) D_x^- W_{ij} \\
&\le \left(\sum_{i=1}^{N} \sum_{j=1}^{N-1} h^2 \left|\varphi_1(x_i, y_j)\right|^2 \right)^{1/2} \left(\sum_{i=1}^{N} \sum_{j=1}^{N-1} h^2 \left|D_x^- W_{ij}\right|^2 \right)^{1/2}.
\end{aligned}$$

We thus deduce that

$$\left|\left(D_x^+ \varphi_1, W\right)_h\right| \le \|\varphi_1\|_{L_2(\Omega_x^h)} \left\|D_x^- W\right\|_{L_2(\Omega_x^h)}. \tag{2.79}$$

Similarly,

$$\left|\left(D_y^+ \varphi_2, W\right)_h\right| \le \|\varphi_2\|_{L_2(\Omega_y^h)} \left\|D_y^- W\right\|_{L_2(\Omega_y^h)}. \tag{2.80}$$

By the Cauchy–Schwarz inequality we also have that

$$\left|(\psi, W)_h\right| \le \|\psi\|_{L_2(\Omega^h)} \|W\|_{L_2(\Omega^h)}. \tag{2.81}$$

Now, by combining (2.79)–(2.81) and noting the elementary inequality

$$|a_1 b_1 + a_2 b_2 + a_3 b_3| \le \left(a_1^2 + a_2^2 + a_3^2\right)^{1/2} \left(b_1^2 + b_2^2 + b_3^2\right)^{1/2},$$

we arrive at the bound

$$\begin{aligned}
&\left|\left(D_x^+ \varphi_1 + D_y^+ \varphi_2 + \psi, W\right)_h\right| \\
&\quad \le \left(\|\varphi_1\|_{L_2(\Omega_x^h)}^2 + \|\varphi_2\|_{L_2(\Omega_y^h)}^2 + \|\psi\|_{L_2(\Omega^h)}^2\right)^{1/2} \\
&\qquad \times \left(\left\|D_x^- W\right\|_{L_2(\Omega_x^h)}^2 + \left\|D_y^- W\right\|_{L_2(\Omega_y^h)}^2 + \|W\|_{L_2(\Omega^h)}^2\right)^{1/2} \\
&\quad = \left(\|\varphi_1\|_{L_2(\Omega_x^h)}^2 + \|\varphi_2\|_{L_2(\Omega_y^h)}^2 + \|\psi\|_{L_2(\Omega^h)}^2\right)^{1/2} \|W\|_{W_2^1(\Omega^h)}.
\end{aligned}$$

Dividing both sides by $\|W\|_{W_2^1(\Omega^h)}$ and taking the supremum over all $W \in S_0^h \setminus \{0\}$ yields the following inequality:

$$\left\| D_x^+\varphi_1 + D_y^+\varphi_2 + \psi \right\|_{W_2^{-1}(\Omega^h)} \le \left(\|\varphi_1\|^2_{L_2(\Omega_x^h)} + \|\varphi_2\|^2_{L_2(\Omega_y^h)} + \|\psi\|^2_{L_2(\Omega^h)}\right)^{1/2}. \tag{2.82}$$

Inserting (2.82) into (2.78) we obtain the following bound on the global error in terms of the truncation error of the scheme.

Lemma 2.24 *The global error, $e := u - U$, of the finite difference scheme* (2.71), (2.72) *satisfies the bound*

$$\|e\|_{W_2^1(\Omega^h)} \le \frac{1}{c_0}\left(\|\varphi_1\|^2_{L_2(\Omega_x^h)} + \|\varphi_2\|^2_{L_2(\Omega_y^h)} + \|\psi\|^2_{L_2(\Omega^h)}\right)^{1/2}, \tag{2.83}$$

where $c_0 = 16/17$, and φ_1, φ_2 and ψ are defined by

$$\varphi_1(x_i, y_j) := \frac{1}{h}\int_{y_j-h/2}^{y_j+h/2} \frac{\partial u}{\partial x}(x_i - h/2, y)\mathrm{d}y - D_x^- u(x_i, y_j), \tag{2.84}$$

$$\varphi_2(x_i, y_j) := \frac{1}{h}\int_{x_i-h/2}^{x_i+h/2} \frac{\partial u}{\partial y}(x, y_j - h/2)\,\mathrm{d}x - D_y^- u(x_i, y_j), \tag{2.85}$$

$$\psi(x_i, y_j) := (cu)(x_i, y_j) - \frac{1}{h^2}\int_{x_i-h/2}^{x_i+h/2}\int_{y_j-h/2}^{y_j+h/2} (cu)(x, y)\,\mathrm{d}x\,\mathrm{d}y, \tag{2.86}$$

with $i = 1, \dots, N$ and $j = 1, \dots, N-1$ in (2.84); *$i = 1, \dots, N-1$ and $j = 1, \dots, N$ in* (2.85); *and $i, j = 1, \dots N-1$ in* (2.86).

To complete the error analysis, it remains to bound φ_1, φ_2 and ψ. Using Taylor series expansions it is easily seen that

$$\left|\varphi_1(x_i, y_j)\right| \le \frac{h^2}{24}\left(\left\|\frac{\partial^3 u}{\partial x \partial y^2}\right\|_{C(\overline{\Omega})} + \left\|\frac{\partial^3 u}{\partial x^3}\right\|_{C(\overline{\Omega})}\right), \tag{2.87}$$

$$\left|\varphi_2(x_i, y_j)\right| \le \frac{h^2}{24}\left(\left\|\frac{\partial^3 u}{\partial x^2 \partial y}\right\|_{C(\overline{\Omega})} + \left\|\frac{\partial^3 u}{\partial y^3}\right\|_{C(\overline{\Omega})}\right), \tag{2.88}$$

$$\left|\psi(x_i, y_j)\right| \le \frac{h^2}{24}\left(\left\|\frac{\partial^2 (cu)}{\partial x^2}\right\|_{C(\overline{\Omega})} + \left\|\frac{\partial^2 (cu)}{\partial y^2}\right\|_{C(\overline{\Omega})}\right), \tag{2.89}$$

which yield the required bounds on $\|\varphi_1\|_{L_2(\Omega_x^h)}$, $\|\varphi_2\|_{L_2(\Omega_y^h)}$ and $\|\psi\|_{L_2(\Omega^h)}$. We thus arrive at the following theorem.

Theorem 2.25 *Let $f \in L_2(\Omega)$, $c \in C^2(\overline{\Omega})$ with $c(x, y) \ge 0$, $(x, y) \in \overline{\Omega}$, and suppose that the corresponding weak solution of the boundary-value problem* (2.55), (2.56) *belongs to* $C^3(\overline{\Omega})$. *Then*,

$$\|u - U\|_{W_2^1(\Omega^h)} \leq \frac{17}{384} h^2 M_3, \tag{2.90}$$

where

$$\begin{aligned} M_3 = \Bigg\{ & \left(\left\| \frac{\partial^3 u}{\partial x \partial y^2} \right\|_{C(\overline{\Omega})} + \left\| \frac{\partial^3 u}{\partial x^3} \right\|_{C(\overline{\Omega})} \right)^2 \\ & + \left(\left\| \frac{\partial^3 u}{\partial x^2 y} \right\|_{C(\overline{\Omega})} + \left\| \frac{\partial^3 u}{\partial y^3} \right\|_{C(\overline{\Omega})} \right)^2 \\ & + \left(\left\| \frac{\partial^2 (cu)}{\partial x^2} \right\|_{C(\overline{\Omega})} + \left\| \frac{\partial^2 (cu)}{\partial y^2} \right\|_{C(\overline{\Omega})} \right)^2 \Bigg\}^{1/2}. \end{aligned}$$

Proof As $1/c_0 = 17/16$, by substituting (2.87)–(2.89) into the right-hand side of (2.83) the estimate (2.90) immediately follows. □

By comparing (2.90) with (2.68) we see that while the smoothness requirement on the solution has been relaxed from $u \in C^4(\overline{\Omega})$ to $u \in C^3(\overline{\Omega})$, second-order convergence has been retained.

The hypothesis $u \in C^3(\overline{\Omega})$ can be further relaxed by using integral representations of φ_1, φ_2 and ψ instead of Taylor series expansions. We show how this is done for φ_1 and ψ; φ_2 is handled analogously to φ_1. The argument is based on repeated use the Newton–Leibniz formula

$$w(b) - w(a) = \int_a^b w'(x)\,\mathrm{d}x.$$

In order to simplify the notation, let us write $x_{i\pm1/2} := x_i \pm h/2$ and $y_{j\pm1/2} := y_j \pm h/2$; we then have that

$$\begin{aligned} \varphi_1(x_i, y_j) &= \frac{1}{h^2} \int_{x_{i-1}}^{x_i} \int_{y_{j-1/2}}^{y_{j+1/2}} \left[\frac{\partial u}{\partial x}(x_{i-1/2}, y) - \frac{\partial u}{\partial x}(x, y_j) \right] \mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{h^2} \int_{x_{i-1}}^{x_i} \int_{y_{j-1/2}}^{y_{j+1/2}} \left[\frac{\partial u}{\partial x}(x_{i-1/2}, y) - \frac{\partial u}{\partial x}(x, y) \right] \mathrm{d}x\,\mathrm{d}y \\ &\quad + \frac{1}{h^2} \int_{x_{i-1}}^{x_i} \int_{y_{j-1/2}}^{y_{j+1/2}} \left[\frac{\partial u}{\partial x}(x, y) - \frac{\partial u}{\partial x}(x, y_j) \right] \mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{h^2} \int_{y_{j-1/2}}^{y_{j+1/2}} \left[\int_{x_{i-1}}^{x_i} \left(\int_x^{x_{i-1/2}} \frac{\partial^2 u}{\partial x^2}(\xi, y)\,\mathrm{d}\xi \right) \mathrm{d}x \right] \mathrm{d}y \\ &\quad + \frac{1}{h^2} \int_{x_{i-1}}^{x_i} \left[\int_{y_{j-1/2}}^{y_{j+1/2}} \left(\int_{y_j}^{y} \frac{\partial^2 u}{\partial x \partial y}(x, \eta)\,\mathrm{d}\eta \right) \mathrm{d}x \right] \mathrm{d}y. \end{aligned}$$

We thus deduce by partial integration that

$$\begin{aligned}
\varphi_1(x_i, y_j) &= \frac{1}{h^2}\int_{y_{j-1/2}}^{y_{j+1/2}}\left[x\int_x^{x_{i-1/2}}\frac{\partial^2 u}{\partial x^2}(\xi, y)\,\mathrm{d}\xi\Big|_{x=x_{i-1}}^{x=x_i}\right.\\
&\quad\left.+\int_{x_{i-1}}^{x_i} x\frac{\partial^2 u}{\partial x^2}(x, y)\,\mathrm{d}x\right]\mathrm{d}y\\
&\quad+\frac{1}{h^2}\int_{x_{i-1}}^{x_i}\left[y\int_{y_j}^{y}\frac{\partial^2 u}{\partial x\partial y}(x, \eta)\,\mathrm{d}\eta\Big|_{y=y_{j-1/2}}^{y=y_{j+1/2}}\right.\\
&\quad\left.-\int_{y_{j-1/2}}^{y_{j+1/2}} y\frac{\partial^2 u}{\partial x\partial y}(x, y)\,\mathrm{d}y\right]\mathrm{d}x\\
&= \frac{1}{h^2}\int_{y_{j-1/2}}^{y_{j+1/2}}\left[\int_{x_{i-1}}^{x_{i-1/2}}(x - x_{i-1})\frac{\partial^2 u}{\partial x^2}(x, y)\,\mathrm{d}x\right.\\
&\quad\left.+\int_{x_{i-1/2}}^{x_i}(x - x_i)\frac{\partial^2 u}{\partial x^2}(x, y)\,\mathrm{d}x\right]\mathrm{d}y\\
&\quad-\frac{1}{h^2}\int_{x_{i-1}}^{x_i}\left[\int_{y_{j-1/2}}^{y_j}(y - y_{j-1/2})\frac{\partial^2 u}{\partial x\partial y}(x, y)\,\mathrm{d}y\right.\\
&\quad\left.+\int_{y_j}^{y_{j+1/2}}(y - y_{j+1/2})\frac{\partial^2 u}{\partial x\partial y}(x, y)\,\mathrm{d}y\right]\mathrm{d}x.
\end{aligned}$$

We define the piecewise quadratic functions

$$A_i(x) = \begin{cases}\frac{1}{2}(x - x_{i-1})^2 & \text{if } x \in [x_{i-1}, x_{i-1/2}],\\ \frac{1}{2}(x - x_i)^2 & \text{if } x \in [x_{i-1/2}, x_i],\end{cases}$$

$$B_j(y) = \begin{cases}\frac{1}{2}(y - y_{j-1/2})^2 & \text{if } y \in [y_{j-1/2}, y_j],\\ \frac{1}{2}(y - y_{j+1/2})^2 & \text{if } y \in [y_j, y_{j+1/2}],\end{cases}$$

and note that A_i and B_j are continuous functions of their respective arguments; furthermore,

$$A_i(x_{i-1}) = A_i(x_i) = 0 \quad \text{and} \quad B_j(y_{j-1/2}) = B_j(y_{j+1/2}) = 0.$$

Integration by parts then yields

$$\begin{aligned}
\varphi_1(x_i, y_j) &= \frac{1}{h^2}\int_{y_{j-1/2}}^{y_{j+1/2}}\left[\int_{x_{i-1}}^{x_i} A_i'(x)\frac{\partial^2 u}{\partial x^2}(x, y)\,\mathrm{d}x\right]\mathrm{d}y\\
&\quad-\frac{1}{h^2}\int_{x_{i-1}}^{x_i}\left[\int_{y_{j-1/2}}^{y_{j+1/2}} B_j'(y)\frac{\partial^2 u}{\partial x\partial y}(x, y)\,\mathrm{d}y\right]\mathrm{d}x
\end{aligned}$$

$$= -\frac{1}{h^2}\int_{y_{j-1/2}}^{y_{j+1/2}}\left[\int_{x_{i-1}}^{x_i} A_i(x)\frac{\partial^3 u}{\partial x^3}(x,y)\,\mathrm{d}x\right]\mathrm{d}y$$

$$+\frac{1}{h^2}\int_{x_{i-1}}^{x_i}\left[\int_{y_{j-1/2}}^{y_{j+1/2}} B_j(y)\frac{\partial^3 u}{\partial x\partial y^2}(x,y)\,\mathrm{d}y\right]\mathrm{d}x. \tag{2.91}$$

Now

$$\left|A_i(x)\right| \le \frac{1}{8}h^2, \quad x\in[x_{i-1},x_i] \quad \text{and} \quad \left|B_j(y)\right| \le \frac{1}{8}h^2, \quad y\in[y_{j-1/2},y_{j+1/2}],$$

and therefore,

$$\left|\varphi_1(x_i,y_j)\right| \le \frac{1}{8}\int_{x_{i-1}}^{x_i}\int_{y_{j-1/2}}^{y_{j+1/2}}\left|\frac{\partial^3 u}{\partial x^3}(x,y)\right|\mathrm{d}x\,\mathrm{d}y$$

$$+\frac{1}{8}\int_{x_{i-1}}^{x_i}\int_{y_{j-1/2}}^{y_{j+1/2}}\left|\frac{\partial^3 u}{\partial x\partial y^2}(x,y)\right|\mathrm{d}x\,\mathrm{d}y.$$

Consequently,

$$\|\varphi_1\|^2_{L_2(\Omega_x^h)} \le \frac{h^4}{32}\left(\left\|\frac{\partial^3 u}{\partial x^3}\right\|^2_{L_2(\Omega)} + \left\|\frac{\partial^3 u}{\partial x\partial y^2}\right\|^2_{L_2(\Omega)}\right). \tag{2.92}$$

Analogously,

$$\|\varphi_2\|^2_{L_2(\Omega_y^h)} \le \frac{h^4}{32}\left(\left\|\frac{\partial^3 u}{\partial y^3}\right\|^2_{L_2(\Omega)} + \left\|\frac{\partial^3 u}{\partial x^2\partial y}\right\|^2_{L_2(\Omega)}\right). \tag{2.93}$$

In order to estimate ψ, we note that

$$\psi(x_i,y_j) = \frac{1}{h^2}\int_{x_{i-1/2}}^{x_{i+1/2}}\int_{y_{j-1/2}}^{y_{j+1/2}}\left(\int_x^{x_i}\frac{\partial w}{\partial x}(s,y)\,\mathrm{d}s\right.$$

$$\left.+\int_y^{y_j}\frac{\partial w}{\partial y}(x,t)\,\mathrm{d}t + \int_x^{x_i}\int_y^{y_j}\frac{\partial^2 w}{\partial x\partial y}(s,t)\,\mathrm{d}s\,\mathrm{d}t\right)\mathrm{d}x\,\mathrm{d}y$$

$$= -\frac{1}{h^2}\int_{x_{i-1/2}}^{x_{i+1/2}}\int_{y_{j-1/2}}^{y_{j+1/2}} C_i(x)\frac{\partial^2 w}{\partial x^2}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

$$-\frac{1}{h^2}\int_{x_{i-1/2}}^{x_{i+1/2}}\int_{y_{j-1/2}}^{y_{j+1/2}} B_j(y)\frac{\partial^2 w}{\partial y^2}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

$$+\frac{1}{h^2}\int_{x_{i-1/2}}^{x_{i+1/2}}\int_{y_{j-1/2}}^{y_{j+1/2}}\left(\int_x^{x_i}\int_y^{y_j}\frac{\partial^2 w}{\partial x\partial y}(s,t)\,\mathrm{d}s\,\mathrm{d}t\right)\mathrm{d}x\,\mathrm{d}y,$$

where $w(x,y)=c(x,y)u(x,y)$ and

$$C_i(x)=\begin{cases}\frac{1}{2}(x-x_{i-1/2})^2 & \text{if } x\in[x_{i-1/2},x_i],\\ \frac{1}{2}(x-x_{i+1/2})^2 & \text{if } x\in[x_i,x_{i+1/2}].\end{cases}$$

Hence,

$$\begin{aligned}\left|\psi(x_i,y_j)\right| \le \frac{1}{8}\Bigg(&\int_{x_{i-1/2}}^{x_{i+1/2}}\int_{y_{j-1/2}}^{y_{j+1/2}}\left|\frac{\partial^2 w}{\partial x^2}(x,y)\right|\mathrm{d}x\,\mathrm{d}y\\ &+\int_{x_{i-1/2}}^{x_{i+1/2}}\int_{y_{j-1/2}}^{y_{j+1/2}}\left|\frac{\partial^2 w}{\partial y^2}(x,y)\right|\mathrm{d}x\,\mathrm{d}y\\ &+2\int_{x_{i-1/2}}^{x_{i+1/2}}\int_{y_{j-1/2}}^{y_{j+1/2}}\left|\frac{\partial^2 w}{\partial x\partial y}\right|\mathrm{d}x\,\mathrm{d}y\Bigg),\end{aligned}$$

so that, with $w:=cu$, we have

$$\|\psi\|^2_{L_2(\Omega^h)}\le\frac{3h^4}{64}\left(\left\|\frac{\partial^2 w}{\partial x^2}\right\|^2_{L_2(\Omega)}+\left\|\frac{\partial^2 w}{\partial y^2}\right\|^2_{L_2(\Omega)}+4\left\|\frac{\partial^2 w}{\partial x\partial y}\right\|^2_{L_2(\Omega)}\right). \quad (2.94)$$

By substituting (2.92)–(2.94) into the right-hand side of (2.83) and noting that $1/c_0=16/17$, we obtain the following result.

Theorem 2.26 *Let $f\in L_2(\Omega)$, $c\in M(W_2^2(\Omega))$, with $c(x,y)\ge 0$ for all (x,y) in $\overline{\Omega}$, and suppose that the corresponding weak solution of the boundary-value problem* (2.55), (2.56) *belongs to $W_2^3(\Omega)\cap\mathring{W}_2^1(\Omega)$. Then,*

$$\|u-U\|_{W_2^1(\Omega^h)}\le Ch^2\|u\|_{W_2^3(\Omega)}, \quad (2.95)$$

where C is a positive constant (*computable from* (2.83) *and* (2.92)–(2.94)), *independent of h and u.*

We note that, by the analogue of Lemma 1.46 on a Lipschitz domain, $M(W_2^2(\Omega))\subset W_2^2(\Omega)$, and therefore, by Sobolev embedding $c\in M(W_2^2(\Omega))$ is a continuous function with well-defined values at the mesh-points.

It can be verified by numerical experiments that the error bound (2.95) is best possible in the sense that further weakening of the regularity hypothesis on u leads to a loss of second-order convergence. Error bounds of the type (2.95), where the highest possible order of convergence is attained under the weakest hypothesis on the smoothness of the solution, are called *optimal* or *compatible with the smoothness of the solution*. Thus, for example, (2.95) is an optimal error bound for the difference scheme (2.71), (2.72), but (2.90) is not. At this point it does not concern us whether the smoothness requirements on the coefficients in the equation are the weakest possible: that issue will be addressed later, in our discussion of optimal error bounds under minimal smoothness hypotheses on the coefficients and the source term f.

We shall now explore the convergence rate of the finite difference scheme in the norm $\|\cdot\|_{W_2^1(\Omega^h)}$ under even weaker regularity hypotheses on the solution, resulting in a loss of second-order convergence established above for $u \in W_2^3(\Omega) \cap \mathring{W}_2^1(\Omega)$. Suppose, for example, that $u \in W_2^2(\Omega) \cap \mathring{W}_2^1(\Omega)$. From (2.91), by noting that

$$\left|A_i'(x)\right| \le \frac{1}{2}h, \quad x \in [x_{i-1}, x_i] \quad \text{and} \quad \left|B_j'(y)\right| \le \frac{1}{2}h, \quad y \in [y_{j-1/2}, y_{j+1/2}],$$

we have that

$$\left|\varphi_1(x_i, y_j)\right| \le \frac{1}{2h}\int_{x_{i-1}}^{x_i}\int_{y_{j-1/2}}^{y_{j+1/2}}\left|\frac{\partial^2 u}{\partial x^2}(x, y)\right| \mathrm{d}x\,\mathrm{d}y + \frac{1}{2h}\int_{x_{i-1}}^{x_i}\int_{y_{j-1/2}}^{y_{j+1/2}}\left|\frac{\partial^2 u}{\partial x \partial y}(x, y)\right| \mathrm{d}x\,\mathrm{d}y.$$

Consequently,

$$\|\varphi_1\|^2_{L_2(\Omega_x^h)} \le \frac{h^2}{2}\left(\left\|\frac{\partial^2 u}{\partial x^2}\right\|^2_{L_2(\Omega)} + \left\|\frac{\partial^2 u}{\partial x \partial y}\right\|^2_{L_2(\Omega)}\right). \tag{2.96}$$

Analogously,

$$\|\varphi_2\|^2_{L_2(\Omega_y^h)} \le \frac{h^2}{2}\left(\left\|\frac{\partial^2 u}{\partial y^2}\right\|^2_{L_2(\Omega)} + \left\|\frac{\partial^2 u}{\partial x \partial y}\right\|^2_{L_2(\Omega)}\right). \tag{2.97}$$

From (2.83), (2.96), (2.97) and (2.94), under the assumptions that $c \in M(W_2^2(\Omega))$, $c \ge 0$ on $\overline{\Omega}$ and $u \in W_2^2(\Omega) \cap \mathring{W}_2^1(\Omega)$, we deduce that:

$$\|u - U\|_{W_2^1(\Omega^h)} \le Ch\|u\|_{W_2^2(\Omega)}, \tag{2.98}$$

where C is a positive constant, independent of h and u.

Application of Function Space Interpolation When $u \in W_2^s(\Omega)$, $2 < s < 3$, an error bound can be obtained from (2.95) and (2.98) by function space interpolation. For the sake of simplicity we shall confine ourselves to Poisson's equation (i.e. $c(x, y) \equiv 0$). In that case the constant C featuring in (2.95) and (2.98) represents an absolute constant (i.e. it is independent of $c(x, y)$). Let us consider the mapping $L : u \mapsto u - U$, with U understood as a linear function of $f = -\Delta u$. Evidently, L is a linear operator. It follows from (2.95) that the operator L, considered as a linear mapping $L : W_2^3(\Omega) \to W_2^1(\Omega^h)$, is bounded and

$$\|L\|_{W_2^3(\Omega)\to W_2^1(\Omega^h)} \le Ch^2.$$

In the same way, it follows from (2.98) that the operator L, considered as a linear mapping $L : W_2^2(\Omega) \to W_2^1(\Omega^h)$, is bounded and

$$\|L\|_{W_2^2(\Omega)\to W_2^1(\Omega^h)} \le Ch.$$

By the results of Sect. 1.1.5, the operator L, considered as a linear mapping $L : (W_2^3(\Omega), W_2^2(\Omega))_{\theta,q} \to (W_2^1(\Omega^h), W_2^1(\Omega^h))_{\theta,q}$, is also bounded and, thanks to (1.8),

$$\|L\|_{(W_2^3(\Omega),W_2^2(\Omega))_{\theta,q}\to(W_2^1(\Omega^h),W_2^1(\Omega^h))_{\theta,q}} \le \left(Ch^2\right)^{1-\theta}(Ch)^\theta = Ch^{2-\theta}.$$

Furthermore,

$$\left(W_2^1(\Omega^h), W_2^1(\Omega^h)\right)_{\theta,q} = W_2^1(\Omega^h),$$
$$\left(W_2^3(\Omega), W_2^2(\Omega)\right)_{\theta,q} = W_2^{3-\theta}(\Omega).$$

Thus we obtain the following error bound:

$$\|u - U\|_{W_2^1(\Omega^h)} \le Ch^{2-\theta}\|u\|_{W_2^{3-\theta}(\Omega)}, \quad 0 < \theta < 1.$$

By writing $3 - \theta = s$ here and supplementing the resulting bounds with the ones corresponding to the limiting cases $s = 2$ and $s = 3$, we deduce that

$$\|u - U\|_{W_2^1(\Omega^h)} \le Ch^{s-1}\|u\|_{W_2^s(\Omega)}, \quad 2 \le s \le 3,$$

where C is a positive real number, independent of h and u.

In the next section we shall show how the tedious use of integral representation theorems can be avoided in the error analysis of finite difference methods by appealing to the Bramble–Hilbert lemma and its variants.

2.3 Convergence Analysis on Uniform Meshes

In the previous section we derived an optimal bound on the global error between the unique weak solution u to a homogeneous Dirichlet boundary-value problem for the generalized Poisson equation and its finite difference approximation U, under the hypothesis that $u \in W_2^s(\Omega) \cap \mathring{W}_2^1(\Omega)$, $s \in [2, 3]$. We used integral representations for $s = 2, 3$ in conjunction with function space interpolation for $s \in (2, 3)$. Here we shall consider the same problem by using a different technique; our main tool will be the Bramble–Hilbert lemma.

2.3.1 The Bramble–Hilbert Lemma

We begin by stating the Bramble–Hilbert lemma in its simplest form, in the case of integer-order Sobolev spaces (cf. [20]). We shall then illustrate its use in the error analysis of simple discretization methods and describe its generalizations to fractional-order and anisotropic Sobolev spaces. We shall also formulate a multilinear version of the Bramble–Hilbert lemma.

Theorem 2.27 (Bramble–Hilbert Lemma) *Let $\Omega \subset \mathbb{R}^n$ be a Lipschitz domain and, for a positive integer m and a real number $p \in [1, \infty]$, let η be a bounded linear functional on the Sobolev space $W_p^m(\Omega)$ such that*

$$\mathcal{P}_{m-1} \subset \mathrm{Ker}(\eta),$$

where $\mathcal{P}_{m-1}$ denotes the set of all polynomials of degree $m-1$ in n variables. Then, there exists a positive real number $C = C(m, p, n, \Omega)$ such that

$$\left|\eta(v)\right| \le C\|\eta\||v|_{W_p^m(\Omega)} \quad \forall v \in W_p^m(\Omega).$$

The proof of this result will be presented below in a more general context. First, however, we consider a series of examples that illustrate the application of Theorem 2.27.

Example 2.6 In this example we apply the Bramble–Hilbert lemma to provide a bound on the error in the numerical quadrature rule

$$\int_{-1}^{1} v(t)\,\mathrm{d}t \approx 2v(0),$$

called the *midpoint rule*. We shall assume that $v \in W_p^2(-1,1)$, $1 \le p \le \infty$. In order to estimate the error committed, let us consider the linear functional

$$\eta(v) := \int_{-1}^{1} v(t)\,\mathrm{d}t - 2v(0)$$

defined on $W_p^2(-1,1)$. Clearly, $\mathcal{P}_1 \subset \mathrm{Ker}(\eta)$ and

$$\begin{aligned}
\left|\eta(v)\right| &\le \int_{-1}^{1} \left|v(t)\right|\mathrm{d}t + 2\left|v(0)\right| \\
&= \int_{-1}^{1} \left|v(t)\right|\mathrm{d}t + \left|\int_{-1}^{1}\int_{t}^{0} v'(s)\,\mathrm{d}s\,\mathrm{d}t + \int_{-1}^{1} v(t)\,\mathrm{d}t\right| \\
&\le 2\int_{-1}^{1} \left|v(t)\right|\mathrm{d}t + 2\int_{-1}^{1} \left|v'(t)\right|\mathrm{d}t \\
&\le 2 \cdot 2^{1-\frac{1}{p}}\left(\|v\|_{L_p(-1,1)} + \|v'\|_{L_p(-1,1)}\right) \\
&\le 2 \cdot 4^{1-\frac{1}{p}}\|v\|_{W_p^1(-1,1)} \le 2 \cdot 4^{1-\frac{1}{p}}\|v\|_{W_p^2(-1,1)}.
\end{aligned}$$

From the Bramble–Hilbert lemma we deduce that there exists a positive constant $C = C(p)$ such that

$$\left|\eta(v)\right| \le C|v|_{W_p^2(-1,1)}.$$

In the next example we consider a similar analysis on the interval $[-h, h]$. Using a scaling argument we shall reduce the problem to the one considered in Example 2.6.

Example 2.7 Let us suppose that we are required to estimate the size of the error in the midpoint rule on the interval $[-h, h]$, for $h > 0$:

$$\int_{-h}^{h} u(x)\,\mathrm{d}x \approx 2hu(0).$$

To do so, we consider the linear functional

$$\eta_h(u) := \int_{-h}^{h} u(x)\,\mathrm{d}x - 2hu(0),$$

and introduce the following change of variable, in order to map $[-h, h]$ on the 'canonical interval' $[-1, 1]$:

$$x = ht, \quad t \in [-1, 1], \qquad v(t) := u(x).$$

Then, with η as in the previous example,

$$\eta_h(u) = h\eta_1(v) = h\eta(v).$$

Therefore, according to the final inequality in Example 2.6, and returning from the interval $[-1, 1]$ to $[-h, h]$,

$$\left|\eta_h(u)\right| \le Ch|v|_{W_p^2(-1,1)} = Ch \cdot h^{2-\frac{1}{p}}|u|_{W_p^2(-h,h)}.$$

In particular, for $p = 2$ we have that

$$\left|\eta_h(u)\right| \le Ch^{5/2}|u|_{W_2^2(-h,h)}.$$

Using the error bound for the midpoint rule on the interval $[-h, h]$ established in this last example by means of the Bramble–Hilbert lemma it is possible to obtain an optimal-order bound on the global error in a finite difference approximation of a two-point boundary-value problem. We shall explain how this is done. In the next section we shall then extend the technique to multiple space dimensions.

Let us consider the two-point boundary-value problem

$$-u'' = f(x), \quad x \in (0, 1),$$
$$u(0) = 0, \qquad u(1) = 0.$$

Given the *nonuniform* finite difference mesh $0 = x_0 < x_1 < \cdots < x_N = 1$ with spacing $h_i := x_i - x_{i-1}$, $i = 1, \ldots, N$, we define $\hbar_i := (h_{i+1} + h_i)/2$, $i = 1, \ldots, N - 1$, and introduce the backward and forward divided difference operators

$$D_x^- V_i := \frac{V_i - V_{i-1}}{h_i}, \qquad D_x^+ V_i := \frac{V_{i+1} - V_i}{\hbar_i},$$

and the following inner products and norms:

$$(V, W)_h := \sum_{i=1}^{N-1} \hbar_i V_i W_i, \quad \|V\|_{L_2(\Omega^h)} := (V, V)_h^{1/2},$$
$$(V, W]_h := \sum_{i=1}^{N} h_i V_i W_i, \quad \|V\|_{L_2(\Omega_+^h)} := (V, V]_h^{1/2},$$

where $\Omega^h := \{x_1, \dots, x_{N-1}\}$ and $\Omega_+^h := \{x_1, \dots, x_N\}$. Let us consider the following finite difference approximation of the two-point boundary-value problem:

$$-D_x^+ D_x^- U_i = T_h^1 f_i, \quad i = 1, \dots, N-1,$$
$$U_0 = 0, \qquad U_N = 0,$$

where $T_h^1 f$ denotes the mollification of f defined by

$$T_h^1 f_i := \frac{1}{\hbar_i} \int_{x_{i-1/2}}^{x_{i+1/2}} f(x) \mathrm{d}x, \quad i = 1, \dots, N-1.$$

In order to derive a bound on the global error $e := u - U$ at the mesh-points, we note that

$$-D_x^+ D_x^- e_i = -D_x^+ \eta_i, \quad i = 1, \dots, N-1,$$
$$e_0 = 0, \qquad e_N = 0,$$

where

$$\begin{aligned}
\eta_i &:= D_x^- u(x_i) - u'(x_{i-1/2}) \\
&= \frac{1}{2h_i} \left[\int_{-h_i}^{h_i} u'\left(x_{i-1/2} + \frac{1}{2} x \right) \mathrm{d}x - 2h_i u'(x_{i-1/2}) \right] \\
&= \frac{1}{2h_i} \eta_{h_i} \left(u'\left(x_{i-1/2} + \frac{1}{2} \cdot \right) \right), \quad i = 1, \dots, N,
\end{aligned}$$

where η_{h_i} is as in Example 2.7. We thus deduce that

$$|\eta_i| \le C h_i^{3/2} |u'|_{W_2^2(x_{i-1}, x_i)},$$

where C is a positive constant, independent of h_i. Consequently,

$$\|\eta\|_{L_2(\Omega_+^h)}^2 = \sum_{i=1}^{N} h_i |\eta_i|^2 \le C^2 \sum_{i=1}^{N} h_i h_i^3 |u'|_{W_2^2(x_{i-1}, x_i)}^2$$

$$= C^2 \sum_{i=1}^{N} h_i^4 |u|^2_{W_2^3(x_{i-1},x_i)} \le C^2 h^4 |u|^2_{W_2^3(0,1)},$$

where $h = \max_i h_i$. We complete the error analysis by showing that the quantity $\|D_x^- e\|_{L_2(\Omega_+^h)}$ can be bounded in terms of $\|\eta\|_{L_2(\Omega_+^h)}$. Indeed, by summation by parts and using the Cauchy–Schwarz inequality we obtain

$$\begin{aligned}\left\|D_x^- e\right\|^2_{L_2(\Omega_+^h)} &= \left(-D_x^+ D_x^- e, e\right)_h = \left(-D_x^+ \eta, e\right)_h \\ &= \left(\eta, D_x^- e\right]_h \le \|\eta\|_{L_2(\Omega_+^h)} \left\|D_x^- e\right\|_{L_2(\Omega_+^h)}.\end{aligned}$$

Hence,

$$\left\|D_x^- e\right\|_{L_2(\Omega_+^h)} \le \|\eta\|_{L_2(\Omega_+^h)},$$

and therefore

$$|u - U|_{W_2^1(\Omega^h)} := \left\|D_x^-(u - U)\right\|_{L_2(\Omega_+^h)} \le Ch^2 |u|_{W_2^3(0,1)},$$

where C is a positive constant, independent of h and u. We note that we did not have to impose any regularity requirements on the nonuniform mesh to prove this error bound; in the next section, we shall develop a similar analysis in two dimensions.

First, however, we formulate a generalization of the Bramble–Hilbert lemma to Sobolev spaces of any positive (not necessarily integer) order.

Theorem 2.28 *Let $\Omega \subset \mathbb{R}^n$ be a Lipschitz domain and, for real numbers $s > 0$ and $p \in [1, \infty]$, let η be a bounded linear functional on the Sobolev space $W_p^s(\Omega)$ such that, by writing $s = m + \alpha$ with m a nonnegative integer and $0 < \alpha \le 1$,*

$$\mathcal{P}_m \subset \operatorname{Ker}(\eta).$$

Then, there exists a positive real number $C = C(s, p, n, \Omega)$ such that

$$\left|\eta(v)\right| \le C\|\eta\| |v|_{W_p^s(\Omega)} \quad \forall v \in W_p^s(\Omega).$$

Proof This result is a simple consequence of Theorem 1.13 with $\mathcal{U}_0 = L_p(\Omega)$, $\mathcal{U}_1 = W_p^s(\Omega)$, $S_0(u) = \|u\|_{L_p(\Omega)}$, $S_1(u) = \|u\|_{W_p^s(\Omega)}$, $S(u) = |\eta(u)|$, by noting that, according to the Theorem 1.36, $W_p^s(\Omega)$ is compactly embedded in $L_p(\Omega)$ for any $s > 0$. □

One can apply this result to the midpoint rule to deduce, in the same manner as in the integer-order case considered earlier, that the linear functional η defined on $W_p^s(-1, 1)$, $1/p < s \le 2$, $1 \le p \le \infty$, by

$$\eta(v) = \int_{-1}^{1} v(t)\,\mathrm{d}t - 2v(0)$$

satisfies the bound

$$\big|\eta(v)\big| \le C|v|_{W_p^s(-1,1)},$$

for any v in $W_p^s(\Omega)$, $1/p < s \le 2$, $1 \le p \le \infty$. Thus in particular, with $p = 2$, we obtain the following bound on the global error in the finite difference approximation of the two-point boundary-value problem considered:

$$|u - U|_{W_2^1(\Omega^h)} \le Ch^s |u'|_{W_2^s(0,1)},$$

where $h = \max_i h_i$, provided that $u \in W_2^{s+1}(0,1)$ (whereby $u' \in W_2^s(0,1)$), $1/2 < s \le 2$. In the next section we extend this result to Poisson's equation on the unit square. First we shall however formulate a generalization of the Bramble–Hilbert lemma to anisotropic Sobolev spaces of the type $W_p^A(\Omega)$.

Let $A \subset \mathbb{R}_+^n$ be a regular set of nonnegative real multi-indices (cf. Sect. 1.5). We denote the convex hull in $\mathbb{R}^n$ of the set A by $\kappa(A)$. Let $\partial_0\kappa(A)$ be the part of the boundary of $\kappa(A)$ that has empty intersection with the co-ordinate hyperplanes, and let $A_\partial = A \cap \overline{\partial_0\kappa(A)}$. Let B be a nonempty subset of A_∂ such that $B \cup \{0\}$ is a regular set of multi-indices, and define

$$\nu(B) := \big\{\beta \in \mathbb{N}_+^n : \partial^{\lfloor\alpha\rfloor} x^\beta \equiv 0 \ \ \forall \alpha \in B\big\}.$$

Let $\mathcal{P}_B$ denote the set of all polynomials in n variables of the form

$$P(x) = \sum_{\alpha \in \nu(B)} p_\alpha x^\alpha.$$

Theorem 2.29 *Suppose that Ω is a Lipschitz domain in $\mathbb{R}^n$ and let the sets A and B of real nonnegative multi-indices satisfy the conditions formulated in the previous paragraph. Then, there exists a positive real number $C = C(A, B, p, n, \Omega)$ such that*

$$\inf_{P \in \mathcal{P}_B} \|v - P\|_{W_p^A(\Omega)} \le C \sum_{\alpha \in B} |v|_{\alpha,p} \quad \forall v \in W_p^A(\Omega).$$

Moreover, if η is a bounded linear functional on $W_p^A(\Omega)$, with norm $\|\eta\|$, such that

$$\mathcal{P}_B \subset \mathrm{Ker}(\eta),$$

then

$$\big|\eta(v)\big| \le C\|\eta\| \sum_{\alpha \in B} |v|_{\alpha,p} \quad \forall v \in W_p^A(\Omega).$$

Proof This result is a simple consequence of Theorem 1.13 with $\mathcal{U}_0 = L_p(\Omega)$, $\mathcal{U}_1 = W_p^A(\Omega)$, $S_0(u) = \|u\|_{L_p(\Omega)}$,

$$S_1(u) = \|u\|_{L_p(\Omega)} + \sum_{\alpha \in B} |v|_{\alpha,p},$$

and $S(u) = |\eta(u)|$, and noting that $W_p^A(\Omega)$, equipped with the norm $S_1(\cdot)$ is compactly embedded in $L_p(\Omega)$. $\square$

As a further generalization, we state the following multilinear version of the Bramble–Hilbert lemma: this will be used extensively in the bilinear case throughout the book.

Lemma 2.30 *Suppose that A_k, B_k and Ω_k satisfy the same conditions in $\mathbb{R}^{n_k}$, $k = 1, \dots, m$, as A, B and Ω did in the previous theorem. Let $(v_1, \dots, v_m) \mapsto \eta(v_1, \dots, v_m)$ be a bounded multilinear functional on the function space*

$$W_{p_1}^{A_1}(\Omega_1) \times \cdots \times W_{p_m}^{A_m}(\Omega_m),$$

which vanishes whenever one of its entries has the form $v_k = x^\alpha$, $x \in \Omega_k$, $\alpha \in \nu(B_k)$. Then, there exists a real number

$$C = C(A_1, B_1, p_1, \Omega_1, n_1, \dots, A_m, B_m, p_m, \Omega_m, n_m)$$

such that

$$\left|\eta(v_1, \dots, v_m)\right| \le C \|\eta\| \prod_{k=1}^{m} \sum_{\alpha \in B_k} |v_k|_{\alpha, p_k}$$

for every $(v_1, \dots, v_m)$ in $W_{p_1}^{A_1}(\Omega_1) \times \cdots \times W_{p_m}^{A_m}(\Omega_m)$.

When $m = 2$, this result will be referred to as the *bilinear version of the Bramble–Hilbert lemma*. In the case of standard, integer-order isotropic Sobolev spaces, the bilinear version of the Bramble–Hilbert lemma can be found in Ciarlet [26], Theorem 4.2.5. In the general case the proof is analogous, and is once again a simple consequence of Theorem 2.29.

2.3.2 Optimal Error Bounds on Uniform Meshes

In this section we shall use the Bramble–Hilbert lemma to derive an optimal bound on the global error of the finite difference (or, more precisely, finite volume) approximation (2.71), (2.72) of the homogeneous Dirichlet boundary-value problem (2.55), (2.56) on a uniform mesh of size h; in the next section we shall extend this analysis to nonuniform meshes. Thus, we consider the following finite difference scheme:

$$-\left(D_x^+ D_x^- + D_y^+ D_y^-\right) U_{ij} + c(x_i, y_j) U_{ij} = \left(T_h^{11} f\right)_{ij}, \quad (x_i, y_j) \in \Omega^h, \tag{2.99}$$

$$U = 0 \quad \text{on } \Gamma^h. \tag{2.100}$$

Let $e := u - U$ denote the global error of the scheme; then, according to Lemma 2.24,

$$\|e\|_{W_2^1(\Omega^h)} \le \frac{1}{c_0}\left(\|\varphi_1\|^2_{L_2(\Omega^h_x)} + \|\varphi_2\|^2_{L_2(\Omega^h_y)} + \|\psi\|^2_{L_2(\Omega^h)}\right)^{1/2}, \qquad (2.101)$$

where φ_1, φ_2, and ψ are defined by

$$\varphi_1(x_i, y_j) := \frac{1}{h}\int_{y_{j-1/2}}^{y_{j+1/2}} \frac{\partial u}{\partial x}(x_{i-1/2}, y)\,\mathrm{d}y - D_x^- u(x_i, y_j),$$

$$\varphi_2(x_i, y_j) := \frac{1}{h}\int_{x_{i-1/2}}^{x_{i+1/2}} \frac{\partial u}{\partial y}(x, y_{j-1/2})\,\mathrm{d}x - D_y^- u(x_i, y_j),$$

$$\psi(x_i, y_j) := (cu)(x_i, y_j) - \frac{1}{h^2}\int_{x_{i-1/2}}^{x_{i+1/2}}\int_{y_{j-1/2}}^{y_{j+1/2}} (cu)(x, y)\,\mathrm{d}x\,\mathrm{d}y,$$

with $x_{i\pm1/2} = x_i \pm h/2$ and $y_{j\pm1/2} = y_j \pm h/2$.

We shall use the Bramble–Hilbert lemma to estimate φ_1, φ_2 and ψ in terms of appropriate powers of the discretization parameter h and suitable Sobolev semi-norms of the analytical solution u. We begin by considering φ_1. Let us introduce the change of variables

$$x = x_{i-1/2} + \tilde{x}h, \quad -\frac{1}{2} \le \tilde{x} \le \frac{1}{2}; \qquad y = y_j + \tilde{y}h, \quad -\frac{1}{2} \le \tilde{y} \le \frac{1}{2},$$

and define $\tilde{v}(\tilde{x}, \tilde{y}) := h\frac{\partial u}{\partial x}(x, y)$. Then,

$$\varphi_1(x_i, y_j) = \frac{1}{h}\tilde{\varphi}_1(\tilde{v}),$$

where

$$\tilde{\varphi}_1(\tilde{v}) := \int_{-1/2}^{1/2}\int_{-1/2}^{1/2}\left[\tilde{v}(0, \tilde{y}) - \tilde{v}(\tilde{x}, 0)\right]\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{y}.$$

Thanks to the trace theorem (Theorem 1.42),

$$\left|\tilde{\varphi}_1(\tilde{v})\right| \le C_s\|\tilde{v}\|_{W_2^s(\tilde{K})}, \quad s > 1/2,$$

where

$$\tilde{K} := \left(-\frac{1}{2}, \frac{1}{2}\right) \times \left(-\frac{1}{2}, \frac{1}{2}\right),$$

and $C_s = C(s)$ is a positive constant. Thus $\tilde{\varphi}_1$ is a bounded linear functional (of the argument $\tilde{v}$) on $W_2^s(\tilde{K})$ for $s > 1/2$.

Moreover, $\tilde{\varphi}_1 = 0$ when $\tilde{v}(\tilde{x}, \tilde{y}) = \tilde{x}^k\tilde{y}^l$, $k, l \in \{0, 1\}$. According to Theorem 2.28, there exists a positive constant $C = C(s)$ such that

$$\left|\tilde{\varphi}_1(\tilde{v})\right| \le C|\tilde{v}|_{W_2^s(\tilde{K})}, \quad 1/2 < s \le 2.$$

Hence, by defining

$$K_{ij} := (x_{i-1/2}, x_{i+1/2}) \times (y_{j-1/2}, y_{j+1/2})$$

and returning from $\tilde{x}$ and $\tilde{y}$ to the original variables x and y, we deduce that

$$\left|\tilde{\varphi}_1(\tilde{v})\right| \le Ch^s \left|\frac{\partial u}{\partial x}\right|_{W_2^s(K_{ij})}, \quad 1/2 < s \le 2,$$

so that

$$\left|\varphi_1(x_i, y_j)\right| \le Ch^{s-1} \left|\frac{\partial u}{\partial x}\right|_{W_2^s(K_{ij})}, \quad 1/2 < s \le 2.$$

By noting that the Sobolev seminorm on the unit square is *superadditive* on the family $\{K_{ij}\}$ of mutually disjoint Lebesgue-measurable subsets K_{ij} of Ω, i.e. for $w \in W_2^s(\Omega)$ one has

$$\left(\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} |w|^2_{W_2^s(K_{ij})}\right)^{1/2} \le |w|_{W_2^s(\cup_{i,j=1}^{N-1} K_{ij})},$$

it follows with $w = \partial u/\partial x$ that

$$\|\varphi_1\|_{L_2(\Omega_x^h)} \le Ch^s \left|\frac{\partial u}{\partial x}\right|_{W_2^s(\Omega)}, \quad 1/2 < s \le 2, \tag{2.102}$$

where C is a positive constant, dependent only on s. Analogously,

$$\|\varphi_2\|_{L_2(\Omega_y^h)} \le Ch^s \left|\frac{\partial u}{\partial y}\right|_{W_2^s(\Omega)}, \quad 1/2 < s \le 2. \tag{2.103}$$

To complete the error analysis it remains to estimate $\psi(x_i, y_j)$. For this purpose we shall write $w := cu$ and note that

$$\psi(x_i, y_j) = w(x_i, y_j) - \frac{1}{h^2}\int_{x_{i-1/2}}^{x_{i+1/2}}\int_{y_{j-1/2}}^{y_{j+1/2}} w(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

Let us also consider the following change of variables:

$$x = x_i + \tilde{x}h, \quad -\frac{1}{2} \le \tilde{x} \le \frac{1}{2}; \qquad y = y_j + \tilde{y}h, \quad -\frac{1}{2} \le \tilde{y} \le \frac{1}{2},$$

and define $\tilde{w}(\tilde{x}, \tilde{y}) := w(x, y)$. Then,

$$\psi(x_i, y_j) = \tilde{\psi}(\tilde{w}),$$

where

$$\tilde{\psi}(\tilde{w}) := \tilde{w}(0, 0) - \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} \tilde{w}(\tilde{x}, \tilde{y})\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{y}.$$

By Sobolev's embedding theorem $\tilde{\psi}$ is a bounded linear functional (of $\tilde{w}$) on $W_2^s(\tilde{K})$ for $s > 1$, where, as before, $\tilde{K} := (-1/2, 1/2) \times (-1/2, 1/2)$. Furthermore, $\tilde{\psi}(\tilde{w}) = 0$ whenever $\tilde{w} = \tilde{x}^k \tilde{y}^l$ with $k, l \in \{0, 1\}$. Thus, by the Bramble–Hilbert lemma,

$$\left|\tilde{\psi}(\tilde{w})\right| \leq C|\tilde{w}|_{W_2^s(\tilde{K})}, \quad 1 < s \leq 2,$$

and consequently, after returning from the $(\tilde{x}, \tilde{y})$ co-ordinate system to the original variables x and y, we obtain the bound

$$\left|\psi(x_i, y_j)\right| \leq Ch^{s-1}|w|_{W_2^s(K_{ij})}, \quad 1 < s \leq 2,$$

and finally, after squaring and summing over $i, j = 1, \dots, N-1$,

$$\|\psi\|_{L_2(\Omega^h)} \leq Ch^s|cu|_{W_2^s(\Omega)}, \quad 1 < s \leq 2. \tag{2.104}$$

Thus, by assuming that the weak solution $u \in W_2^s(\Omega) \cap \mathring{W}_2^1(\Omega)$ and that $c \in M(W_2^s(\Omega))$, for $1 < s \leq 2$, after substituting (2.102), (2.103) and (2.104) into (2.101), we arrive at the following bound on the global error:

$$\|u - U\|_{W_2^1(\Omega^h)} \leq Ch^s\left(\left|\frac{\partial u}{\partial x}\right|_{W_2^s(\Omega)} + \left|\frac{\partial u}{\partial y}\right|_{W_2^s(\Omega)} + \|c\|_{M(W_2^s(\Omega))}\|u\|_{W_2^s(\Omega)}\right),$$

where C is a positive constant depending on s, but independent of h; or, more crudely, after bounding $|\partial u/\partial x|_{W_2^s(\Omega)} + |\partial u/\partial y|_{W_2^s(\Omega)}$ by $\|u\|_{W_2^{s+1}(\Omega)}$, and writing $s-1$ instead s, we obtain

$$\|u - U\|_{W_2^1(\Omega^h)} \leq Ch^{s-1}\|u\|_{W_2^s(\Omega)}, \quad 2 < s \leq 3.$$

This should be compared with the error bound derived in the previous section using integral representations based on the Newton–Leibniz formula for $s = 2$ and $s = 3$ and by function space interpolation for $2 < s < 3$.

2.4 Convergence Analysis on Nonuniform Meshes

Our objective in this section is to develop the error analysis of finite difference (or, more precisely, finite volume) approximations on nonuniform meshes for the model Poisson equation with homogeneous Dirichlet boundary condition:

$$-\Delta u = f \quad \text{in } \Omega, \tag{2.105}$$

$$u = 0 \quad \text{on } \Gamma = \partial\Omega, \tag{2.106}$$

where $\Omega := (0, 1) \times (0, 1)$. When $f \in W_2^{-1}(\Omega)$, this boundary-value problem has a unique weak solution u in $\mathring{W}_2^1(\Omega)$; furthermore, if $f \in W_2^s(\Omega)$ then u belongs to $W_2^{s+2}(\Omega)$, $-1 \leq s < 1$, $s \neq \pm 1/2$ (see, Theorem 2.8).

As has already been indicated earlier, the key idea behind the construction of a finite volume method for (2.105), (2.106) is to make use of the divergence form of the differential operator $\Delta = \nabla \cdot \nabla$ appearing in the equation $-\Delta u = f$ by integrating both sides over mutually disjoint 'cells' $K_{ij} \subset \Omega$, and use the divergence theorem to convert integrals over the cells K_{ij} into contour integrals along their boundaries, which are then discretized by means of numerical quadrature rules. This construction gives rise to a finite difference scheme whose right-hand side involves the integral average of f over individual cells, the particular form of the difference scheme being dependent on the shapes of the cells and the numerical quadrature formula used. For example, if Ω has been partitioned by a uniform square mesh of mesh-size h, then the resulting scheme coincides with (2.71), (2.72) (with $c \equiv 0$).

2.4.1 Cartesian-Product Nonuniform Meshes

We begin by considering Cartesian-product nonuniform meshes. For the purposes of the error analysis it is helpful to reformulate the finite volume scheme as a Petrov–Galerkin finite element method based on bilinear or piecewise linear trial functions on the underlying mesh and piecewise constant test functions on the dual 'box mesh'. We shall prove that, as in the case of uniform meshes considered in the previous section, the scheme is stable in the discrete W_2^1 norm. This stability result will then, similarly to the arguments in the previous section, lead to an optimal-order error bound in the discrete W_2^1 norm under minimal smoothness requirements on the exact solution and without any additional assumptions on the spacing of the mesh. In particular, the mesh is not required to be quasi-uniform (in a sense that will be made precise). If quasi-uniformity is assumed, then an additional error bound holds, in the discrete maximum norm. In the next section similar results will be derived for a general one-parameter family of schemes.

The problem (2.105), (2.106) is approximated on the nonuniform mesh $\overline{\Omega}^h$, which is the Cartesian product of the one-dimensional meshes

$$\{x_i,\ i = 0, \ldots, M : x_0 = 0,\ x_i - x_{i-1} = h_i, x_M = 1\},$$
$$\{y_j,\ j = 0, \ldots, N : y_0 = 0,\ y_j - y_{j-1} = k_j, y_N = 1\}.$$

We then define

$$\Omega^h := \Omega \cap \overline{\Omega}^h, \qquad \Gamma^h := \Gamma \cap \overline{\Omega}^h,$$
$$\Omega_x^h := \overline{\Omega}^h \cap \big((0,1] \times (0,1)\big), \qquad \Omega_y^h := \overline{\Omega}^h \cap \big((0,1) \times (0,1]\big),$$
$$\Gamma_x^h := \overline{\Omega}^h \cap \big(\{0,1\} \times (0,1)\big), \qquad \Gamma_y^h := \overline{\Omega}^h \cap \big((0,1) \times \{0,1\}\big).$$

To each mesh-point (x_i, y_j) in Ω^h we assign a cell

$$K_{ij} := (x_{i-1/2}, x_{i+1/2}) \times (y_{j-1/2}, y_{j+1/2}),$$

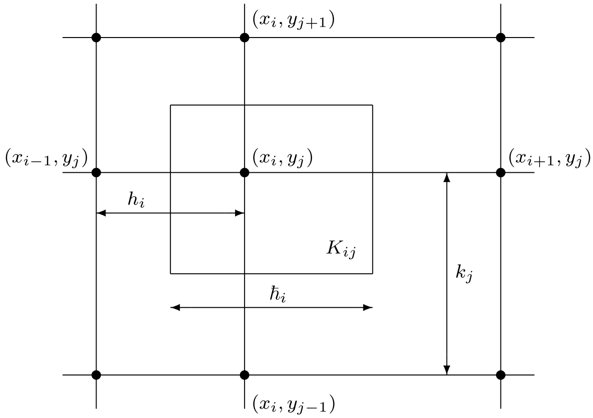

Fig. 2.4 Section of the Cartesian-product nonuniform mesh $\overline{\Omega}^h$, showing nine mesh-points and the cell K_{ij} associated with the mesh-point (x_i, y_j)

as shown in Fig. 2.4, where

$$x_{i-1/2} := x_i - \frac{1}{2}h_i, \qquad x_{i+1/2} := x_i + \frac{1}{2}h_{i+1},$$
$$y_{j-1/2} := y_j - \frac{1}{2}k_j, \qquad y_{j+1/2} := y_j + \frac{1}{2}k_{j+1},$$

and we denote the edge-lengths of the cell K_{ij} by

$$\hbar_i := \frac{1}{2}(h_i + h_{i+1}) \quad \text{and} \quad \bar{k}_j := \frac{1}{2}(k_j + k_{j+1}).$$

A simple calculation based on the definition of the fractional-order Sobolev norm shows that χ_{ij}, the characteristic function of the set $(-h_{i+1}/2, h_i/2) \times (-k_{j+1}/2, k_j/2)$, belongs to $W_2^\tau(\mathbb{R}^2)$ for all $\tau < 1/2$. Assuming that f belongs to $W_2^s(\Omega)$ for some $s > -1/2$, and extending f from Ω onto $\mathbb{R}^2$ by preserving its Sobolev class, we deduce from Theorem 1.69 that the convolution $\chi_{ij} * f$ is a continuous function on $\mathbb{R}^2$ (whose values on Ω^h are independent of the particular form of the extension). Convolution of (2.105) with χ_{ij} then yields

$$-\frac{1}{\text{meas } K_{ij}} \int_{\partial K_{ij}} \frac{\partial u}{\partial \nu} \, \mathrm{d}s = \frac{1}{\text{meas } K_{ij}} (\chi_{ij} * f)(x_i, y_j), \tag{2.107}$$

where ν denotes the unit outward normal vector to ∂K_{ij}.

We remark that if f is a locally integrable function on Ω then, similarly as in the case of uniform meshes considered earlier, the right-hand side of (2.107) is simply

$$\left(T_h^{11} f\right)_{ij} = \frac{1}{\hbar_i \bar{k}_j} \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{j-1/2}}^{y_{j+1/2}} f(x, y) \, \mathrm{d}x \, \mathrm{d}y.$$

Let $\mathcal{S}^h$ signify the set of all real-valued continuous piecewise bilinear functions defined on the rectangular partition of $\overline{\Omega}$ induced by $\overline{\Omega}^h$, and let $\mathcal{S}_0^h$ be the subset of

$\mathcal{S}^h$ consisting of those functions that vanish on Γ. Motivated by the form of (2.107), we define the finite volume approximation of u as $U \in \mathcal{S}_0^h$ satisfying

$$-\frac{1}{\hbar_i \hbar_j} \int_{\partial K_{ij}} \frac{\partial U}{\partial \nu} \,\mathrm{d}s = \frac{1}{\hbar_i \hbar_j} (\chi_{ij} * f)(x_i, y_j) \quad \text{for } (x_i, y_j) \in \Omega^h. \tag{2.108}$$

First, we shall prove that this method is stable by proceeding in the same way as in the case of uniform meshes considered in the previous section. To this end, we shall rewrite (2.108) as a finite difference scheme on $\overline{\Omega}^h$ by using the *averaging operator* μ_x defined by

$$\mu_x V_{ij} := \frac{1}{8\hbar_i} (h_i V_{i-1,j} + 6\hbar_i V_{ij} + h_{i+1} V_{i+1,j}) \tag{2.109}$$

and the divided differences

$$D_x^- V_{ij} := \frac{V_{ij} - V_{i-1,j}}{h_i} \quad \text{and} \quad D_x^+ V_{ij} := \frac{V_{i+1,j} - V_{ij}}{\hbar_i},$$

with analogous definitions for μ_y, D_y^- and D_y^+. With these notational conventions,

$$-\int_{\partial K_{ij}} \frac{\partial U}{\partial \nu} \,\mathrm{d}s = -\hbar_i \hbar_j \big(D_x^+ D_x^- \mu_y + D_y^+ D_y^- \mu_x\big) U_{ij}. \tag{2.110}$$

By inserting (2.110) into (2.108), the finite volume method (2.108) can be restated as the finite difference scheme

$$-\big(D_x^+ D_x^- \mu_y + D_y^+ D_y^- \mu_x\big) U = T_h^{11} f \quad \text{in } \Omega^h, \tag{2.111}$$

$$U = 0 \quad \text{on } \Gamma^h, \tag{2.112}$$

where

$$\big(T_h^{11} f\big)_{ij} := \frac{1}{\hbar_i \hbar_j} (\chi_{ij} * f)(x_i, y_j).$$

We begin the analysis of the scheme (2.111), (2.112) by investigating its stability in the discrete W_2^1 norm, $\|\cdot\|_{W_2^1(\Omega^h)}$, defined by

$$\|V\|_{W_2^1(\Omega^h)} := \big(\|V\|_{L_2(\Omega^h)}^2 + |V|_{W_2^1(\Omega^h)}^2\big)^{1/2},$$

where $\|\cdot\|_{L_2(\Omega^h)}$ is the discrete L_2 norm on the linear space of real-valued mesh-functions defined on Ω^h:

$$\|V\|_{L_2(\Omega^h)} := (V, V)_h^{1/2}, \qquad (V, W)_h := \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \hbar_i \hbar_j V_{ij} W_{ij},$$

and $|\cdot|_{W_2^1(\Omega^h)}$ is the discrete W_2^1 seminorm defined by

$$|V|_{W_2^1(\Omega^h)} := \big(\|D_x^- V\|^2_{L_2(\Omega_x^h)} + \|D_y^- V\|^2_{L_2(\Omega_y^h)}\big)^{1/2},$$

with

$$\|V\|^2_{L_2(\Omega_x^h)} := (V, V]_x^2, \qquad (V, W]_x := \sum_{i=1}^{M}\sum_{j=1}^{N-1} h_i \hbar k_j V_{ij} W_{ij},$$

$$\|V\|^2_{L_2(\Omega_y^h)} := (V, V]_y^2, \qquad (V, W]_y := \sum_{i=1}^{M-1}\sum_{j=1}^{N} \hbar_i k_j V_{ij} W_{ij}.$$

The associated discrete W_2^{-1} norm is then defined by

$$\|V\|_{W_2^{-1}(\Omega^h)} := \sup_{W\in\mathcal{S}_0^h\setminus\{0\}} \frac{|(V, W)_h|}{\|W\|_{W_2^1(\Omega^h)}}.$$

Lemma 2.31 *Suppose that V is a mesh-function defined on $\overline{\Omega}_h$.*

(a) *If $V = 0$ on Γ_x^h, then*

$$(\mu_x V, V]_y \geq \frac{1}{2}\|V\|^2_{L_2(\Omega_y^h)}. \tag{2.113}$$

(b) *If $V = 0$ on Γ_y^h, then*

$$(\mu_y V, V]_x \geq \frac{1}{2}\|V\|^2_{L_2(\Omega_x^h)}. \tag{2.114}$$

Proof We shall only prove inequality (2.113), the proof of (2.114) being analogous. Let us assume for a moment that j is fixed, $1 \leq j \leq N$. Then,

$$\begin{aligned}\sum_{i=1}^{M-1} \hbar_i (\mu_x V_{ij}) V_{ij} &= \frac{1}{8}\sum_{i=1}^{M-1}\big(h_i V_{i-1,j} V_{ij} + 6\hbar_i V_{ij}^2 + h_{i+1} V_{i+1,j} V_{ij}\big)\\ &\geq \frac{1}{8}\left(\sum_{i=1}^{M-1} 5\hbar_i V_{ij}^2 - \frac{1}{2}\sum_{i=2}^{M} h_i V_{ij}^2 - \frac{1}{2}\sum_{i=0}^{M-2} h_{i+1} V_{ij}^2\right)\\ &\geq \frac{1}{2}\sum_{i=1}^{M-1}\hbar_i V_{ij}^2.\end{aligned}$$

We then multiply this by k_j and sum through the index $j \in \{1, \ldots, N\}$ to deduce the desired inequality. □

We shall also require the following discrete analogue of the Friedrichs inequality on Cartesian-product nonuniform meshes.

Lemma 2.32 *Suppose that V is a mesh-function defined on $\overline{\Omega}_h$ such that $V = 0$ on Γ_h. Then,*

$$\|V\|^2_{W^1_2(\Omega^h)} \le \frac{3}{2}|V|^2_{W^1_2(\Omega^h)}. \tag{2.115}$$

Proof Let V be a mesh-function defined on $\overline{\Omega}_h$ such that $V = 0$ on Γ_h. Then, the expression

$$\|V\|^2_{L_2(\Omega^h)} = \sum_{i=1}^{M-1}\sum_{j=1}^{N-1} \hbar_i \bar{k}_j V_{ij}^2$$

can be bounded as follows:

$$\begin{aligned}
\|V\|^2_{L_2(\Omega^h)} &= \frac{1}{2}\sum_{i=1}^{M-1}\sum_{j=1}^{N-1} \hbar_i \bar{k}_j V_{ij}^2 + \frac{1}{2}\sum_{i=1}^{M-1}\sum_{j=1}^{N-1} \hbar_i \bar{k}_j V_{ij}^2 \\
&= \frac{1}{2}\left(\sum_{i=1}^{M-1}\sum_{j=1}^{N-1} \hbar_i \bar{k}_j \left|\sum_{m=1}^{i} h_m D_x^- V_{mj}\right|^2 + \sum_{i=1}^{M-1}\sum_{j=1}^{N-1} \hbar_i \bar{k}_j \left|\sum_{n=1}^{j} k_n D_y^- V_{in}\right|^2\right) \\
&\le \frac{1}{2}\sum_{i=1}^{M-1}\sum_{j=1}^{N-1} \hbar_i \bar{k}_j \left(\sum_{m=1}^{i} h_m\right)\left(\sum_{m=1}^{i} h_m \left|D_x^- V_{mj}\right|^2\right) \\
&\quad + \frac{1}{2}\sum_{i=1}^{M-1}\sum_{j=1}^{N-1} \hbar_i \bar{k}_j \left(\sum_{n=1}^{j} k_n\right)\left(\sum_{n=1}^{j} k_n \left|D_y^- V_{in}\right|^2\right) \\
&\le \frac{1}{2}\left(\sum_{m=1}^{M}\sum_{j=1}^{N-1} h_m \bar{k}_j \left|D_x^- V_{mj}\right|^2\right)\left(\sum_{i=1}^{M-1} \hbar_i \sum_{m=1}^{i} h_m\right) \\
&\quad + \frac{1}{2}\left(\sum_{i=1}^{M-1}\sum_{n=1}^{N} \hbar_i k_n \left|D_y^- V_{in}\right|^2\right)\left(\sum_{j=1}^{N-1} \bar{k}_j \sum_{n=1}^{j} k_n\right) \\
&\le \frac{1}{2}\left(\left\|D_x^- V\right\|^2_{L_2(\Omega_x^h)} + \left\|D_x^- V\right\|^2_{L_2(\Omega_y^h)}\right) = \frac{1}{2}|V|^2_{W_2^1(\Omega^h)}.
\end{aligned}$$

Adding $|V|^2_{W^1_2(\Omega^h)}$ to both sides completes the proof of the lemma. □

By using this discrete Friedrichs inequality we shall now prove that the finite difference scheme is stable; the key to the proof is the following result.

Theorem 2.33 *Let $L^h V := -(D_x^+ D_x^- \mu_y + D_y^+ D_y^- \mu_x)V$. Then,*

$$\|V\|_{W_2^1(\Omega^h)} \le 3\big\|L^h V\big\|_{W_2^{-1}(\Omega^h)} \tag{2.116}$$

for any mesh-function V defined on $\overline{\Omega}^h$ and such that $V = 0$ on Γ^h.

Proof By taking the $(\cdot,\cdot)_h$ inner product of $L^h V$ with V we obtain

$$\big(-\big(D_x^+ D_x^- \mu_y\big)V, V\big)_h + \big(-\big(D_y^+ D_y^- \mu_x\big)V, V\big)_h = \big(L^h V, V\big)_h.$$

By performing summations by parts in the two terms on the left-hand side we get

$$\big(D_x^- \mu_y V, D_x^- V\big]_x + \big(D_y^- \mu_x V, D_y^- V\big]_y = \big(L^h V, V\big)_h.$$

Since D_x^- commutes with μ_y and D_y^- commutes with μ_x, we can apply (2.113) and (2.114) to obtain

$$\frac{1}{2}\big(\big\|D_x^- V\big\|^2_{L_2(\Omega_x^h)} + \big\|D_y^- V\big\|^2_{L_2(\Omega_y^h)}\big) \le \big(L^h V, V\big)_h.$$

By recalling (2.115) and the definition of $\|\cdot\|_{W_2^{-1}(\Omega^h)}$ we get (2.116). $\square$

Theorem 2.33 now implies the stability of the scheme.

Theorem 2.34 *For any $f \in W_2^s(\Omega)$, $s > -1/2$, the scheme* (2.108) (*or, equivalently*, (2.111), (2.112)) *has a unique solution U. Moreover,*

$$\|U\|_{W_2^1(\Omega^h)} \le 3\big\|T_h^{11} f\big\|_{W_2^{-1}(\Omega^h)}.$$

Having proved stability, we are now ready to embark on the error analysis of the scheme. We shall derive an optimal-order error bound for the finite difference method (2.111), (2.112), which can also be seen as a superconvergence result for the finite volume method (2.108) considered as a Petrov–Galerkin finite element method, on a family of Cartesian-product nonuniform meshes. By *superconvergence* we mean that $\mathcal{O}(h^2)$ convergence of the error between u and its continuous piecewise bilinear approximation U is observed in the *discrete* W_2^1 norm while only $\mathcal{O}(h)$ convergence will be seen if $u - U$ is measured in the norm of the Sobolev space $W_2^1(\Omega)$. The result will be shown to hold without any additional assumptions on the spacing of the mesh: in particular the mesh is not required to be quasi-uniform (the definition of quasi-uniform mesh will be given in the statement of Theorem 2.38).

Theorem 2.35 *Suppose that $u \in W_2^{s+1}(\Omega) \cap \mathring{W}_2^1(\Omega)$, $1/2 < s \le 2$. Then,*

$$\|u - U\|_{W_2^1(\Omega^h)} \le C h^s |u|_{W_2^{s+1}(\Omega)}, \tag{2.117}$$

where $h = \max_{i,j}(h_i, k_j)$, and $C = C(s)$ is a positive constant independent of u and the discretization parameters.

In the proof of Theorem 2.35 we shall make use of anisotropic Sobolev spaces on rectangular subdomains of $\mathbb{R}^2$. For $\omega = (a, b) \times (c, d)$ and a pair (r, s) of nonnegative real numbers, we denote by $W_2^{r,s}(\omega)$ the anisotropic Sobolev space consisting of all functions $u \in L_2(\omega)$ such that

$$|u|_{W_2^{r,0}(\omega)} := \left(\int_c^d \left|u(\cdot, y)\right|^2_{W_2^r(a,b)} \,\mathrm{d}y\right)^{1/2} < \infty,$$

$$|u|_{W_2^{0,s}(\omega)} := \left(\int_a^b \left|u(x, \cdot)\right|^2_{W_2^s(c,d)} \,\mathrm{d}x\right)^{1/2} < \infty.$$

The linear space $W_2^{r,s}(\omega)$ is a Banach space equipped with the norm

$$\|u\|_{W_2^{r,s}(\omega)} := \left(\|u\|^2_{L_2(\omega)} + |u|^2_{W_2^{r,0}(\omega)} + |u|^2_{W_2^{0,s}(\omega)}\right)^{1/2}.$$

For $s \geq 0$, $W_2^{s,s}(\omega)$ coincides with the standard (isotropic) Sobolev space $W_2^s(\omega)$, and the norm $\|\cdot\|_{W_2^{s,s}(\omega)}$ is equivalent to the Sobolev norm $\|\cdot\|_{W_2^s(\omega)}$ (cf. Sect. 18 of Besov, Il'in and Nikol'skiĭ [13]).

Proof of Theorem 2.35 Let us define the global error as $e := u - U$. Then, by applying the difference operator L^h defined in Theorem 2.33 to e and noting the definition of the finite difference scheme, we deduce that

$$L^h e = \left(T_h^{11} \frac{\partial^2 u}{\partial x^2} - D_x^+ D_x^- \mu_y u\right) + \left(T_h^{11} \frac{\partial^2 u}{\partial y^2} - D_y^+ D_y^- \mu_x u\right).$$

However,

$$\begin{aligned}\left(T_h^{11} \frac{\partial^2 u}{\partial x^2}\right)_{ij} &= \frac{1}{\hbar_i \bar{k}_j} \int_{y_{j-1/2}}^{y_{j+1/2}} \left[\frac{\partial u}{\partial x}(x_{i+1/2}, y) - \frac{\partial u}{\partial x}(x_{i-1/2}, y)\right] \mathrm{d}y \\ &= D_x^+ \left(T_-^{01} \frac{\partial u}{\partial x}\right)_{ij},\end{aligned}$$

where

$$\left(T_-^{01} w\right)_{ij} = \frac{1}{\bar{k}_j} \int_{y_{j-1/2}}^{y_{j+1/2}} w(x_{i-1/2}, y)\, \mathrm{d}y.$$

Consequently,

$$\begin{aligned} L^h e &= D_x^+ \eta_1 + D_y^+ \eta_2 \quad \text{in } \Omega^h, \\ e &= 0 \quad \text{on } \Gamma^h, \end{aligned} \tag{2.118}$$

where

$$\eta_1 := T_-^{01} \frac{\partial u}{\partial x} - D_x^- \mu_y u, \qquad \eta_2 := T_-^{10} \frac{\partial u}{\partial y} - D_y^- \mu_x u,$$

and T_-^{10} is defined analogously to T_-^{01} above. By applying Theorem 2.33 to the finite difference equations (2.118) we have that

$$\|e\|_{W_2^1(\Omega^h)} \le 3\big\|D_x^+\eta_1 + D_y^+\eta_2\big\|_{W_2^{-1}(\Omega^h)}.$$

It remains to bound the right-hand side of this inequality. We observe to this end that, for any mesh-function V defined on $\overline{\Omega}^h$ and vanishing on Γ^h,

$$-\big(D_x^+\eta_1 + D_y^+\eta_2, V\big)_h = \big(\eta_1, D_x^- V\big]_x + \big(\eta_2, D_y^- V\big]_y.$$

By noting the definition of the norm $\|\cdot\|_{W_2^{-1}(\Omega^h)}$ we thus deduce that

$$\big\|D_x^+\eta_1 + D_y^+\eta_2\big\|_{W_2^{-1}(\Omega^h)} \le \|\eta_1\|_{L_2(\Omega_x^h)} + \|\eta_2\|_{L_2(\Omega_y^h)}.$$

Hence,

$$\|u - U\|_{W_2^1(\Omega^h)} \le 3\big(\|\eta_1\|_{L_2(\Omega_x^h)} + \|\eta_2\|_{L_2(\Omega_y^h)}\big). \tag{2.119}$$

It remains to bound the right-hand side of (2.119). We only consider the term involving η_1; the norm of η_2 is bounded analogously.

To this end, we first define

$$(\mu_y u)(x, y_j) := \frac{1}{8\hbar_j}\big[k_j u(x, y_{j-1}) + 6\hbar_j u(x, y_j) + k_{j+1} u(x, y_{j+1})\big],$$

and for fixed x, $0 \le x \le 1$, we let $I_y w(x,\cdot)$ denote the univariate continuous piecewise linear interpolant of $w(x,\cdot)$ on the mesh $\overline{\Omega}_y^h$. Then,

$$(\mu_y w)(x, y_j) = \frac{1}{\hbar_j}\int_{y_{j-1/2}}^{y_{j+1/2}} (I_y w)(x, y)\,\mathrm{d}y,$$

and therefore,

$$\begin{aligned}
(\mu_y u)_{ij} - (\mu_y u)_{i-1,j} &= \int_{x_{i-1}}^{x_i} \frac{\partial}{\partial x}(\mu_y u)(x, y_j)\,\mathrm{d}x \\
&= \int_{x_{i-1}}^{x_i} \frac{\partial}{\partial x}\frac{1}{\hbar_j}\int_{y_{j-1/2}}^{y_{j+1/2}} (I_y u)(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{\hbar_j}\int_{x_{i-1}}^{x_i}\int_{y_{j-1/2}}^{y_{j+1/2}} \frac{\partial}{\partial x}(I_y u)(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{\hbar_j}\int_{x_{i-1}}^{x_i}\int_{y_{j-1/2}}^{y_{j+1/2}} I_y\left(\frac{\partial u}{\partial x}\right)(x, y)\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}$$

Thus we find that $(\eta_1)_{ij}$ can be expressed as

$$(\eta_1)_{ij} = \frac{1}{h_i\hbar_j}\int_{x_{i-1}}^{x_i}\int_{y_{j-1/2}}^{y_{j+1/2}} \left[\frac{\partial u}{\partial x}(x_{i-1/2}, y) - \left(I_y\frac{\partial u}{\partial x}\right)(x, y)\right]\mathrm{d}x\,\mathrm{d}y.$$

By splitting η_1 as the sum of η_{11} and η_{12}, where

$$(\eta_{11})_{ij} := \frac{1}{h_i \hbar_j} \int_{x_{i-1}}^{x_i} \int_{y_j}^{y_{j+1/2}} \left[\frac{\partial u}{\partial x}(x_{i-1/2}, y) - \left(I_y \frac{\partial u}{\partial x} \right)(x, y) \right] \mathrm{d}x \, \mathrm{d}y,$$

$$(\eta_{12})_{ij} := \frac{1}{h_i \hbar_j} \int_{x_{i-1}}^{x_i} \int_{y_{j-1/2}}^{y_j} \left[\frac{\partial u}{\partial x}(x_{i-1/2}, y) - \left(I_y \frac{\partial u}{\partial x} \right)(x, y) \right] \mathrm{d}x \, \mathrm{d}y,$$

the task of estimating η_1 is reduced to bounding η_{11} and η_{12}.

Let us first consider η_{11}. By introducing the change of variables

$$x = x_{i-1/2} + \tilde{x} h_i, \quad -\frac{1}{2} \le \tilde{x} \le \frac{1}{2}; \qquad y = y_j + \tilde{y} k_{j+1}, \quad 0 \le \tilde{y} \le 1,$$

and defining $\tilde{v}(\tilde{x}, \tilde{y}) := h_i \frac{\partial u}{\partial x}(x, y)$, we can write

$$(\eta_{11})_{ij} = \frac{k_{j+1}}{h_i \hbar_j} \tilde{\eta}_{11}(\tilde{v}),$$

where

$$\tilde{\eta}_{11}(\tilde{v}) := \int_{-1/2}^{1/2} \int_0^{1/2} \left[\tilde{v}(0, \tilde{y}) - \tilde{v}(\tilde{x}, 0)(1 - \tilde{y}) - \tilde{v}(\tilde{x}, 1)\tilde{y} \right] \mathrm{d}\tilde{x} \, \mathrm{d}\tilde{y}.$$

Now $\tilde{\eta}_{11}$ can be regarded as a linear functional (with the argument $\tilde{v}$) defined on $W_2^s(\tilde{K}^*)$, where $s > 1/2$ and

$$\tilde{K}^* := \left(-\frac{1}{2}, \frac{1}{2} \right) \times (0, 1).$$

Thanks to the trace theorem (Theorem 1.42),

$$\left| \tilde{\eta}_{11}(\tilde{v}) \right| \le C \|\tilde{v}\|_{W_2^s(\tilde{K}^*)}, \quad s > 1/2,$$

and therefore $|\tilde{\eta}_{11}(\cdot)|$ is a bounded sublinear functional on $W_2^s(\tilde{K}^*)$. Moreover, if $\tilde{v}(\tilde{x}, \tilde{y}) = \tilde{x}^k \tilde{y}^l$, $k, l \in \{0, 1\}$, then $\tilde{\eta}_{11}(\tilde{v}) = 0$. By applying Theorem 1.9 with

$$\mathcal{U}_1 = W_2^s\big(\tilde{K}^*\big), \qquad \mathcal{U}_0 = L_2\big(\tilde{K}^*\big),$$

$$S = |\tilde{\eta}_{11}|, \qquad S_1 = \big(|\cdot|^2_{W_2^{0,s}(\tilde{K}^*)} + |\cdot|^2_{W_2^{s,0}(\tilde{K}^*)} \big)^{1/2}, \qquad S_0 = \|\cdot\|_{L_2(\tilde{K}^*)},$$

and noting that for $s > 0$ the Sobolev space $W_2^s(\tilde{K}^*)$ is compactly embedded in $L_2(\tilde{K}^*)$, we deduce that

$$\left| \tilde{\eta}_{11}(\tilde{v}) \right| \le C \big(|\tilde{v}|^2_{W_2^{0,s}(\tilde{K}^*)} + |\tilde{v}|^2_{W_2^{s,0}(\tilde{K}^*)} \big)^{1/2}$$

for $1/2 < s \leq 2$. By defining $K_{ij}^- := (x_{i-1}, x_i) \times (y_{j-1}, y_j)$, $K_{i,j+1}^- := (x_{i-1}, x_i) \times (y_j, y_{j+1})$ and returning from the $(\tilde{x}, \tilde{y})$–variables to the original (x, y) co-ordinates, we thus have that

$$\left|\tilde{\eta}_{11}(\tilde{v})\right|^2 \leq C\left(\frac{h_i^2 k_{j+1}^{2s}}{h_i k_{j+1}}\left|\frac{\partial u}{\partial x}\right|^2_{W_2^{0,s}(K_{i,j+1}^-)} + \frac{h_i^{2(s+1)}}{h_i k_{j+1}}\left|\frac{\partial u}{\partial x}\right|^2_{W_2^{s,0}(K_{i,j+1}^-)}\right),$$

and therefore

$$\left|(\eta_{11})_{ij}\right|^2 \leq C\left(\frac{k_{j+1}^{2s+1}}{h_i \hbar_j^2}\left|\frac{\partial u}{\partial x}\right|^2_{W_2^{0,s}(K_{i,j+1}^-)} + \frac{h_i^{2s-1} k_{j+1}}{\hbar_j^2}\left|\frac{\partial u}{\partial x}\right|^2_{W_2^{s,0}(K_{i,j+1}^-)}\right).$$

Analogously,

$$\left|(\eta_{12})_{ij}\right|^2 \leq C\left(\frac{k_{j}^{2s+1}}{h_i \hbar_j^2}\left|\frac{\partial u}{\partial x}\right|^2_{W_2^{0,s}(K_{ij}^-)} + \frac{h_i^{2s-1} k_{j}}{\hbar_j^2}\left|\frac{\partial u}{\partial x}\right|^2_{W_2^{s,0}(K_{ij}^-)}\right).$$

By noting the superadditivity of the Sobolev seminorm on a family of mutually disjoint Lebesgue-measurable subsets of Ω, we thus have that

$$\|\eta_1\|^2_{L_2(\Omega_x^h)} \leq Ch^{2s}\left(\left|\frac{\partial u}{\partial x}\right|^2_{W_2^{0,s}(\Omega)} + \left|\frac{\partial u}{\partial x}\right|^2_{W_2^{s,0}(\Omega)}\right), \tag{2.120}$$

where $h = \max_{i,j}(h_i, k_j)$. Analogously,

$$\|\eta_2\|^2_{L_2(\Omega_y^h)} \leq Ch^{2s}\left(\left|\frac{\partial u}{\partial y}\right|^2_{W_2^{0,s}(\Omega)} + \left|\frac{\partial u}{\partial y}\right|^2_{W_2^{s,0}(\Omega)}\right). \tag{2.121}$$

By substituting (2.120) and (2.121) into (2.119) we thus obtain the desired error bound

$$\|u - U\|_{W_2^1(\Omega^h)} \leq Ch^s |u|_{W_2^{s+1}(\Omega)}, \quad 1/2 < s \leq 2.$$

That completes the proof of the theorem. □

On a quasi-uniform mesh, the finite volume method (2.108) can be shown to be (almost) optimally accurate in the *discrete maximum norm* $\|\cdot\|_{\infty,h}$ defined by

$$\|V\|_{\infty,h} := \max_{(x,y)\in\overline{\Omega}^h}\left|V(x, y)\right|.$$

We shall say that $\{\overline{\Omega}^h\}$ is a family of *quasi-uniform Cartesian-product meshes* on $\overline{\Omega} = [0, 1] \times [0, 1]$ if there exists a positive constant $C_\star$ such that

$$h := \max_{i,j}(h_i, k_j) \leq C_\star \min_{i,j}(h_i, k_j).$$

Some auxiliary results are required to prove an error bound in the discrete maximum norm; these are formulated in the next two lemmas, the first of which states a version of the *inverse inequality* (see, for example, Ciarlet [26], Theorem 3.2.6).

Lemma 2.36 *Suppose that* $\{\overline{\Omega}^h\}$ *is a family of quasi-uniform Cartesian-product meshes on* $\overline{\Omega} = [0, 1] \times [0, 1]$, *and let* $\mathcal{S}^h$ *be the linear space of continuous piecewise bilinear polynomials defined on the partition of* Ω *induced by* $\overline{\Omega}^h$. *Suppose that* $1 \le q, r \le \infty$. *Then, there exists a positive constant* $C = C(C_\star, q, r)$, *independent of the discretization parameter* h, *such that*

$$\|V\|_{L_q(\Omega)} \le C h^{\min(0,(2/q)-(2/r))} \|V\|_{L_r(\Omega)} \quad \forall V \in \mathcal{S}^h.$$

Proof Consider the rectangle $K_{ij}^- := (x_{i-1}, x_i) \times (y_{j-1}, y_j)$, $1 \le i \le M$, $1 \le j \le N$, and the mapping $(\tilde{x}, \tilde{y}) \mapsto (x, y)$ defined by

$$x = x_{i-1} + \tilde{x} h_i, \qquad y = y_{j-1} + \tilde{y} k_j, \tag{2.122}$$

which maps the unit square $\tilde{K}^+ := (0, 1)^2$ onto K_{ij}^-. Let us define

$$\tilde{V}(\tilde{x}, \tilde{y}) := V(x, y),$$

where (x, y) is the image of $(\tilde{x}, \tilde{y})$ under the transformation (2.122). Now

$$\|\tilde{V}\|_{L_r(\tilde{K}^+)} = (h_i k_j)^{-1/r} \|V\|_{L_r(K_{ij}^-)},$$

and

$$\|V\|_{L_q(K_{ij}^-)} = (h_i k_j)^{1/q} \|\tilde{V}\|_{L_q(\tilde{K}^+)}.$$

Let $P(\tilde{K}^+)$ denote the linear space of all bilinear polynomials defined on the square $\tilde{K}^+$:

$$P\big(\tilde{K}^+\big) := \big\{(a + b\tilde{x})(c + d\tilde{y}) : a, b, c, d \in \mathbb{R}, 0 \le \tilde{x}, \tilde{y} \le 1\big\}.$$

Since $P(\tilde{K}^+)$ is finite-dimensional (in fact, the dimension of $P(\tilde{K}^+)$ is 4), the norms $\|\cdot\|_{L_q(\tilde{K}^+)}$ and $\|\cdot\|_{L_r(\tilde{K}^+)}$ are equivalent on $P(\tilde{K}^+)$. Hence, there is a constant $C_0 = C_0(q, r)$ such that

$$\|\tilde{V}\|_{L_q(\tilde{K}^+)} \le C_0 \|\tilde{V}\|_{L_r(\tilde{K}^+)},$$

for all $\tilde{V}$ in $P(\tilde{K}^+)$. Combining this with the two previous equalities yields

$$\|V\|_{L_q(K_{ij}^-)} \le C_0 (h_i k_j)^{(1/q)-(1/r)} \|V\|_{L_r(K_{ij}^-)},$$

and thus, by defining $C_1 = C_0 C_\star^{\max(0,(2/r)-(2/q))}$, we get

$$\|V\|_{L_q(K_{ij}^-)} \le C_1 h^{(2/q)-(2/r)} \|V\|_{L_r(K_{ij}^-)}. \tag{2.123}$$

Let us suppose that $q = \infty$; then, there exist i_0 and j_0, $1 \le i_0 \le M$, $1 \le j_0 \le N$, such that

$$\|V\|_{L_\infty(\Omega)} = \|V\|_{L_\infty(K_{i_0 j_0})} \le C_1 h^{-2/r} \|V\|_{L_r(K_{i_0 j_0})} \le C_1 h^{-2/r} \|V\|_{L_r(\Omega)},$$

which is the required result in the case of $q = \infty$.

Let us suppose now that $q < \infty$. It follows from (2.123) that

$$\left(\sum_{i,j} \|V\|^q_{L_q(K^-_{ij})}\right)^{1/q} \le C_1 h^{(2/q)-(2/r)} \left(\sum_{i,j} \|V\|^q_{L_r(K^-_{ij})}\right)^{1/q}, \qquad (2.124)$$

where the sums are taken over all i and j, $1 \le i \le M$, $1 \le j \le N$.

We shall consider three cases. When $r \le q$, by noting that $s \mapsto (\sum_{i,j} a^s_{ij})^{1/s}$ is monotonic decreasing on $[1, \infty)$ when $0 < a_{ij} \le 1$, we have, with $a_{ij} = \|V\|_{L_r(K^-_{ij})}/\|V\|_{L_r(\Omega)}$, that

$$\left(\sum_{i,j} \|V\|^q_{L_r(K^-_{ij})}\right)^{1/q} \le \left(\sum_{i,j} \|V\|^r_{L_r(K^-_{ij})}\right)^{1/r}.$$

When $q < r < \infty$, Hölder's inequality for finite sums gives

$$\begin{aligned}\left(\sum_{i,j} \|V\|^q_{L_r(K^-_{ij})}\right)^{1/q} &\le (MN)^{(1/q)-(1/r)} \left(\sum_{i,j} \|V\|^r_{L_r(K^-_{ij})}\right)^{1/r} \\ &\le \left(\frac{C_\star}{h}\right)^{(2/q)-(2/r)} \left(\sum_{i,j} \|V\|^r_{L_r(K^-_{ij})}\right)^{1/r}.\end{aligned}$$

Finally, when $r = \infty$, we have that

$$\left(\sum_{i,j} \|V\|^q_{L_\infty(K^-_{ij})}\right)^{1/q} \le \left(\frac{C_\star}{h}\right)^{2/q} \max_{ij} \|V\|_{L_\infty(K^-_{ij})}.$$

It remains to combine (2.124) with one of the three inequalities corresponding to $r \le q$, $q < r < \infty$ and $r = \infty$ respectively to complete the proof. □

Lemma 2.37 *Suppose that $\{\overline{\Omega}^h\}$ is a family of quasi-uniform Cartesian-product meshes, and let $\mathcal{S}^h_0$ be the linear space of continuous piecewise bilinear polynomials defined on the partition of Ω induced by $\overline{\Omega}^h$ that vanish on Γ. Then, there exists a positive constant C, independent of the discretization parameter h, such that,*

$$\|V\|_{L_\infty(\Omega)} \le C|\log h|^{1/2} \|\nabla V\|_{L_2(\Omega)} \quad \forall V \in \mathcal{S}^h_0.$$

Proof By Sobolev's embedding theorem on a Lipschitz domain $D \subset \mathbb{R}^n$,

$$\|v\|_{L_p(D)} \leq q\left(\frac{nq}{p}\right)^{1/n+1/p-1} \omega_n^{-1/n} n^{-1/p} \|\nabla v\|_{L_q(D)} \quad \forall v \in \mathring{W}_2^1(D),$$

where $q = np/(n+p)$, and $\omega_n := 2\pi^{n/2}/\Gamma(n/2)$ is the surface area of the unit ball in $\mathbb{R}^n$ (see inequality (2.3.21) in Maz'ya [136]). Specifically, by taking $n = 2$ and $D = \Omega$,

$$\|v\|_{L_p(\Omega)} \leq Cq\left(\frac{2q}{p}\right)^{1/2+1/p-1} 2^{-1/p} \|\nabla v\|_{L_q(\Omega)} \quad \forall v \in \mathring{W}_2^1(\Omega),$$

with $q = 2p/(2+p)$. Also, by the previous lemma,

$$\|V\|_{L_\infty(\Omega)} \leq Ch^{-2/p}\|V\|_{L_p(\Omega)}$$

and, by an analogous argument to that in the proof of the previous lemma,

$$\|\nabla V\|_{L_q(\Omega)} \leq Ch^{\min(0,2/q-1)}\|\nabla V\|_{L_2(\Omega)},$$

for all V in $\mathcal{S}_0^h$. Setting $p = |\log h| (> 1)$, for sufficiently small h, and combining the last three inequalities, we obtain the required result. □

Theorem 2.38 *Suppose that* $\{\overline{\Omega}^h\}$ *is a family of quasi-uniform Cartesian-product meshes*, *i.e. there exists a positive constant* $C_\star$ *such that*

$$h = \max_{i,j}(h_i, k_j) \leq C_\star \min_{i,j}(h_i, k_j),$$

and let $u \in W_2^{s+1}(\Omega) \cap \mathring{W}_2^1(\Omega)$, $1/2 < s \leq 2$. *Then,*

$$\|u - U\|_{\infty,h} \leq Ch^s |\log h|^{1/2} |u|_{W_2^{s+1}(\Omega)},$$

where $C = C(s)$ *is a positive constant depending on* $C_\star$, *but independent of* u *and the discretization parameter* h.

Proof Let $I^h : \mathring{W}_2^1(\Omega) \cap C(\overline{\Omega}) \to S_0^h$ denote the interpolation projector onto $\mathcal{S}_0^h$ defined by $(I^h u)(x_i, y_j) = u(x_i, y_j)$ for all $(x_i, y_j) \in \overline{\Omega}^h$. Then,

$$\|u - U\|_{\infty,h} = \left\|I^h u - U\right\|_{\infty,h} \leq \left\|I^h u - U\right\|_{L_\infty(\Omega)}.$$

Thanks to Lemma 2.37,

$$\|V\|_{L_\infty(\Omega)} \leq C|\log h|^{1/2}\|V\|_{W_2^1(\Omega)} \quad \forall V \in \mathcal{S}_0^h.$$

Also, the equivalence of the norms $\|\cdot\|_{W_2^1(\Omega)}$ and $\|\cdot\|_{W_2^1(\Omega^h)}$ on $\mathcal{S}_0^h$ implies that

$$\|V\|_{W_2^1(\Omega)} \leq C\|V\|_{W_2^1(\Omega^h)} \quad \forall V \in \mathcal{S}_0^h.$$

Hence,

$$\|u - U\|_{\infty,h} \le C|\log h|^{1/2}\|u - U\|_{W_2^1(\Omega^h)},$$

and therefore Theorem 2.35 yields

$$\|u - U\|_{\infty,h} \le Ch^s|\log h|^{1/2}|u|_{W_2^{s+1}(\Omega)}. \qquad \square$$

In the next section we extend the error analysis developed here to a more general class of schemes.

2.4.2 An Alternative Scheme

Hitherto it was assumed that the trial space $\mathcal{S}^h$ in the finite volume method (which was subsequently rewritten as a finite difference scheme) consisted of continuous piecewise bilinear functions on the rectangular partition of $\overline{\Omega}$ induced by the Cartesian-product mesh $\overline{\Omega}^h$. One can construct an alternative method, based on continuous piecewise linear trial functions on triangles; to this end, we consider a triangulation of $\overline{\Omega}$ obtained from the original rectangular partition by subdividing each rectangle into two triangles by the diagonal of positive slope. Let $\mathcal{S}^h$ denote the set of all continuous piecewise linear functions on this triangulation, and let $\mathcal{S}_0^h$ be the subset of $\mathcal{S}^h$ consisting of all those functions that vanish on Γ.

Similarly to (2.108), we define the finite volume approximation of u as $U \in \mathcal{S}_0^h$ satisfying

$$-\frac{1}{\hbar_i \hbar_j}\int_{\partial K_{ij}} \frac{\partial U}{\partial \nu}\,\mathrm{d}s = \frac{1}{\hbar_i \hbar_j}(\chi_{ij} * f)(x_i, y_j) \quad \text{for } (x_i, y_j) \in \Omega^h. \tag{2.125}$$

This scheme resembles the finite volume method (2.108). Indeed, a simple calculation reveals that (2.125) can be rewritten as the finite difference scheme

$$-\big(D_x^+ D_x^- + D_y^+ D_y^-\big)U = T_h^{11} f \quad \text{in } \Omega^h, \tag{2.126}$$

$$U = 0 \quad \text{on } \Gamma^h. \tag{2.127}$$

In fact, both (2.111), (2.112) and (2.126), (2.127) can be embedded in the following one-parameter family of finite difference schemes:

$$-\big(D_x^+ D_x^- \mu_y^\theta + D_y^+ D_y^- \mu_x^\theta\big)U = T_h^{11} f \quad \text{in } \Omega^h, \tag{2.128}$$

$$U = 0 \quad \text{on } \Gamma^h, \tag{2.129}$$

where $\theta \in [0, 1]$, and

$$\mu_x^\theta U_{ij} := \frac{1}{\hbar_i}\big[\theta h_i U_{i-1,j} + (1 - 2\theta)\hbar_i U_{ij} + \theta h_{i+1} U_{i+1,j}\big],$$

with μ_y^θ defined analogously. The scheme (2.111), (2.112) (resp. (2.126), (2.127)) is obtained from (2.128), (2.129) with $\theta = 1/8$ (resp. $\theta = 0$). The rest of this section is devoted to the analysis of the one-parameter family of schemes (2.126), (2.127).

By proceeding similarly as in the proofs of Lemmas 2.31, 2.32 and Theorems 2.33 and 2.34 we arrive at the following set of results, whose proofs have been omitted for the sake of brevity.

Lemma 2.39 *Suppose that V is a mesh-function defined on $\overline{\Omega}^h$, and let $\theta \in [0, 1/4)$.*

(a) *If $V = 0$ on Γ_x^h, then*

$$\big(\mu_x^\theta V, V\big]_y \geq (1-4\theta)\|V\|^2_{L_2(\Omega_y^h)}.$$

(b) *If $V = 0$ on Γ_y^h, then*

$$\big(\mu_y^\theta V, V\big]_x \geq (1-4\theta)\|V\|^2_{L_2(\Omega_x^h)}.$$

Theorem 2.40 *Let $L^h V := -(D_x^+ D_x^- \mu_y^\theta + D_y^+ D_y^- \mu_x^\theta)V$, and suppose that $\theta \in [0, 1/4)$. Then,*

$$\|V\|_{W_2^1(\Omega^h)} \leq \frac{3}{2(1-4\theta)}\big\|L^h V\big\|_{W_2^{-1}(\Omega^h)},$$

for any mesh-function V defined on $\overline{\Omega}^h$ and such that $V = 0$ on Γ^h.

Theorem 2.41 *Suppose that $\theta \in [0, 1/4)$. For any $f \in W_2^s(\Omega)$, $s > -1/2$,* (2.128), (2.129) *has a unique solution U. Moreover,*

$$\|U\|_{W_2^1(\Omega^h)} \leq \frac{3}{2(1-4\theta)}\big\|T_h^{11} f\big\|_{W_2^{-1}(\Omega^h)}.$$

The central result of this section is the following error bound for the finite difference scheme (2.128), (2.129).

Theorem 2.42 *Suppose that $u \in W_2^3(\Omega) \cap \mathring{W}_2^1(\Omega)$, and let $\theta \in [0, 1/4)$. Then,*

$$\|u - U\|_{W_2^1(\Omega^h)} \leq Ch^2 |u|_{W_2^3(\Omega)},$$

where $h = \max_{i,j}(h_i, k_j)$ and $C = C(\theta)$ is a positive constant independent of u and the discretization parameters.

Proof Let us define the global error as $e := u - U$. We then have that

$$-\big(D_x^+ D_x^- \mu_y^\theta + D_y^+ D_y^- \mu_x^\theta\big)e = D_x^+ \eta_1^\theta + D_y^+ \eta_2^\theta \quad \text{in } \Omega^h, \tag{2.130}$$

$$e = 0 \quad \text{on } \Gamma^h, \tag{2.131}$$

where

$$\eta_1^\theta := \eta_1 + \left(\frac{1}{8} - \theta\right)\zeta_1, \qquad \eta_2^\theta := \eta_2 + \left(\frac{1}{8} - \theta\right)\zeta_2,$$

and

$$\eta_1 := T_-^{01}\frac{\partial u}{\partial x} - D_x^- \mu_y u, \qquad \eta_2 := T_-^{10}\frac{\partial u}{\partial y} - D_y^- \mu_x u,$$

as in the proof of Theorem 2.35, and

$$(\zeta_1)_{ij} := h_i^2 D_x^- D_y^+ D_y^- u_{ij}, \qquad (\zeta_2)_{ij} := k_j^2 D_y^- D_x^+ D_x^- u_{ij}.$$

By applying Theorem 2.40 to (2.130), (2.131) we then deduce that

$$\|e\|_{W_2^1(\Omega^h)} \le \frac{3}{2(1-4\theta)} \left\| D_x^+ \eta_1^\theta + D_y^+ \eta_2^\theta \right\|_{W_2^{-1}(\Omega^h)}.$$

Consequently,

$$\begin{aligned} \|u - U\|_{W_2^1(\Omega^h)} \le{}& \frac{3}{2(1-4\theta)}\left(\|\eta_1\|_{L_2(\Omega_x^h)} + \|\eta_2\|_{L_2(\Omega_y^h)}\right) \\ &+ \frac{3|1-8\theta|}{16(1-4\theta)}\left(\|\zeta_1\|_{L_2(\Omega_x^h)} + \|\zeta_2\|_{L_2(\Omega_y^h)}\right). \end{aligned} \tag{2.132}$$

The first two terms on the right-hand side have already been bounded in the proof of Theorem 2.35; we showed there that

$$\|\eta_1\|_{L_2(\Omega_x^h)} \le Ch^2 \left(\left|\frac{\partial u}{\partial x}\right|_{W_2^{0,2}(\Omega)} + \left|\frac{\partial u}{\partial x}\right|_{W_2^{2,0}(\Omega)} \right) \tag{2.133}$$

and

$$\|\eta_2\|_{L_2(\Omega_y^h)} \le Ch^2 \left(\left|\frac{\partial u}{\partial y}\right|_{W_2^{0,2}(\Omega)} + \left|\frac{\partial u}{\partial y}\right|_{W_2^{2,0}(\Omega)} \right). \tag{2.134}$$

It therefore remains to bound the norms of ζ_1 and ζ_2. We observe in passing that for $\theta = 1/8$ the terms involving ζ_1 and ζ_2 are absent from (2.132).

To this end, let $\phi_i(x)$ (resp. $\psi_j(y)$) denote the standard continuous piecewise linear finite element basis function on $\overline{\Omega}_x^h$ (resp. $\overline{\Omega}_y^h$) such that $\phi_i(x_k) = \delta_{ik}$ (resp. $\psi_j(y_k) = \delta_{jk}$); $(\zeta_1)_{ij}$ and $(\zeta_2)_{ij}$ can then be rewritten as

$$(\zeta_1)_{ij} = h_i^2 \frac{1}{h_i \hbar_j} \int_{x_{i-1}}^{x_i} \int_{y_{j-1}}^{y_{j+1}} \psi_j(y) \frac{\partial^3 u}{\partial x \partial y^2}(x, y)\, \mathrm{d}x\, \mathrm{d}y,$$

$$(\zeta_2)_{ij} = k_j^2 \frac{1}{\hbar_i k_j} \int_{x_{i-1}}^{x_{i+1}} \int_{y_{j-1}}^{y_j} \phi_i(x) \frac{\partial^3 u}{\partial x^2 \partial y}(x, y)\, \mathrm{d}x\, \mathrm{d}y.$$

Clearly,

$$\|\zeta_1\|_{L_2(\Omega_x^h)} \le Ch^2 \left|\frac{\partial u}{\partial x}\right|_{W_2^{0,2}(\Omega)} \tag{2.135}$$

and

$$\|\zeta_2\|_{L_2(\Omega_y^h)} \le Ch^2 \left|\frac{\partial u}{\partial y}\right|_{W_2^{2,0}(\Omega)}. \tag{2.136}$$

Inserting (2.133)–(2.136) in (2.132) we obtain the desired error bound. □

On a quasi-uniform mesh the scheme (2.128), (2.129) can be shown to be (almost) optimally accurate in the discrete maximum norm $\|\cdot\|_{\infty,h}$ for any $\theta \in [0, 1/4)$, by proceeding analogously as in the case of $\theta = 1/8$.

Theorem 2.43 *Suppose that* $\{\overline{\Omega}^h\}$ *is a family of quasi-uniform meshes*, $\theta \in [0, 1/4)$, *and let* $u \in W_2^3(\Omega) \cap \mathring{W}_2^1(\Omega)$. *Then,*

$$\|u - U\|_{\infty,h} \le C(\theta) h^2 |\log h|^{1/2} |u|_{W_2^3(\Omega)}.$$

The proof of this result is analogous to that of Theorem 2.38.

2.4.3 The Rotated Discrete Laplacian

In the previous section we considered the analysis of a one-parameter family of finite difference schemes, parametrized by θ. For $\theta \in [0, 1/4)$ we showed there that the scheme is stable and we proved optimal-order error bounds in various norms. A natural question is: *what happens when* $\theta = 1/4$*?* This section is devoted to the analysis of the resulting discretization.

Let us consider the finite difference scheme (2.128), (2.129), with $\theta = 1/4$. For the sake of notational simplicity we define

$$\hat{\mu}_x V_{ij} := \frac{1}{4\hbar_i}(h_i V_{i-1,j} + 2\hbar_i V_{ij} + h_{i+1} V_{i+1,j}),$$

and $\hat{\mu}_y$ is defined analogously. In fact, by introducing

$$\nu_x V_{ij} := \frac{1}{2}(V_{ij} + V_{i-1,j})$$

we can write

$$\hat{\mu}_x V_{ij} = \frac{1}{2\hbar_i}(h_i \nu_x V_{ij} + h_{i+1} \nu_x V_{i+1,j}).$$

Analogously, by letting

$$\nu_y V_{ij} := \frac{1}{2}(V_{ij} + V_{i,j-1})$$

we have that

$$\hat{\mu}_y V_{ij} = \frac{1}{2\hbar_j}(k_j \nu_y V_{ij} + k_{j+1}\nu_y V_{i,j+1}).$$

In terms of this new notation, for $\theta = 1/4$ the finite difference scheme (2.128), (2.129) can be rewritten as follows:

$$-\big(D_x^+ D_x^- \hat{\mu}_y + D_y^+ D_y^- \hat{\mu}_x\big)U = T_h^{11} f \quad \text{in } \Omega^h, \tag{2.137}$$

$$U = 0 \quad \text{on } \Gamma^h. \tag{2.138}$$

In particular, on a uniform mesh of size h, the resulting five-point finite difference operator is given by

$$-\frac{1}{2h^2}(U_{i-1,j-1} + U_{i-1,j+1} + U_{i+1,j-1} + U_{i+1,j+1} - 4U_{ij})$$

and is usually referred to as the *rotated discrete Laplace operator*.

We begin by showing that the scheme (2.137), (2.138) is stable. A preliminary result in this direction stated in the next lemma concerns the averaging operators $\hat{\mu}_x$, ν_x, $\hat{\mu}_y$ and ν_y.

Lemma 2.44 *Suppose that V is a function defined on the mesh $\overline{\Omega}_h$.*

(a) *If $V_{0j} = V_{Mj} = 0$ for $j = 1, \dots, N$, then*

$$(\hat{\mu}_x V, V]_y = \sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j |\nu_x V_{ij}|^2;$$

(b) *If $V_{i0} = V_{iM} = 0$ for $i = 1, \dots, M$, then*

$$(\hat{\mu}_y V, V]_x = \sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j |\nu_y V_{ij}|^2.$$

Proof We shall prove (a); the proof of (b) is completely analogous. By noting the definition of $\hat{\mu}_x$ we have that

$$\begin{aligned}(\hat{\mu}_x V, V]_y &= \frac{1}{4}\sum_{i=1}^{M-1}\sum_{j=1}^{N} k_j\big[h_{i+1}V_{i+1,j} + V_{ij}(h_{i+1}+h_i) + h_i V_{i-1,j}\big]V_{ij} \\ &= \frac{1}{4}\sum_{j=1}^{N} k_j\left[\sum_{i=1}^{M-1}(h_{i+1}+h_i)V_{ij}^2 + 2\sum_{i=1}^{M} h_i V_{ij} V_{i-1,j}\right]\end{aligned}$$

$$= \frac{1}{4}\sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j (V_{ij} + V_{i-1,j})^2,$$

and that completes the proof of (a). □

Lemma 2.45 *Let* $L^h V := -(D_x^+ D_x^- \hat{\mu}_y + D_y^+ D_y^- \hat{\mu}_x)V$. *Then,*

$$\big(L^h V, V\big)_h = \sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j \big(\big|\nu_y D_x^- V_{ij}\big|^2 + \big|\nu_x D_y^- V_{ij}\big|^2\big)$$

for any mesh-function V *defined on* $\overline{\Omega}^h$ *such that* $V = 0$ *on* Γ^h.

Proof This identity is a straightforward consequence of Lemma 2.44 by observing that

$$\begin{aligned}\big(L^h V, V\big)_h &= \big(D_x^- \hat{\mu}_y V, D_x^- V\big]_x + \big(D_y^- \hat{\mu}_x V, D_y^- V\big]_y \\ &= \big(\hat{\mu}_y D_x^- V, D_x^- V\big]_x + \big(\hat{\mu}_x D_y^- V, D_y^- V\big]_y,\end{aligned}$$

where the first equality follows by summation by parts and the second by noting that D_x^- commutes with $\hat{\mu}_y$ and D_y^- commutes with $\hat{\mu}_x$. □

We deduce from Lemma 2.45 that

$$\sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j \big(\big|\nu_y D_x^- V_{ij}\big|^2 + \big|\nu_x D_y^- V_{ij}\big|^2\big) = \big(L^h V, V\big)_h,$$

for any function V defined on $\overline{\Omega}^h$ such that $V = 0$ on Γ^h. Therefore, by applying the Cauchy–Schwarz inequality on the right-hand side, noting that

$$V_{ij} = \nu_x V_{ij} + \frac{1}{2} h_i D_x^- V_{ij}, \tag{2.139}$$

and letting

$$W_{ij} := h_i D_x^- V_{ij},$$

we deduce that, for any such mesh-function V,

$$\begin{aligned}&\sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j \big(\big|\nu_y D_x^- V_{ij}\big|^2 + \big|\nu_x D_y^- V_{ij}\big|^2\big) \\ &\quad \le \big\|L^h V\big\|_{L_2(\Omega^h)} \bigg(\|\nu_x V\|_{L_2(\Omega^h)} + \frac{1}{2}\|W\|_{L_2(\Omega^h)}\bigg).\end{aligned} \tag{2.140}$$

Now to complete the stability analysis of the finite difference scheme (2.137), (2.138) it remains to relate the two norms in the brackets on the right-hand side of (2.140) to the expression on the left. To do so, we state and prove two lemmas.

Lemma 2.46 *Suppose that $\{\overline{\Omega}^h\}$ is a family of quasi-uniform meshes, i.e. there exists a positive constant $C_\star$ such that*

$$h = \max_{i,j}(h_i, k_j) \leq C_\star \min_{i,j}(h_i, k_j).$$

Let V be a function defined on $\overline{\Omega}^h$ such that $V = 0$ on Γ^h; then,

$$\|\nu_x V\|_{L_2(\Omega^h)} \leq \frac{1}{2}(1 + C_\star)^{1/2}\left(\sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j \left|\nu_x D_y^- V_{ij}\right|^2\right)^{1/2}.$$

Proof Let $Z_{ij} = \nu_x V_{ij}$; then, because $Z_{i0} = 0$ for $i = 1, \dots, M$, we have that

$$\begin{aligned}|Z_{ij}|^2 &= \left(\sum_{n=1}^{j} k_n D_y^- Z_{in}\right)^2 \\ &\leq \left(\sum_{n=1}^{j} k_n\right)\left(\sum_{n=1}^{j} k_n \left|D_y^- Z_{in}\right|^2\right) \leq \sum_{n=1}^{j} k_n \left|D_y^- Z_{in}\right|^2\end{aligned}$$

for $i = 1, \dots, M-1$ and $j = 1, \dots, N$. Hence,

$$\|Z\|^2_{L_2(\Omega^h)} \leq \sum_{i=1}^{M-1}\sum_{n=1}^{j} \hbar_i k_n \left|D_y^- Z_{in}\right|^2, \quad 1 \leq j \leq N.$$

Similarly, since $Z_{iN} = 0$ for $i = 1, \dots, M$, we also have that

$$\|Z\|^2_{L_2(\Omega^h)} \leq \sum_{i=1}^{M-1}\sum_{n=j+1}^{N} \hbar_i k_n \left|D_y^- Z_{in}\right|^2, \quad 0 \leq j \leq N-1.$$

By adding the last two inequalities we deduce that

$$\|Z\|^2_{L_2(\Omega^h)} \leq \frac{1}{2}\sum_{i=1}^{M-1}\sum_{n=1}^{N} \hbar_i k_n \left|D_y^- Z_{in}\right|^2.$$

Because ν_x commutes with D_y^- this yields

$$\|\nu_x V\|^2_{L_2(\Omega^h)} \leq \frac{1}{2}\sum_{i=1}^{M-1}\sum_{j=1}^{N} \hbar_i k_j \left|\nu_x D_y^- V_{ij}\right|^2.$$

Since $\hbar_i \le \frac{1}{2} h_i (1 + C_\star)$ it follows that

$$\|\nu_x V\|^2_{L_2(\Omega^h)} \le \frac{1}{4}(1 + C_\star) \sum_{i=1}^{M-1} \sum_{j=1}^{N} h_i k_j \left|\nu_x D_y^- V_{ij}\right|^2,$$

and hence, by increasing the right-hand side of this inequality further by extending the upper limit of the sum over i from $M - 1$ to M, we obtain the desired inequality. □

Our next result is concerned with bounding $W_{ij} := h_i D_x^- V_{ij}$.

Lemma 2.47 *Suppose that* $\{\overline{\Omega}^h\}$ *is a family of quasi-uniform meshes*, *i.e. there exists a positive constant* $C_\star$ *such that*

$$h = \max_{i,j}(h_i, k_j) \le C_\star \min_{i,j}(h_i, k_j).$$

Let V *be a function defined on* $\overline{\Omega}^h$ *such that* $V = 0$ *on* Γ^h *and let* $W_{ij} = h_i D_x^- V_{ij}$; *then*,

$$\|W\|_{L_2(\Omega^h)} \le 2C_\star \left(\sum_{i=1}^{M} \sum_{j=1}^{N} h_i k_j \left|\nu_y D_x^- V_{ij}\right|^2 \right)^{1/2}.$$

Proof By noting that $W_{i0} = 0$ for $i = 1, \ldots, M$, we have that

$$W_{ij} = \sum_{n=1}^{j} (-1)^{j-n} (W_{in} + W_{i,n-1})$$

for $i = 1, \ldots, M$ and $j = 1, \ldots, N$. Therefore,

$$|W_{ij}|^2 \le 4j \sum_{n=1}^{j} |\nu_y W_{in}|^2 \le 4j \sum_{n=1}^{N} h_i^2 \left|\nu_y D_x^- V_{in}\right|^2.$$

As $h_i^2 \le h h_i$ and $\hbar_i \le C_\star k_n$ for all $i \in \{1, \ldots, M\}$ and all $n \in \{1, \ldots, N\}$, and $h \sum_{j=1}^{N} j \bar{k}_j \le N h \le C_\star$, we deduce that

$$\|W\|^2_{L_2(\Omega^h)} \le 4C_\star^2 \sum_{i=1}^{M} \sum_{n=1}^{N} h_i k_n \left|\nu_y D_x^- V_{in}\right|^2,$$

and hence the desired inequality upon renaming the index n into j. □

By combining (2.139), (2.140) and Lemmas 2.46 and 2.47, we deduce that

$$\sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j \left(\left|\nu_y D_x^- V_{ij}\right|^2 + \left|\nu_x D_y^- V_{ij}\right|^2\right)$$
$$\leq C\left\|L^h V\right\|_{L_2(\Omega^h)} \left(\sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j \left(\left|\nu_y D_x^- V_{ij}\right|^2 + \left|\nu_x D_y^- V_{ij}\right|^2\right)\right)^{1/2}.$$

This yields the inequality

$$\left(\sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j \left(\left|\nu_y D_x^- V_{ij}\right|^2 + \left|\nu_x D_y^- V_{ij}\right|^2\right)\right)^{1/2} \leq C\left\|L^h V\right\|_{L_2(\Omega^h)}, \tag{2.141}$$

and thereby the difference scheme is stable in the discrete W_2^1 norm defined by the left-hand side of this inequality.

Remark 2.11 We note that stability has been proved in a weaker sense here, for $\theta = 1/4$, than in the previous section for $\theta \in [0, 1/4)$. Indeed, for $\theta \in [0, 1/4)$ we deduce from Theorem 2.40 the stronger bound

$$\left[\sum_{i=1}^{M}\sum_{j=1}^{N} h_i k_j \left(\left|D_x^- V_{ij}\right|^2 + \left|D_y^- V_{ij}\right|^2\right)\right]^{1/2} \leq C(\theta)\left\|L^h V\right\|_{W_2^{-1}(\Omega^h)}, \tag{2.142}$$

whose left-hand side is an upper bound on the left-hand side of (2.141).

Worse still, the stability of the scheme (2.137), (2.138) is not robust, in the sense that when the homogeneous Dirichlet boundary condition is replaced by 1-periodic boundary conditions in the two co-ordinate directions, on a uniform mesh with spacing $h = 1/(2M)$, $M > 1$, the resulting difference scheme is ill-posed for any 1-periodic f. To see this, first take $f = 0$ and note that, in addition to the trivial constant solution (which is, incidentally, also a solution to the boundary-value problem), the difference scheme has the oscillatory chequer-board-like solution $U_{ij}^\star = (-1)^{i+j}$. Thus if U is a solution of the difference scheme with $f \neq 0$ subject to 1-periodic boundary conditions in the two co-ordinate directions, then $U + \alpha U^\star$ is also a solution, for any real number α. In other words, the solution is not unique. In fact, the finite difference scheme (2.137), corresponding to the choice of $\theta = 1/4$ in (2.128), with 1-periodic boundary condition, has infinitely many solutions for any f. This is consistent with the fact that, with a 1-periodic boundary condition, the expression appearing on the left-hand side of (2.142) has a nontrivial kernel in the set of mesh-functions defined on a uniform mesh with spacing $h = 1/(2M)$, $M > 1$, and is therefore only a seminorm in that case rather than a norm; and it is also consistent with the fact that, with $\theta \in [0, 1/4)$, the stability constant $C(\theta)$ of the scheme (2.128), (2.129) in the discrete $W_2^1(\Omega^h)$ norm, appearing in (2.142), tends to $+\infty$ as $\theta \to 1/4 - 0$.

2.5 Convergence Analysis in L_p Norms

Hitherto, with the exception of various error bounds in the discrete maximum norm, we have been concerned with the error analysis of finite difference schemes in mesh-dependent analogues of Hilbertian Sobolev norms, i.e. discrete Sobolev norms that are induced by inner products.

In this section we develop a framework for the error analysis of finite difference schemes in mesh-dependent versions of the Sobolev and Bessel-potential norms W_p^s and H_p^s, respectively. For the sake of simplicity, we shall confine ourselves to finite difference approximations of the homogeneous Dirichlet boundary-value problem for Poisson's equation on an open square Ω, assuming that the weak solution of the boundary-value problem belongs to $W_p^s(\Omega)$, $0 \le s \le 4$, $1 < p < \infty$. We shall make extensive use of the theory of discrete Fourier multipliers to investigate the stability of the difference schemes considered, in conjunction with the Bramble–Hilbert lemma in fractional-order Sobolev spaces to derive error bounds of optimal order. The presentation in this section is based on the following sources: the journal papers by Mokin [140] and Süli, Jovanović, Ivanović [173] and the monograph of Samarskiĭ, Lazarov and Makarov [160].

2.5.1 Discrete Fourier Multipliers

In previous sections we relied on the use of energy estimates based on Hilbert space techniques to show the stability of the finite difference schemes considered. In order to extend these stability results to L_p norms, $p \ne 2$, we require a new tool – discrete Fourier multipliers. To this end, we shall state and prove below a discrete counterpart of the Marcinkiewicz multiplier theorem. First, however, we shall introduce the notion of *discrete Fourier transform*.

Suppose that N is a positive integer and $h = \pi/N$. We consider the mesh

$$\mathbb{R}_h^n = h\mathbb{Z}^n := \left\{x \in \mathbb{R}^n : x = hk,\ k \in \mathbb{Z}^n\right\}$$

and the set of all 2π-periodic mesh-functions defined on $\mathbb{R}_h^n$. We let

$$\mathbb{I} := \{-N+1, \dots, -1, 0, \dots, N\}.$$

Then, any 2π-periodic function V defined on $\mathbb{R}_h^n$ is completely determined by its values on the 'basic cell'

$$\omega^h = h\mathbb{I}^n := \left\{hk : k \in \mathbb{I}^n\right\}.$$

With each mesh-function V defined on ω^h we associate its *discrete Fourier transform* $\mathcal{F}V$ given by

$$(\mathcal{F}V)(k) := h^n \sum_{x \in \omega^h} V(x) \mathrm{e}^{-\imath x \cdot k}, \quad k \in \mathbb{I}^n. \tag{2.143}$$

In order to distinguish the discrete Fourier transform from its integral counterpart F defined in Chap. 1 we have used the calligraphic letter $\mathcal{F}$ here instead of F. Clearly, $\mathcal{F}V$ is a $2N$-periodic function of its variables $k_1, \dots, k_n$, and $2N$ is the minimum period; thus it suffices to consider $\mathcal{F}V$ on the basic cell $\mathbb{I}^n$. Hence our choice of $k \in \mathbb{I}^n$ in (2.143).

For $x \in \omega^h$ the following *discrete Fourier inversion formula* holds:

$$V(x) = \frac{1}{(2\pi)^n} \sum_{k \in \mathbb{I}^n} (\mathcal{F}V)(k) \mathrm{e}^{\imath x \cdot k}. \tag{2.144}$$

Indeed, substituting (2.143) into the right-hand side of (2.144), we have that

$$\frac{1}{(2\pi)^n} \sum_{k \in \mathbb{I}^n} \mathrm{e}^{\imath x \cdot k} \sum_{y \in \omega^h} h^n V(y) \mathrm{e}^{-\imath y \cdot k} = \frac{1}{(2\pi)^n} \sum_{y \in \omega^h} h^n V(y) \sum_{k \in \mathbb{I}^n} \mathrm{e}^{\imath (x-y) \cdot k}.$$

However, for any $x, y \in \omega^h$ we have that

$$\sum_{k \in \mathbb{I}^n} \mathrm{e}^{\imath (x-y) \cdot k} = \begin{cases} (2N)^n & \text{if } x = y, \\ 0 & \text{otherwise,} \end{cases}$$

and hence (2.144), by noting that $h^n (2N)^n = (2\pi)^n$.

We can write (2.144) as $V = \mathcal{F}^{-1}\mathcal{F}V$ where, for a sequence $a = \{a(k)\}_{k \in \mathbb{I}^n}$, the *inverse discrete Fourier transform* $\mathcal{F}^{-1}a$ of a is defined by

$$\big(\mathcal{F}^{-1}a\big)(x) := \frac{1}{(2\pi)^n} \sum_{k \in \mathbb{I}^n} a(k) \mathrm{e}^{\imath x \cdot k}, \quad x \in \omega^h.$$

Assuming that V is a function defined on the mesh ω^h, we consider the trigonometric polynomial T_V given by

$$T_V(x) = \frac{1}{(2\pi)^n} \sum_{k \in \mathbb{I}^n} (\mathcal{F}V)(k) \mathrm{e}^{\imath x \cdot k}, \quad x \in (-\pi, \pi]^n. \tag{2.145}$$

According to the discrete Fourier inversion formula,

$$T_V(x) = V(x) \quad \forall x \in \omega^h;$$

in other words, T_V *interpolates* V over the mesh ω^h.

Next we introduce the space $L_p(\omega^h)$, $1 \le p < \infty$, consisting all mesh-functions V defined on ω^h such that, for some constant M, independent of the discretization parameter h,

$$\|V\|_{L_p(\omega^h)} = \left(h^n \sum_{x \in \omega^h} \big|V(x)\big|^p \right)^{1/p} \le M.$$

The following lemma establishes a useful relationship between the L_p norm of a mesh-function V defined on ω^h and the L_p norm of the associated trigonometric interpolant T_V on $\omega = \mathbb{T}^n := (-\pi, \pi)^n$.

Lemma 2.48 *Suppose that $V \in L_p(\omega^h)$, $1 \leq p < \infty$, and let $\omega = (-\pi, \pi)^n$. Then,*

$$\|V\|_{L_p(\omega^h)} \leq (1+\pi)^n \|T_V\|_{L_p(\omega)}.$$

Proof Let us suppose for simplicity that $n = 1$; for $n > 1$ the proof follows from the case of $n = 1$ by induction over n. We shall first show that there exists a real number ξ_0 in the interval $(-h, 0)$ such that

$$\|T_V\|_{L_p(\omega)} = \left(h \sum_{x \in \omega^h} \left|T_V(x+\xi_0)\right|^p\right)^{1/p}, \tag{2.146}$$

where now $\omega = (-\pi, \pi)$ and $\omega^h = h\mathbb{I}$.

Indeed,

$$\begin{aligned}
\|T_V\|^p_{L_p(\omega)} &= \int_{-\pi}^{\pi} \left|T_V(x)\right|^p \mathrm{d}x = \sum_{k=-N+1}^{N} \int_{x_{k-1}}^{x_k} \left|T_V(x)\right|^p \mathrm{d}x \\
&= \sum_{k=-N+1}^{N} \int_0^h \left|T_V(y + x_{k-1})\right|^p \mathrm{d}y \\
&= \int_0^h \sum_{k \in \mathbb{I}} \left|T_V(x_k + y - h)\right|^p \mathrm{d}y.
\end{aligned}$$

Now the integrand is a continuous function of y on $[0, h]$; therefore, by the integral mean-value theorem, there exists a ξ in $(0, h)$ such that

$$\int_0^h \sum_{k \in \mathbb{I}} \left|T_V(x_k + y - h)\right|^p \mathrm{d}y = h \sum_{k \in \mathbb{I}} \left|T_V(x_k + \xi - h)\right|^p.$$

Letting $\xi_0 := \xi - h$ and noting that $k \in \mathbb{I}$ if, and only if, $x = x_k \in \omega^h = h\mathbb{I}$, we deduce (2.146).

Now consider

$$\mathcal{D} := \left|\left(h \sum_{x \in \omega^h} \left|V(x)\right|^p\right)^{1/p} - \left(h \sum_{x \in \omega^h} \left|T_V(x+\xi_0)\right|^p\right)^{1/p}\right|.$$

We shall prove that

$$\mathcal{D} \leq h \left\|T_V'\right\|_{L_p(\omega)}. \tag{2.147}$$

This follows by noting that $V(x) = T_V(x)$ for x in ω^h, and observing that by the reverse triangle inequality, the Newton–Leibniz formula and Hölder's inequality we have that

$$\begin{aligned}
\mathcal{D} &\le \left(\sum_{x\in\omega^h} h \left| T_V(x) - T_V(x+\xi_0) \right|^p \right)^{1/p} \\
&= \left(\sum_{x\in\omega^h} h \left| \int_{x+\xi_0}^{x} T_V'(t)\,\mathrm{d}t \right|^p \right)^{1/p} \\
&\le \left(\sum_{x\in\omega^h} h \left(\int_{x-h}^{x} \left| T_V'(t) \right| \mathrm{d}t \right)^p \right)^{1/p} \\
&\le \left(\sum_{x\in\omega^h} h h^{p-1} \int_{x-h}^{x} \left| T_V'(t) \right|^p \mathrm{d}t \right)^{1/p} = h \left\| T_V' \right\|_{L_p(\omega)}.
\end{aligned}$$

Now using (2.146) and (2.147) we deduce that

$$\begin{aligned}
\|V\|_{L_p(\omega^h)} &= \|V\|_{L_p(\omega^h)} - \left(h \sum_{x\in\omega^h} \left| T_V(x+\xi_0) \right|^p \right)^{1/p} + \|T_V\|_{L_p(\omega)} \\
&\le \mathcal{D} + \|T_V\|_{L_p(\omega)} \\
&\le h \left\| T_V' \right\|_{L_p(\omega)} + \|T_V\|_{L_p(\omega)}.
\end{aligned}$$

We bound the first term on the right-hand side further by applying Bernstein's inequality to the trigonometric polynomial T_V of degree N (see, Nikol'skiĭ [144], p. 115):

$$\left\| T_V' \right\|_{L_p(\omega)} \le N \|T_V\|_{L_p(\omega)},$$

and noting that $hN = \pi$. Hence the required result for $n = 1$. □

After this brief preparation, we are now ready to discuss a discrete counterpart of the Marcinkiewicz multiplier theorem, Theorem 1.75, due to Mokin [140] (see also Samarskiĭ, Lazarov, Makarov [160]), which will be our main tool in the stability analysis of finite difference schemes in discrete L_p norms. In order to state it, we require the notion of *total variation*. For a $2N$-periodic function a defined on $\mathbb{Z}^n$, the total variation of a over $\mathbb{I}^n$ is defined by

$$\mathrm{var}(a) := \sup_{k\in\mathbb{Z}^n} \max_{0\ne\alpha\in\{0,1\}^n} \sum_{\nu}^{\alpha} \left| \Delta^\alpha a(\nu) \right|.$$

Here $\varDelta^\alpha := \varDelta_1^{\alpha_1} \cdots \varDelta_n^{\alpha_n}$, as in Theorem 1.75, and, for $\alpha \in \{0, 1\}^n$, we have used the multi-index notation

$$\sum_\nu^\alpha := \sum_{\nu_1}^{\alpha_1} \cdots \sum_{\nu_n}^{\alpha_n}$$

where now, in contrast with the notational convention in Theorem 1.75,

$$\sum_{\nu_j}^{\alpha_j} := \begin{cases} \max_{\nu_j = \pm 2^{|k_j|-1}, \dots, \pm 2^{|k_j|}-1 \text{ such that } \nu_j \in \mathbb{I}} & \text{if } \alpha_j = 0, \\ \sum_{\nu_j = \pm 2^{|k_j|-1}, \dots, \pm 2^{|k_j|}-1 \text{ such that } \nu_j \in \mathbb{I}} & \text{if } \alpha_j = 1. \end{cases}$$

In order to distinguish the total variation of a $2N$-periodic function over $\mathbb{I}^n$ defined here from total variation of a function on $\mathbb{Z}^n$ (as in the statement of the Marcinkiewicz multiplier theorem, Theorem 1.75, stated in the previous chapter), we have used the symbol 'var' here instead of our earlier notation 'Var'. The set of $k \in \mathbb{Z}^n$ for which the index set of $\sum_\nu^\alpha$ is nonempty is finite. Therefore, 'sup' in the definition of $\mathrm{var}(a)$ can be replaced with 'max'.

Theorem 2.49 (Discrete Marcinkiewicz Multiplier Theorem) *Let a be a $2N$-periodic function defined on $\mathbb{Z}^n$, and suppose that one of the following two conditions holds*:

(a) *a is a bounded function on $\mathbb{I}^n$ with bounded variation*; *i.e. there exists a constant M_0 such that*

$$\max_{k \in \mathbb{I}^n} \left| a(k) \right| \le M_0, \qquad \mathrm{var}(a) \le M_0;$$

(b) *a can be extended to a function, still denoted by a, which is defined and continuous on $[-N+1, N]^n$, with $\partial^\alpha a \in C([-N+1, N]^n \setminus \mathbb{I}^n)$ for every multi-index $\alpha \in \{0, 1\}^n$, and such that $\xi^\alpha \partial^\alpha a(\xi)$ is bounded for every $\alpha \in \{0, 1\}^n$; i.e. there exists a constant M_0 such that*

$$\max_{\alpha \in \{0,1\}^n} \sup_{\xi \in [-N+1, N]^n \setminus \mathbb{I}^n} \left| \xi^\alpha \partial^\alpha a(\xi) \right| \le M_0.$$

Then, a is a discrete Fourier multiplier on $L_p(\omega^h)$, $1 < p < \infty$; that is,

$$\left\| \mathcal{F}^{-1}(a \mathcal{F} V) \right\|_{L_p(\omega^h)} \le C \|V\|_{L_p(\omega^h)},$$

for all V in $L_p(\omega^h)$, where $C = C_p M_0$ and C_p is a positive constant, independent of a, h and V.

A simple sufficient condition for $\mathrm{var}(a) \le M_0$ in part (a) of this theorem is that $\mathrm{var}_*(a) \le M_0$, where $\mathrm{var}_*(a)$ is defined analogously to $\mathrm{var}(a)$, except that $\sum_{\nu_j}^{\alpha_j}$ is defined as $\max_{\nu_j \in \mathbb{I}}$ when $\alpha_j = 0$ and as $\sum_{\nu_j \in \mathbb{I}}$ when $\alpha_j = 1$. As there is then no dependence on the diadic sets $\{\pm(2^{|k_j|} - 1), \dots, \pm(2^{|k_j|} - 1)\}$, the symbol $\sup_{k \in \mathbb{Z}^n}$ can be omitted from the definition of $\mathrm{var}_*(a)$.

The proof of the theorem relies on the following result.

Lemma 2.50 *Let $\omega = \mathbb{T}^n := (-\pi, \pi)^n$.*

1. *Suppose that $a(k)$ satisfies the hypotheses in part* (a) *of Theorem 2.49. Then, the sequence $\{\tilde{a}(k)\}_{k\in\mathbb{Z}^n}$ defined by*

$$\tilde{a}(k) = \begin{cases} a(k) & \text{for } k \in \mathbb{I}^n, \\ 0 & \text{otherwise}, \end{cases}$$

is a Fourier multiplier on $L_p(\omega)$, $1 < p < \infty$.

2. *Consider the sequence $\{\tilde{b}(k)\}_{k\in\mathbb{Z}^n}$ defined by $\tilde{b}(k) = b(k_1)\cdots b(k_n)$, with*

$$b(m) = \begin{cases} 1 & \text{if } m = 0, \\ \frac{mh/2}{\sin(mh/2)} & \text{if } m \in \mathbb{I}\setminus\{0\}, \\ \pi/2 & \text{otherwise}. \end{cases}$$

Then, $\{\tilde{b}(k)\}_{k\in\mathbb{Z}^n}$ is a Fourier multiplier on $L_p(\omega)$, $1 < p < \infty$.

3. *The sequence $\{\tilde{a}(k)\tilde{b}(k)\}_{k\in\mathbb{Z}^n}$ is a Fourier multiplier on $L_p(\omega)$, $1 < p < \infty$.*

Proof The proof of this lemma is straightforward and proceeds as follows.

1. The stated result is obtained by noting that

$$\sup_{k\in\mathbb{Z}^n} \left|\tilde{a}(k)\right| = \max_{k\in\mathbb{I}^n}\left|a(k)\right| \le M_0,$$

and

$$\mathrm{Var}(\tilde{a}) \le \max\left\{\max_{k\in\mathbb{I}^n}\left|a(k)\right|, \mathrm{var}(a)\right\} \le M_0 =: M_0(a),$$

and by applying Theorem 1.75 to the sequence $\tilde{a} = \{\tilde{a}(k)\}_{k\in\mathbb{Z}^n}$.

2. The result is proved by noting that

$$\sup_{k\in\mathbb{Z}^n} \left|\tilde{b}(k)\right| \le \left(\frac{\pi}{2}\right)^n,$$

and

$$\mathrm{Var}(\tilde{b}) \le \left(\frac{\pi^2}{2}\right)^n =: M_0(b),$$

and applying Theorem 1.75 to the sequence $\tilde{b} = \{\tilde{b}(k)\}_{k\in\mathbb{Z}^n}$.

3. The stated result follows by observing that

$$\sup_{k\in\mathbb{Z}^n} \left|\tilde{a}(k)\tilde{b}(k)\right| \le \left(\frac{\pi}{2}\right)^n \max_{k\in\mathbb{I}^n}\left|a(k)\right| \le M_0(a)M_0(b),$$

and

$$\mathrm{Var}(\tilde{a}\tilde{b}) \le 2^n M_0(a)M_0(b) = \pi^{2n} M_0(a) =: M_0(ab),$$

and applying Theorem 1.75 to the sequence $\tilde{a}\tilde{b} = \{\tilde{a}(k)\tilde{b}(k)\}_{k\in\mathbb{Z}^n}$. □

We are now ready to prove Theorem 2.49.

Proof of Theorem 2.49 (a) Let us suppose that u is defined on ω^h, and consider its piecewise constant extension w to $\mathbb{R}^n$, defined as follows:

$$w(x) := \begin{cases} u(y), & \text{for } \|x - y\|_\infty < h/2,\ y \in \mathbb{I}^n, \\ 2\pi\text{-periodically extended to } \mathbb{R}^n, \end{cases}$$

where $\|\cdot\|_\infty$ denotes the norm on $\mathbb{R}^n$ defined by $\|x\|_\infty := \max_{1\le j\le n} |x_j|$. Clearly,

$$\|w\|_{L_p(\omega)} = \|u\|_{L_p(\omega^h)}, \quad \omega = \mathbb{T}^n := (-\pi, \pi)^n.$$

Furthermore, with the same notational conventions as in Sect. 1.9.5.1, w has the Fourier series expansion

$$w(x) = \frac{1}{(2\pi)^n} \sum_{k\in\mathbb{Z}^n} \hat{w}(k) \mathrm{e}^{\imath x\cdot k}, \quad x \in \omega,$$

with Fourier coefficients

$$\hat{w}(k) = \int_\omega w(x) \mathrm{e}^{-\imath x\cdot k}\,\mathrm{d}x = \tilde{c}(k) h^n \sum_{x\in\omega^h} u(x) \mathrm{e}^{-\imath x\cdot k},$$

where $\tilde{c}(k) = c(k_1)\cdots c(k_n)$ and

$$c(m) = \begin{cases} 1 & \text{if } m = 0, \\ \frac{\sin(mh/2)}{mh/2} & \text{if } m \in \mathbb{Z}\setminus\{0\}. \end{cases}$$

By noting from Lemma 2.50, part (2), that $c(k) = 1/b(k)$ for $k \in \mathbb{I}$ and therefore $\tilde{c}(k) = 1/\tilde{b}(k)$ for $k \in \mathbb{I}^n$, we have that

$$\hat{w}(k) = \tilde{c}(k)(\mathcal{F}u)(k) = \frac{1}{\tilde{b}(k)}(\mathcal{F}u)(k) \quad \text{for } k \in \mathbb{I}^n.$$

Now, the trigonometric polynomial of degree N defined by

$$T_V : x \in \omega \mapsto \frac{1}{(2\pi)^n} \sum_{k\in\mathbb{I}^n} a(k)(\mathcal{F}u)(k) \mathrm{e}^{\imath x\cdot k}, \quad x \in (-\pi, \pi]^n,$$

is the trigonometric interpolant of the mesh-function $V := \mathcal{F}^{-1}(a\mathcal{F}u)$ defined on ω^h. Therefore, by Lemma 2.48, we have that

$$\begin{aligned} \left\|\mathcal{F}^{-1}(a\mathcal{F}u)\right\|_{L_p(\omega^h)} &\le (1+\pi)^n \left\| \frac{1}{(2\pi)^n} \sum_{k\in\mathbb{I}^n} a(k)(\mathcal{F}u)(k) \mathrm{e}^{\imath x\cdot k} \right\|_{L_p(\omega)} \\ &= (1+\pi)^n \left\| \frac{1}{(2\pi)^n} \sum_{k\in\mathbb{I}^n} a(k)\tilde{b}(k)\hat{w}(k) \mathrm{e}^{\imath x\cdot k} \right\|_{L_p(\omega)} \end{aligned}$$

$$= (1+\pi)^n \left\| \frac{1}{(2\pi)^n} \sum_{k\in\mathbb{Z}^n} \tilde{a}(k)\tilde{b}(k)\hat{w}(k)\mathrm{e}^{\imath x\cdot k} \right\|_{L_p(\omega)}$$
$$= (1+\pi)^n \left\| (\tilde{a}\tilde{b}\hat{w})^\vee \right\|_{L_p(\omega)}.$$

Here $\hat{\cdot}$ and $\cdot^\vee$ denote the Fourier transform of a periodic distribution and its inverse transform, defined in Sect. 1.9.5. Finally, by recalling from Lemma 2.50, part (3), that the sequence $\{\tilde{a}(k)\tilde{b}(k)\}_{k\in\mathbb{Z}^n}$ is a Fourier multiplier on $L_p(\omega)$ it follows that

$$\left\| \mathcal{F}^{-1}(a\mathcal{F}u) \right\|_{L_p(\omega^h)} \le (1+\pi)^n C_p M_0(ab) \|w\|_{L_p(\omega)}$$
$$= (1+\pi)^n C_p M_0(ab) \|u\|_{L_p(\omega^h)},$$

where C_p is as in Theorem 1.75 and $M_0(ab) = \pi^{2n} M_0(a)$, as in the proof of Lemma 2.50. Thus we have shown that

$$\left\| \mathcal{F}^{-1}(a\mathcal{F}u) \right\|_{L_p(\omega^h)} \le C_1 M_0 \|u\|_{L_p(\omega^h)},$$

where $C_1 = (1+\pi)^n \pi^{2n} C_p$ is a positive constant and $M_0 = M_0(a)$ is the constant from the statement of the theorem.

(b) This is a direct consequence of part (a), using the mean-value theorem in those variables x_j for which $\alpha_j = 1$ for a certain $\alpha \in \{0, 1\}^n$. □

We shall now prove the converse of the inequality stated in Lemma 2.48, which will be required in our subsequent considerations.

Lemma 2.51 *Suppose that V is a mesh-function defined on ω^h, and let T_V be its trigonometric interpolant defined by (2.145). Then, for $1 < p < \infty$, there exists a positive constant C_p, independent of h and V, such that*

$$\|T_V\|_{L_p(\omega)} \le C_p \|V\|_{L_p(\omega^h)}.$$

Proof We shall prove this result in one dimension ($n = 1$); the case of $n > 1$ is dealt with by induction over n, starting from $n = 1$. In the proof of Lemma 2.48 we showed that there exists a ξ_0 in the interval $(-h_0, 0)$ such that

$$\|T_V\|_{L_p(\omega)} = \left(h \sum_{x\in\omega^h} \left| T_V(x+\xi_0) \right|^p \right)^{1/p} = \left\| T_V(\cdot + \xi_0) \right\|_{L_p(\omega^h)}$$
$$= \left\| \sum_{k\in\mathbb{I}} (\mathcal{F}V)(k)\mathrm{e}^{\imath xk}\mathrm{e}^{\imath\xi_0 k} \right\|_{L_p(\omega^h)}. \tag{2.148}$$

Next we shall prove that the sequence $\{\lambda(k)\}_{k\in\mathbb{I}}$, with $\lambda(k) := \mathrm{e}^{\imath\xi_0 k}$, is a discrete Fourier multiplier on $L_p(\omega^h)$. First, note that $|\mathrm{e}^{\imath\xi_0 k}| = 1$; furthermore,

$$\sum_{k=-N+1}^{N} \left|\mathrm{e}^{\imath\xi_0 k} - \mathrm{e}^{\imath\xi_0 (k-1)}\right| = \sum_{k=-N+1}^{N} \left|1 - \mathrm{e}^{-\imath\xi_0}\right|$$
$$\leq 2N|\xi_0| \leq 2Nh = 2\pi.$$

Hence, $\mathrm{var}(\lambda) \leq 2\pi$ and, by Theorem 2.49, $\{\lambda(k)\}_{k\in\mathbb{I}}$ is a discrete Fourier multiplier on $L_p(\omega^h)$. Thanks to (2.148) we then have that

$$\|T_V\|_{L_p(\omega)} = \left\|\mathcal{F}^{-1}(\lambda\mathcal{F}V)\right\|_{L_p(\omega^h)} \leq 2\pi C_p \|V\|_{L_p(\omega^h)},$$

where C_p is a positive constant, and hence the required result (with the constant $2\pi C_p$ relabelled as C_p). $\square$

After this interlude on discrete Fourier multipliers, we are ready to embark on the error analysis of finite difference approximations to our elliptic model problem in discrete L_p spaces.

2.5.2 The Model Problem and Its Approximation

Suppose that $\Omega = (0,\pi)^2$. For $f \in W_2^{-1}(\Omega)$, we consider the homogeneous Dirichlet boundary-value problem

$$-\Delta u = f \quad \text{in } \Omega, \tag{2.149}$$
$$u = 0 \quad \text{on } \Gamma = \partial\Omega. \tag{2.150}$$

Throughout the section we shall suppose that the unique weak solution $u \in \mathring{W}_2^1(\Omega)$ of (2.149), (2.150) belongs to $W_p^s(\Omega)$ for some $s \geq 0$ and $p \in (1,\infty)$ (other than $s = 1$ and $p = 2$, of course).

For a nonnegative integer $N \geq 2$ let $h := \pi/N$, and define the meshes:

$$\Omega^h := \left\{(x_i, y_j) : x_i = ih,\ y_j = jh,\ 1 \leq i, j \leq N-1\right\},$$
$$\overline{\Omega}^h := \left\{(x_i, y_j) : x_i = ih,\ y_j = jh,\ 0 \leq i, j \leq N\right\},$$
$$\Gamma^h := \overline{\Omega}^h \setminus \Omega^h.$$

In addition to these, we shall also require the following meshes:

$$\Gamma_x^h := \Gamma^h \cap \big(\{0,\pi\} \times (0,\pi)\big),$$
$$\Gamma_y^h := \Gamma^h \cap \big((0,\pi) \times \{0,\pi\}\big),$$

$$\Gamma_+^h := \Gamma^h \cap \big(\{\pi\} \times (0,\pi) \cup (0,\pi) \times \{\pi\}\big),$$
$$\Omega_+^h := \Omega^h \cup \Gamma_+^h,$$
$$\Omega_x^h := \Omega^h \cup \big(\Gamma_+^h \cap \Gamma_x^h\big),$$
$$\Omega_y^h := \Omega^h \cup \big(\Gamma_+^h \cap \Gamma_y^h\big).$$

As before, we approximate the Laplace operator $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ by

$$D_x^+ D_x^- + D_y^+ D_y^-.$$

Since f has not been assumed to be a continuous function on Ω, we shall mollify it before sampling it at the mesh-points. To do so, we shall use the mollifier $T^\nu = T_h^\nu$ with $\nu = (\nu_1, \nu_2)$ and $h = \pi/N$, defined in (1.35); for the sake of notational simplicity, we shall write $T_h^{\nu_1\nu_2}$, or simply $T^{\nu_1\nu_2}$, instead of the more cumbersome symbol $T_h^{(\nu_1,\nu_2)}$.

First we shall suppose that the weak solution of the boundary-value problem (2.149), (2.150) belongs to $W_p^s(\Omega)$, $s > 2/p$, $1 < p < \infty$; then, by Sobolev's embedding theorem, u is almost everywhere on $\overline{\Omega}$ equal to a continuous function on $\overline{\Omega}$, and

$$\left(T_h^{20} \frac{\partial^2 u}{\partial x^2}\right)(x,y) = D_x^+ D_x^- u(x,y), \quad (x,y) \in \Omega^h,$$
$$\left(T_h^{02} \frac{\partial^2 u}{\partial y^2}\right)(x,y) = D_y^+ D_y^- u(x,y), \quad (x,y) \in \Omega^h.$$

Therefore,

$$-\big(D_x^+ D_x^- T_h^{02} + D_y^+ D_y^- T_h^{20}\big)u = T_h^{22} f \quad \text{on } \Omega^h, \tag{2.151}$$
$$u = 0 \quad \text{on } \Gamma^h. \tag{2.152}$$

This identity motivates us to consider the difference scheme

$$-\big(D_x^+ D_x^- + D_y^+ D_y^-\big)U = T_h^{22} f \quad \text{on } \Omega^h, \tag{2.153}$$
$$U = 0 \quad \text{on } \Gamma^h. \tag{2.154}$$

The rest of this section is devoted to the error analysis of the finite difference scheme (2.153), (2.154). First we introduce the natural discrete analogues of the L_p spaces on Ω^h.

A function V defined on Ω^h (or on $\overline{\Omega}^h$ and equal to zero on Γ^h) is said to belong to $L_p(\Omega^h)$, $1 < p < \infty$, if there exists a positive constant M, independent of h, such that

$$\|V\|_{L_p(\Omega^h)} := \left(h^2 \sum_{(x,y)\in\Omega^h} \big|V(x,y)\big|^p\right)^{1/p} \le M.$$

If V is defined on Ω^h_+ (or on $\overline{\Omega}^h$ and equal to zero on $\Gamma^h \setminus \Gamma^h_+$), the norm $\|\cdot\|_{L_p(\Omega^h)}$ is replaced by

$$\|V\|_{L_p(\Omega^h_+)} := \left(h^2 \sum_{(x,y)\in\Omega^h_+} \left|V(x,y)\right|^p\right)^{1/p}.$$

For mesh-functions defined on Ω^h_x and Ω^h_y the norms $\|\cdot\|_{L_p(\Omega^h_x)}$ and $\|\cdot\|_{L_p(\Omega^h_y)}$ are defined analogously.

The discrete analogues of the Sobolev norms $W^1_p(\Omega)$ and $W^2_p(\Omega)$ are defined, respectively, by

$$\|V\|_{W^1_p(\Omega^h)} := \left(\|V\|^p_{L_p(\Omega^h)} + |V|^p_{W^1_p(\Omega^h)}\right)^{1/p},$$

where

$$|V|_{W^1_p(\Omega^h)} := \left(\left\|D^-_x V\right\|^p_{L_p(\Omega^h_x)} + \left\|D^-_y V\right\|^p_{L_p(\Omega^h_y)}\right)^{1/p};$$

and

$$\|V\|_{W^2_p(\Omega^h)} := \left(\|V\|^p_{W^1_p(\Omega^h)} + |V|^p_{W^2_p(\Omega^h)}\right)^{1/p},$$

where

$$|V|_{W^2_p(\Omega^h)} := \left(\left\|D^+_x D^-_x V\right\|^p_{L_p(\Omega^h)} + \left\|D^-_x D^-_y V\right\|^p_{L_p(\Omega^h_+)} + \left\|D^+_y D^-_y V\right\|^p_{L_p(\Omega^h)}\right)^{1/p}.$$

Let us recall the notion of discrete Fourier transform from the previous section. However, as we are now working on $(0,\pi)^2$ rather than $(-\pi,\pi)^2$ and the functions we shall be dealing with will satisfy a homogeneous Dirichlet boundary condition rather then a periodic boundary condition, some adjustments have to be made before the techniques developed in the previous section can be applied.

Suppose that V is defined on Ω^h (or on $\overline{\Omega}^h$ and equal to zero on Γ^h). We shall consider the *odd extension* $\tilde{V}$ of the mesh-function V to the mesh

$$\omega^h = h\mathbb{I}^2 = \left\{(x_i, y_j) : x_i = ih, y_j = jh,\ i,j = -N+1,\dots,N\right\}$$

contained in $(-\pi,\pi]^2$. Thus

$$\tilde{V}(-x,y) = -\tilde{V}(x,y) \quad \text{and} \quad \tilde{V}(x,-y) = -\tilde{V}(x,y) \quad \text{for all } (x,y) \text{ in } \Omega^h.$$

After such an extension, $\tilde{V}$ is further extended 2π-periodically in each co-ordinate direction to the whole of $h\mathbb{Z}^2$. Let us note that

$$\|\tilde{V}\|_{L_p(\omega^h)} = 4^{1/p}\|V\|_{L_p(\Omega^h)}. \tag{2.155}$$

Lemma 2.52 *Let us suppose that V is defined on Ω^h (or on $\overline{\Omega}^h$ and equal to zero on Γ^h), and consider its odd extension $\tilde{V}$. The discrete Fourier transform $\mathcal{F}\tilde{V}$ has the following properties*:

1. *For any $k = (k_1, k_2) \in \mathbb{I}^2$,*

$$\mathcal{F}\tilde{V}(k_1, k_2) = -4h^2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} V(x_i, y_j) \sin(k_1 x_i) \sin(k_2 y_j);$$

2. *$\mathcal{F}\tilde{V}$ is an odd function on $\mathbb{I}^2$; that is,*

$$\mathcal{F}\tilde{V}(-k_1, k_2) = -\mathcal{F}\tilde{V}(k_1, k_2) \quad \textit{and} \quad \mathcal{F}\tilde{V}(k_1, -k_2) = -\mathcal{F}\tilde{V}(k_1, k_2)$$

   for all $k = (k_1, k_2) \in \mathbb{I}^2$. Also, $\mathcal{F}\tilde{V}(0, k_2) = \mathcal{F}\tilde{V}(k_1, 0) = \mathcal{F}\tilde{V}(0, 0) = 0$;
3. *For $1 \le i, j \le N - 1$,*

$$V(x_i, y_j) = -\frac{1}{\pi^2} \sum_{k_1=1}^{N-1} \sum_{k_2=1}^{N-1} \mathcal{F}\tilde{V}(k_1, k_2) \sin(k_1 x_i) \sin(k_2 y_j).$$

The proof of this result is elementary and is left to the reader.

Lemma 2.52 implies that the values of $\mathcal{F}\tilde{V}$ on $\mathbb{I}^2$ are completely determined by the values of V on Ω^h; conversely, V can be completely characterized on Ω^h (and $\tilde{V}$ on ω^h) by the values $\mathcal{F}\tilde{V}(k_1, k_2)$, $k_1, k_2 = 1, \ldots, N-1$. Consequently, it is meaningful to consider the *discrete Fourier sine-transform* $\mathcal{F}_\sigma V$ of a mesh-function V defined on Ω^h (or on $\overline{\Omega}^h$ and equal to zero on Γ^h). Indeed, we let

$$\mathcal{F}_\sigma V := -\frac{1}{4} \mathcal{F}\tilde{V},$$

and, for a function W defined on the set $\{(i, j) : 1 \le i, j \le N-1\}$ with odd extension $\tilde{W}$ to $\mathbb{I}^2$, we put

$$\mathcal{F}_\sigma^{-1} W := -4\mathcal{F}^{-1}\tilde{W}.$$

Thus,

$$\mathcal{F}_\sigma V(k_1, k_2) = h^2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} V(x_i, y_j) \sin(k_1 x_i) \sin(k_2 y_j)$$

and

$$\mathcal{F}_\sigma^{-1} W(x, y) = \left(\frac{2}{\pi}\right)^2 \sum_{k_1=1}^{N-1} \sum_{k_2=1}^{N-1} W(k_1, k_2) \sin(k_1 x) \sin(k_2 y).$$

In order to derive error bounds for the finite difference scheme under consideration we shall need the following stability result.

Lemma 2.53 *Suppose that η_1 and η_2 are two functions defined on $\overline{\Omega}^h$ that vanish on Γ^h. Further, let e be the solution to the problem*

$$-\big(D_x^+ D_x^- + D_y^+ D_y^-\big)e = D_x^+ D_x^- \eta_1 + D_y^+ D_y^- \eta_2 \quad \text{in } \Omega^h, \tag{2.156}$$

$$e = 0 \quad \text{on } \Gamma^h. \tag{2.157}$$

Then, for any $p \in (1,\infty)$,

$$\|e\|_{L_p(\Omega^h)} \le C_p\big(\|\eta_1\|_{L_p(\Omega^h)} + \|\eta_2\|_{L_p(\Omega^h)}\big), \tag{2.158}$$

$$|e|_{W_p^1(\Omega^h)} \le C_p\big(\big\|D_x^- \eta_1\big\|_{L_p(\Omega_x^h)} + \big\|D_y^- \eta_2\big\|_{L_p(\Omega_y^h)}\big), \tag{2.159}$$

$$|e|_{W_p^2(\Omega^h)} \le C_p\big(\big\|D_x^+ D_x^- \eta_1\big\|_{L_p(\Omega^h)} + \big\|D_y^+ D_y^- \eta_2\big\|_{L_p(\Omega^h)}\big), \tag{2.160}$$

where C_p is a positive constant, independent of h, e, η_1 and η_2.

Proof (1) Let us first prove (2.158). As

$$\mathcal{F}_\sigma\big(D_x^+ D_x^- e\big) = -\lambda_1^2 \mathcal{F}_\sigma e \quad \text{and} \quad \mathcal{F}_\sigma\big(D_y^+ D_y^- e\big) = -\lambda_2^2 \mathcal{F}_\sigma e,$$

where

$$\lambda_1 = \lambda_1(k_1) := \frac{2}{h}\sin\frac{k_1 h}{2} \quad \text{and} \quad \lambda_2 = \lambda_1(k_2) := \frac{2}{h}\sin\frac{k_2 h}{2},$$

with $k := (k_1, k_2)$, $1 \le k_1, k_2 \le N-1$, it follows that

$$e = \mathcal{F}_\sigma^{-1}(a_1 \mathcal{F}_\sigma \eta_1) + \mathcal{F}_\sigma^{-1}(a_2 \mathcal{F}_\sigma \eta_2),$$

where

$$a_l(k_1, k_2) := \frac{\lambda_l^2(k_l)}{\lambda_1^2(k_1) + \lambda_2^2(k_2)}, \quad 1 \le k_1, k_2 \le N-1,\ l = 1, 2.$$

We note that $a_1(k_1, k_2)$ and $a_2(k_1, k_2)$ can be defined for all $k \in \mathbb{I}^2 \setminus \{0\}$ by letting

$$a_l(-k_1, k_2) := a_l(k_1, k_2),$$
$$a_l(k_1, -k_2) := a_l(k_1, k_2),$$
$$a_l(-k_1, -k_2) := a_l(k_1, k_2),$$

for all $k = (k_1, k_2)$, $1 \le k_1, k_2 \le N-1$, $l = 1, 2$.

Let $\tilde{e}$, $\tilde{\eta}_1$ and $\tilde{\eta}_2$ denote the odd extensions of the mesh-functions e, η_1 and η_2, respectively, from $\overline{\Omega}_h$ to ω^h. Then,

$$\tilde{e} = \mathcal{F}^{-1}(a_1 \mathcal{F}\tilde{\eta}_1) + \mathcal{F}^{-1}(a_2 \mathcal{F}\tilde{\eta}_2).$$

The fact that a_1 and a_2 are not defined at $(0,0)$ is of no significance, since

$$(\mathcal{F}\tilde{\eta}_l)(0,0) = h^2 \sum_{i=-N+1}^{N} \sum_{j=-N+1}^{N} \tilde{\eta}_i(x_i, y_j) = 0, \quad l = 1, 2,$$

which follows from our assumption that η_l, $l = 1, 2$, vanish on Γ^h, by noting that $\tilde{\eta}_l$ is the odd extension of η_l.

Hence, by the triangle inequality,

$$\|\tilde{e}\|_{L_p(\omega^h)} \le \left\|\mathcal{F}^{-1}(a_1 \mathcal{F}\tilde{\eta}_1)\right\|_{L_p(\omega^h)} + \left\|\mathcal{F}^{-1}(a_2 \mathcal{F}\tilde{\eta}_2)\right\|_{L_p(\omega^h)}.$$

Next we show that a_1 and a_2 are discrete Fourier multipliers on $L_p(\omega^h)$.

Clearly $0 \le a_1 \le 1$ on $\mathbb{I}^2$. Further, as $a_1 + a_2 = 1$,

$$x \frac{\partial a_1}{\partial x} = 2a_1(1 - a_1) \frac{xh}{2} \cot \frac{xh}{2}.$$

Thus, noting that $|t \cot t| \le 1$ for $|t| \le \pi/2$, we have that

$$\left| x \frac{\partial a_1}{\partial x}(x, y) \right| \le \frac{1}{2} \quad \text{for } (x, y) \in \mathbb{I}^2.$$

Similarly, noting again that $a_1 + a_2 = 1$,

$$y \frac{\partial a_1}{\partial y} = -2a_1(1 - a_1) \frac{yh}{2} \cot \frac{yh}{2}.$$

Therefore,

$$\left| y \frac{\partial a_1}{\partial y}(x, y) \right| \le \frac{1}{2} \quad \text{for } (x, y) \in \mathbb{I}^2.$$

Finally,

$$xy \frac{\partial^2 a_1}{\partial x \partial y} = 4\left(y \frac{\partial a_1}{\partial y} \right) \frac{xh}{2} \cot \frac{xh}{2},$$

and so,

$$\left| xy \frac{\partial^2 a_1}{\partial x \partial y}(x, y) \right| \le 2 \quad \text{for } (x, y) \in \mathbb{I}^2.$$

Hence, by Theorem 2.49, a_1 is a discrete Fourier multiplier on $L_p(\omega^h)$. By symmetry, the same is true of a_2.

Therefore,

$$\|\tilde{e}\|_{L_p(\omega^h)} \le C_p \big(\|\tilde{\eta}_1\|_{L_p(\omega^h)} + \|\tilde{\eta}_2\|_{L_p(\omega^h)} \big),$$

from which (2.158) immediately follows by noting (2.155).

(2) As we have seen in part (1),

$$\mathcal{F}\tilde{e} = a_1 \mathcal{F}\tilde{\eta}_1 + a_2 \mathcal{F}\tilde{\eta}_2.$$

Multiplying this identity by $(1 - \exp(-\imath k_1 h))/h$ we deduce that

$$D_x^- \tilde{e} = \mathcal{F}^{-1}\big(a_1 \mathcal{F}\big(D_x^- \tilde{\eta}_1\big)\big) + \mathcal{F}^{-1}\big(b_2 \mathcal{F}\big(D_y^- \tilde{\eta}_2\big)\big),$$

where

$$b_2(k_1, k_2) := a_2(k_1, k_2) \frac{1 - \mathrm{e}^{-\imath k_1 h}}{1 - \mathrm{e}^{-\imath k_2 h}}.$$

We have already shown in part (1) that a_1 and a_2 are discrete Fourier multipliers on $L_p(\omega^h)$. Similarly, using Theorem 2.49 we deduce that the same is true of $(1 - \mathrm{e}^{-\imath k_1 h})/(1 - \mathrm{e}^{-\imath k_2 h})$, and therefore of b_2. Hence,

$$\big\| D_x^- \tilde{e} \big\|_{L_p(\omega^h)} \le C_p \big(\big\| D_x^- \tilde{\eta}_1 \big\|_{L_p(\omega^h)} + \big\| D_y^- \tilde{\eta}_2 \big\|_{L_p(\omega^h)} \big),$$

which yields

$$\big\| D_x^- e \big\|_{L_p(\Omega_x^h)} \le C_p \big(\big\| D_x^- \eta_1 \big\|_{L_p(\Omega_x^h)} + \big\| D_y^- \eta_2 \big\|_{L_p(\Omega_y^h)} \big).$$

An identical bound holds for $\| D_y^- e \|_{L_p(\Omega_y^h)}$, which, when added to the last inequality, yields (2.159).

(3) To prove (2.160), we note, by recalling the definitions of a_1 and a_2 from part (1) of the proof, that

$$-\lambda_1^2 \mathcal{F}\tilde{e} = \frac{\lambda_1^2}{\lambda_1^2 + \lambda_2^2} \big(-\lambda_1^2 \mathcal{F}\tilde{\eta}_1\big) + \frac{\lambda_1^2}{\lambda_1^2 + \lambda_2^2} \big(-\lambda_2^2 \mathcal{F}\tilde{\eta}_2\big).$$

Thus,

$$\mathcal{F}\big(D_x^+ D_x^- \tilde{e}\big) = \frac{\lambda_1^2}{\lambda_1^2 + \lambda_2^2} \mathcal{F}\big(D_x^+ D_x^- \tilde{\eta}_1\big) + \frac{\lambda_1^2}{\lambda_1^2 + \lambda_2^2} \mathcal{F}\big(D_y^+ D_y^- \tilde{\eta}_2\big).$$

Equivalently,

$$D_x^+ D_x^- \tilde{e} = \mathcal{F}^{-1}\left(\frac{\lambda_1^2}{\lambda_1^2 + \lambda_2^2} \mathcal{F}\big(D_x^+ D_x^- \tilde{\eta}_1\big) \right) + \mathcal{F}^{-1}\left(\frac{\lambda_1^2}{\lambda_1^2 + \lambda_2^2} \mathcal{F}\big(D_y^+ D_y^- \tilde{\eta}_2\big) \right).$$

As $\lambda_l^2/(\lambda_1^2 + \lambda_2^2)$, $l = 1, 2$, are discrete Fourier multipliers on $L_p(\omega^h)$, it follows that

$$\big\| D_x^+ D_x^- \tilde{e} \big\|_{L_p(\omega^h)} \le C_p \big(\big\| D_x^+ D_x^- \tilde{\eta}_1 \big\|_{L_p(\omega^h)} + \big\| D_y^+ D_y^- \tilde{\eta}_2 \big\|_{L_p(\omega^h)} \big),$$

which gives

$$\big\| D_x^+ D_x^- e \big\|_{L_p(\Omega^h)} \le C_p \big(\big\| D_x^+ D_x^- \eta_1 \big\|_{L_p(\Omega^h)} + \big\| D_y^+ D_y^- \eta_2 \big\|_{L_p(\Omega^h)} \big).$$

Identical bounds hold for $\|D_y^+ D_y^- e\|_{L_p(\Omega^h)}$ and $\|D_x^- D_y^- e\|_{L_p(\Omega_+^h)}$, which, when added to the last inequality, yield (2.160). $\square$

It is possible to derive bounds analogous to (2.159) and (2.160), but with e measured in a norm rather than a seminorm. To see this, we need the following preliminary result that relates discrete Sobolev seminorms to the corresponding discrete Sobolev norms.

Lemma 2.54 *Suppose that V is a function defined on the mesh $\overline{\Omega}^h$ such that $V = 0$ on Γ^h. Then, the following bounds hold:*

(a) *Assuming that* $1 < p < \infty$,

$$\|V\|_{L_p(\Omega^h)} \leq 2^{-1/p}\pi |V|_{W_p^1(\Omega^h)};$$

(b) *There exists a constant C_p, independent of V and h, such that*

$$|V|_{W_p^1(\Omega^h)} \leq C_p |V|_{W_p^2(\Omega^h)}, \quad 1 < p < \infty;$$

(c) *Assuming that* $1 < p < \infty$,

$$\|V\|_{W_p^1(\Omega^h)} \leq \left(1 + \frac{1}{2}\pi^p\right)^{1/p} |V|_{W_p^1(\Omega^h)};$$

(d) *With C_p denoting the constant from part* (*b*),

$$\|V\|_{W_p^2(\Omega^h)} \leq \left(1 + \left(1 + \frac{1}{2}\pi^p\right)C_p^p\right)^{1/p} |V|_{W_p^2(\Omega^h)}, \quad 1 < p < \infty.$$

Proof Part (c) is a direct consequence of (a), while (d) follows by combining (c) and (b). We note that (c) is a discrete Friedrichs inequality, which generalizes Lemma 2.19. It remains to prove (a) and (b).

(a) As $V = 0$ on Γ^h, we can write

$$V_{ij} = \sum_{k=1}^{i} h D_x^- V_{kj}.$$

By Hölder's inequality for finite sums,

$$|V_{ij}|^p \leq (ih)^{p/q} \sum_{k=1}^{i} h |D_x^- V_{kj}|^p, \quad \text{where } \frac{1}{p} + \frac{1}{q} = 1.$$

Multiplying by h^2, increasing the upper limit in the sum on the right to N, and summing through $i, j = 1, \dots, N-1$, we get that

$$\|V\|_{L_p(\Omega^h)}^p \leq h^p \left(\sum_{i=1}^{N-1} i^{p/q}\right) \|D_x^- V\|_{L_p(\Omega_x^h)}^p.$$

Now

$$\sum_{i=1}^{N-1} i^{p/q} \leq 1 + \int_1^{N-1} x^{p/q}\,\mathrm{d}x = 1 + \frac{(N-1)^{(p/q)+1} - 1}{(p/q)+1} \leq N^{(p/q)+1} = N^p,$$

and therefore, since $Nh = \pi$, we deduce that

$$\|V\|^p_{L_p(\Omega^h)} \leq \pi^p \left\| D_x^- V \right\|^p_{L_p(\Omega^h_x)}.$$

Analogously,

$$\|V\|^p_{L_p(\Omega^h)} \leq \pi^p \left\| D_y^- V \right\|^p_{L_p(\Omega^h_y)}.$$

By adding the last two inequalities we deduce (a).

(b) Let $W := -(D_x^+ D_x^- + D_y^+ D_y^-)V$. Using the same technique and the same notation as in the proof of Lemma 2.53, and observing that

$$\left(\frac{1 - \mathrm{e}^{-\imath k_1 x}}{h}\right)\left(\frac{1}{\lambda_1^2(k_1, k_2) + \lambda_2^2(k_1, k_2)}\right)$$

and

$$\left(\frac{1 - \mathrm{e}^{-\imath k_2 y}}{h}\right)\left(\frac{1}{\lambda_1^2(k_1, k_2) + \lambda_2^2(k_1, k_2)}\right)$$

are discrete Fourier multipliers on $L_p(\omega^h)$, we deduce from Theorem 2.49 that

$$\left\| D_x^- V \right\|_{L_p(\Omega^h_x)} \leq C_p \|W\|_{L_p(\Omega^h)},$$

and

$$\left\| D_y^- V \right\|_{L_p(\Omega^h_y)} \leq C_p \|W\|_{L_p(\Omega^h)}.$$

Hence

$$|V|_{W^1_p(\Omega^h)} \leq 2^{1/p} C_p \|W\|_{L_p(\Omega^h)},$$

and therefore, by the triangle inequality,

$$|V|_{W^1_p(\Omega^h)} \leq 2^{1/p} C_p \big(\left\| D_x^+ D_x^- V \right\|_{L_p(\Omega^h)} + \left\| D_y^+ D_y^- V \right\|_{L_p(\Omega^h)} \big).$$

Thus, by noting the inequality $a + b \leq 2^{1-(1/p)}(a^p + b^p)^{1/p}$ for $a, b \geq 0$,

$$|V|_{W^1_p(\Omega^h)} \leq 2 C_p |V|_{W^2_p(\Omega^h)}.$$

Renaming the constant $2C_p$ into C_p then yields the stated inequality. □

Combining the last two lemmas, we arrive at the following result.

Lemma 2.55 *Suppose that η_1 and η_2 are two functions defined on $\overline{\Omega}^h$ that vanish on Γ^h. Let further e denote the solution of the problem*

$$-\left(D_x^+ D_x^- + D_y^+ D_y^-\right)e = D_x^+ D_x^- \eta_1 + D_y^+ D_y^- \eta_2 \quad \textit{in } \Omega^h, \tag{2.161}$$

$$e = 0 \quad \textit{on } \Gamma^h. \tag{2.162}$$

Then, there exists a positive constant C_p, independent of h, such that

$$\|e\|_{L_p(\Omega^h)} \le C_p\left(\|\eta_1\|_{L_p(\Omega^h)} + \|\eta_2\|_{L_p(\Omega^h)}\right), \tag{2.163}$$

$$\|e\|_{W_p^1(\Omega^h)} \le C_p\left(\left\|D_x^-\eta_1\right\|_{L_p(\Omega_x^h)} + \left\|D_y^-\eta_2\right\|_{L_p(\Omega_y^h)}\right), \tag{2.164}$$

$$\|e\|_{W_p^2(\Omega^h)} \le C_p\left(\left\|D_x^+ D_x^-\eta_1\right\|_{L_p(\Omega^h)} + \left\|D_y^+ D_y^-\eta_2\right\|_{L_p(\Omega^h)}\right). \tag{2.165}$$

Now we are ready to state the main result of this section.

Theorem 2.56 *Let u be the weak solution of the boundary-value problem* (2.149), (2.150), *let U be the solution of the finite difference scheme* (2.153), (2.154) *and suppose that $m \in \{0, 1, 2\}$. Assuming that u belongs to $W_p^s(\Omega)$, with $m \le s$, $2/p < s \le m + 2$, $1 < p < \infty$, the following error bound holds*:

$$\|u - U\|_{W_p^m(\Omega^h)} \le C h^{s-m} |u|_{W_p^s(\Omega)},$$

with a positive constant $C = C(p, m, s)$, independent of h.

Proof (a) Let us first suppose that $m = 2$ and $s \ge 2$. We define the global error e on $\overline{\Omega}^h$ by $e_{ij} := u(x_i, y_j) - U_{ij}$. It follows from (2.151)–(2.154) that e satisfies (2.161), (2.162) with

$$\eta_1 = u - T_h^{02}u \quad \text{and} \quad \eta_2 = u - T_h^{20}u.$$

Now η_1 (resp. η_2) is defined on the mesh $\Omega^h \cup \Gamma_x^h$ (resp. $\Omega^h \cup \Gamma_y^h$) and equal to zero on Γ_x^h (resp. Γ_y^h). According to (2.165), in order to obtain the desired error bound for $m = 2$, it suffices to estimate $\|D_x^+ D_x^- \eta_1\|_{L_p(\Omega^h)}$ and $\|D_y^+ D_y^- \eta_2\|_{L_p(\Omega^h)}$. To do so, we define the squares

$$K_{ij}^0 := (x_{i-1}, x_{i+1}) \times (y_{j-1}, y_{j+1}),$$
$$\tilde{K}^0 := (-1, 1) \times (-1, 1),$$

and consider the affine mapping $(x, y) \in K_{ij}^0 \mapsto (\tilde{x}, \tilde{y}) \in \tilde{K}^0$, where

$$x = x(\tilde{x}) := (i + \tilde{x})h, \qquad y = y(\tilde{y}) := (j + \tilde{y})h.$$

Let $\tilde{u}(\tilde{x}, \tilde{y}) = u(x(\tilde{x}), y(\tilde{y}))$. We then have the following equalities:

$$\begin{aligned}
&\left(D_x^+ D_x^- \eta_1\right)_{ij} \\
&\quad = \frac{u(x_{i+1}, y_j) - 2u(x_i, y_j) + u(x_{i-1}, y_j)}{h^2} \\
&\qquad - \int_{-1}^{1} \theta_2(\tilde{y}) \frac{u(x_{i+1}, y_j + \tilde{y}h) - 2u(x_i, y_j + \tilde{y}h) + u(x_{i-1}, y_j + \tilde{y}h)}{h^2} \, \mathrm{d}\tilde{y} \\
&\quad = \frac{1}{h^2} \bigg\{ \tilde{u}(1,0) - 2\tilde{u}(0,0) + \tilde{u}(-1,0) \\
&\qquad - \int_{-1}^{1} \theta_2(\tilde{y}) \big[\tilde{u}(1,\tilde{y}) - 2\tilde{u}(0,\tilde{y}) + \tilde{u}(-1,\tilde{y})\big] \, \mathrm{d}\tilde{y} \bigg\},
\end{aligned}$$

where $\theta_2(\tilde{y}) = 1 - |\tilde{y}|$, $\tilde{y} \in (-1, 1)$.

Now $(D_x^+ D_x^- \eta_1)_{ij}$ is a bounded linear functional on $W_p^s(\tilde{K}^0)$, $s > 2/p$, whose kernel contains $\mathcal{P}_3(\tilde{K}^0)$. According to the Bramble–Hilbert lemma,

$$\left|\left(D_x^+ D_x^- \eta_1\right)_{ij}\right| \le C h^{-2} |\tilde{u}|_{W_p^s(\tilde{K}^0)}$$

for $2/p < s \le 4$. Thus, by changing from the $(\tilde{x}, \tilde{y})$ to the (x, y) co-ordinate system, we have that

$$\left|\left(D_x^+ D_x^- \eta_1\right)_{ij}\right| \le C h^{-2} h^{s-2/p} |u|_{W_p^s(K_{ij}^0)}$$

for $2/p < s \le 4$. Hence,

$$\left\| D_x^+ D_x^- \eta_1 \right\|_{L_p(\Omega^h)} \le C h^{s-2} |u|_{W_p^s(\Omega)}, \quad 2/p < s \le 4.$$

Likewise,

$$\left\| D_y^+ D_y^- \eta_2 \right\|_{L_p(\Omega^h)} \le C h^{s-2} |u|_{W_p^s(\Omega)}, \quad 2/p < s \le 4,$$

which, after insertion into (2.165), completes the proof for the case $m = 2$.

(b) Let $m = 1$ and $s \ge 1$. By (2.164) it suffices to bound $\|D_x^- \eta_1\|_{L_p(\Omega_x^h)}$ and $\|D_y^- \eta_2\|_{L_p(\Omega_y^h)}$. We proceed in the same way as in part (a) to deduce that

$$\left(D_x^- \eta_1\right)_{ij} = \frac{1}{h} \left\{ \tilde{u}(1,0) - \tilde{u}(0,0) - \int_{-1}^{1} \theta_2(\tilde{y}) \big[\tilde{u}(1,\tilde{y}) - \tilde{u}(0,\tilde{y})\big] \, \mathrm{d}\tilde{y} \right\}$$

is a bounded linear functional on $W_p^s(\tilde{K}^0)$, $s > 2/p$, whose kernel contains $\mathcal{P}_2(\tilde{K}^0)$. Therefore,

$$\left\| D_x^- \eta_1 \right\|_{L_p(\Omega_x^h)} \le C h^{s-1} |u|_{W_p^s(\Omega)}, \quad 2/p < s \le 3,$$

and, similarly,

$$\left\| D_y^- \eta_2 \right\|_{L_p(\Omega_y^h)} \le C h^{s-1} |u|_{W_p^s(\Omega)}, \quad 2/p < s \le 3.$$

Inserting these into (2.164) we obtain the desired error bound for $m = 1$.

(c) Let $m = 0$ and $s \geq 0$. We need to estimate $\|\eta_1\|_{L_p(\Omega^h)}$ and $\|\eta_2\|_{L_p(\Omega^h)}$. Since

$$(\eta_1)_{ij} = \tilde{u}(0,0) - \int_{-1}^{1} \theta_2(\tilde{y})\tilde{u}(0,\tilde{y})\,\mathrm{d}\tilde{y}$$

is a bounded linear functional on $W_p^s(\tilde{K}^0)$, $s > 2/p$, whose kernel contains $\mathcal{P}_1(\tilde{K}^0)$, it follows that

$$\|\eta_1\|_{L_p(\Omega^h)} \leq Ch^s |u|_{W_p^s(\Omega)}, \quad 2/p < s \leq 2,$$

and, likewise,

$$\|\eta_2\|_{L_p(\Omega^h)} \leq Ch^s |u|_{W_p^s(\Omega)}, \quad 2/p < s \leq 2.$$

Substituting these into (2.163) we obtain the desired error bound for $m = 0$. That completes the proof of the theorem. □

In the remainder of this section we shall discuss the rate of convergence of the finite difference scheme (2.153), (2.154) in the case when $0 \leq s < 1 + 1/p$, which also covers the case $0 \leq s \leq 2/p$. Let us define the function space $\tilde{W}_p^s(\Omega)$, $1 < p < \infty$, by

$$\tilde{W}_p^s(\Omega) = \begin{cases} W_p^s(\Omega), & 0 \leq s \leq 1/p, \\ \{w : w \in W_p^s(\Omega),\, w = 0 \text{ on } \Gamma\}, & 1/p < s < 1 + 1/p. \end{cases}$$

We observe that if u, the weak solution of the boundary-value problem (2.149), (2.150) belongs $W_p^s(\Omega)$ then $u \in \tilde{W}_p^s(\Omega)$. Let $\Omega^* := (-\pi, 2\pi) \times (-\pi, 2\pi)$; the extension of u by 0 is a continuous linear operator from $\tilde{W}_p^s(\Omega)$ into $W_p^s(\Omega^*)$, $0 \leq s < 1 + 1/p$, $s \neq 1/p$, $1 < p < \infty$ (cf. Triebel [182], Sect. 2.10.2, Lemma and Remark 1 on p. 227 and Theorem 1 on p. 228). Hence

$$u \mapsto u^* = \text{odd extension of } u$$

is a continuous mapping from $\tilde{W}_p^s(\Omega)$ into $W_p^s(\Omega^*)$, $0 \leq s < 1 + 1/p$, $s \neq 1/p$, $1 < p < \infty$. Moreover, $(T_h^{11} u^*)(x, y) = 0$ for $(x, y) \in \Gamma^h$.

Theorem 2.57 *Let u be the weak solution of the boundary-value problem* (2.149), (2.150), *let U be the solution of the finite difference scheme* (2.153), (2.154) *and suppose that $m \in \{0, 1\}$. Assuming that u belongs to $W_p^s(\Omega)$ with $m \leq s$, $0 \leq s < 1 + 1/p$, $s \neq 1/p$ and $1 < p < \infty$, the following error bound holds*:

$$\left\| T_h^{11} u - U \right\|_{W_p^m(\Omega^h)} \leq Ch^{s-m} |u|_{W_p^s(\Omega)},$$

with a positive constant $C = C(p, m, s)$, independent of h.

Proof The proof is completely analogous to that of Theorem 2.56, except that we now define the global error e on $\overline{\Omega}^h$ by

$$e_{ij} = \left(T_h^{11} u^*\right)(x_i, y_j) - U_{ij}.$$

Clearly $e_{ij} = 0$ for $(x_i, y_j) \in \Gamma^h$, and $e_{ij} = (T_h^{11} u)(x_i, y_j) - U_{ij}$ when $(x_i, y_j) \in \Omega^h$. In addition, it follows from (2.151)–(2.154) that e satisfies (2.156), (2.157) with

$$\eta_1 = T_h^{11} u^* - T_h^{02} u^* \quad \text{and} \quad \eta_2 = T_h^{11} u^* - T_h^{20} u^*.$$

Again, η_1 (resp. η_2) is defined on the mesh $\Omega^h \cup \Gamma_x^h$ (resp. $\Omega^h \cup \Gamma_y^h$) and is equal to zero on Γ_x^h (resp. Γ_y^h). The rest of the proof is the same as in the case of Theorem 2.56, except that now $s \in [0, 1/p) \cup (1/p, 1 + 1/p)$. □

2.5.3 Convergence in Discrete Bessel-Potential Norms

This section is devoted to error estimation in discrete Bessel-potential norms. A function v defined on $\Omega^h \subset (0, \pi)^2$ (or on $\overline{\Omega}^h \subset [0, \pi]^2$ and equal to zero on Γ^h) is said to belong to the discrete Bessel-potential space $H_p^s(\Omega^h)$, with $-\infty < s < \infty$, $1 < p < \infty$, if there exists a function $V \in L_p(\Omega^h)$ such that

$$v = I_{s,h} V := \mathcal{F}_\sigma^{-1}\left(\left(1 + |k|^2\right)^{-s/2} \mathcal{F}_\sigma V\right) = \mathcal{F}^{-1}\left(\left(1 + |k|^2\right)^{-s/2} \mathcal{F} \tilde{V}\right),$$

where $\tilde{V}$ is the odd extension of V from Ω^h to $\omega^h = h\mathbb{I}^2$, defined to be zero on Γ^h, and further extended 2π-periodically to the whole of $h\mathbb{Z}^2$. We then define (compare with the definition in Sect. 1.9.5.3)

$$\|v\|_{H_p^s(\Omega^h)} := \|V\|_{L_p(\Omega^h)} = 4^{-1/p} \|\tilde{V}\|_{L_p(\omega^h)},$$

where the last equality is a consequence of (2.155).

First we shall prove equivalence of the discrete Sobolev norm $\|\cdot\|_{W_p^m(\Omega^h)}$ and the norm $\|\cdot\|_{H_p^m(\Omega^h)}$ for integer m; then, the error bounds in discrete Bessel-potential norms of integer order will follow from the error bounds derived in Theorems 2.56 and 2.57. Error bounds in fractional-order discrete Bessel-potential norms will be derived from these by function space interpolation. We need the following preliminary result in the univariate case.

Lemma 2.58 *Let W be a mesh-function defined on $\omega^h = h\mathbb{I}$, where $\mathbb{I} = \{-N + 1, \dots, N\}$, and let T_W be the trigonometric interpolant of W on $(-\pi, \pi]$ given by* (2.145), *with $n = 1$. Then, there exists a constant C_p, independent of h and W, such that the following inequalities hold, with $\omega = (-\pi, \pi)$:*

(a) $\|D_x^- W\|_{L_p(\omega^h)} \le \|T_W'\|_{L_p(\omega)} \le C_p \|D_x^- W\|_{L_p(\omega^h)}$;

(b) $\|D_x^+ D_x^- W\|_{L_p(\omega^h)} \le \|T_W''\|_{L_p(\omega)} \le C_p \|D_x^+ D_x^- W\|_{L_p(\omega^h)}$.

Proof (a) Since W and T_W coincide at the mesh-points,

$$D_x^- W(x_i) = D_x^- T_W(x_i) = \frac{1}{h}\int_{x_{i-1}}^{x_i} T_W'(x)\,\mathrm{d}x.$$

Thus,

$$h\big|D_x^- W(x_i)\big|^p \le \int_{x_{i-1}}^{x_i} \big|T_W'(x)\big|^p\,\mathrm{d}x.$$

Summing over all x_i in ω^h, we deduce that

$$\big\|D_x^- W\big\|_{L_p(\omega^h)} \le \big\|T_W'\big\|_{L_p(\omega)}.$$

To deduce the second inequality, let us note that

$$T_W'(x) = \frac{1}{2\pi}\sum_{k\in\mathbb{I}} (\imath k)\mathcal{F}W(k)\mathrm{e}^{\imath x k},$$

and

$$\mathcal{F}\big(D_x^- W\big)(k) = \frac{1-\mathrm{e}^{-\imath k h}}{h}\mathcal{F}W(k).$$

Therefore,

$$T_W'(x) = \frac{1}{2\pi}\sum_{k\in\mathbb{I}} \frac{\imath k h}{1-\mathrm{e}^{-\imath k h}}\mathcal{F}\big(D_x^- W\big)(k)\mathrm{e}^{\imath x k}.$$

Since T_W' is a trigonometric polynomial of degree N, it follows from (2.146) that there is a ξ_0 in $(-h, 0)$ such that

$$\big\|T_W'\big\|_{L_p(\omega)} = \big\|T_W'(\cdot + \xi_0)\big\|_{L_p(\omega^h)}.$$

Letting

$$\lambda(kh) := \frac{\imath k h}{1-\mathrm{e}^{-\imath k h}}$$

and

$$\mu(k) := \lambda(kh)\mathrm{e}^{\imath k \xi_0},$$

the last equality can be rewritten as follows:

$$\big\|T_W'\big\|_{L_p(\omega)} = \big\|\mathcal{F}^{-1}\big(\mu\mathcal{F}\big(D_x^- W\big)\big)\big\|_{L_p(\omega^h)}.$$

A simple calculation shows that both λ and $\mathrm{var}(\lambda)$ are bounded by a constant, independent of h. It remains to apply part (a) of Theorem 2.49 to deduce that λ is a discrete Fourier multiplier on $L_p(\omega^h)$, and therefore the same is true of μ. Hence the upper bound in part (a).

(b) Let us define $Z = D_x^+ W$. Then, $D_x^+ D_x^- W = D_x^- Z$ and by part (a) of this lemma we have that

$$\left\| D_x^+ D_x^- W \right\|_{L_p(\omega^h)} = \left\| D_x^- Z \right\|_{L_p(\omega^h)} \leq \left\| T_Z' \right\|_{L_p(\omega)}.$$

Since

$$\mathcal{F}Z(k) = \mathcal{F}\big(D_x^+ W\big)(k) = \frac{\mathrm{e}^{\imath kh} - 1}{h} \mathcal{F}W(k),$$

it follows that

$$\begin{aligned} T_Z(x) &= \frac{1}{2\pi} \sum_{k \in \mathbb{I}} \mathcal{F}Z(k) \mathrm{e}^{\imath xk} \\ &= \frac{1}{2\pi} \sum_{k \in \mathbb{I}} \frac{\mathrm{e}^{\imath kh} - 1}{h} \mathcal{F}W(k) = D_x^+ T_W(x). \end{aligned}$$

By noting that $T_W(x)$ is a 2π-periodic function of x we deduce that

$$\begin{aligned} \left\| T_Z' \right\|_{L_p(\omega)}^p &= \left\| D_x^+ T_W' \right\|_{L_p(\omega)}^p = h^{-p} \int_{-\pi}^{\pi} \left| T_W'(x+h) - T_W'(x) \right|^p \mathrm{d}x \\ &= h^{-p} \int_{-\pi}^{\pi} \left| \int_x^{x+h} T_W''(\xi)\, \mathrm{d}\xi \right|^p \mathrm{d}x \leq \frac{1}{h} \int_{-\pi}^{\pi} \int_x^{x+h} \left| T_W''(t) \right|^p \mathrm{d}t\, \mathrm{d}x \\ &= \frac{1}{h} \int_{-\pi}^{\pi} \left| T_W''(t) \right|^p \left(\int_{t-h}^{t} \mathrm{d}x \right) \mathrm{d}t = \int_{-\pi}^{\pi} \left| T_W''(t) \right|^p \mathrm{d}t = \left\| T_W'' \right\|_{L_p(\omega)}^p. \end{aligned}$$

Hence we obtain the first inequality in (b). The second inequality is proved in the same way as in part (a), by observing that

$$T_W''(x) = \frac{1}{2\pi} \sum_{k \in \mathbb{I}} \frac{(\imath kh)^2}{(\mathrm{e}^{\imath kh} - 1)(1 - \mathrm{e}^{-\imath kh})} \mathcal{F}\big(D_x^+ D_x^- W\big)(k) \mathrm{e}^{\imath xk}.$$

Thus, by noting that with $\xi_0 \in (-h, 0]$ as in part (a) the function μ_1 defined on $\mathbb{I}$ by

$$\mu_1(k) := \frac{(\imath kh)^2}{(\mathrm{e}^{\imath kh} - 1)(1 - \mathrm{e}^{-\imath kh})} \mathrm{e}^{\imath k \xi_0} = \left(\frac{\frac{kh}{2}}{\sin \frac{kh}{2}} \right)^2 \mathrm{e}^{\imath k \xi_0}$$

is bounded by $\pi/2$ and $\mathrm{var}(\mu_1)$ is bounded by a constant, independent of h, it follows from part (a) of Theorem 2.49 that μ_1 is a discrete Fourier multiplier on $L_p(\omega^h)$, and hence the upper bound stated in part (b). □

Lemma 2.58 has the following extension to two space dimensions.

Lemma 2.59 *Let W be a mesh-function defined on $\omega^h = h\mathbb{I}^2$, where $\mathbb{I} = \{-N+1, \dots, N\}$, and let T_W be the trigonometric interpolant of W on $(-\pi, \pi]^2$ given by* (2.145), *with $n = 2$. There is a constant $C_p > 0$, independent of h and W, such that the following inequalities hold, with $\omega = (-\pi, \pi)^2$:*

(a)

$$\frac{1}{1+\pi}\|D_x^- W\|_{L_p(\omega^h)} \le \left\|\frac{\partial}{\partial x} T_W\right\|_{L_p(\omega)} \le C_p \|D_x^- W\|_{L_p(\omega^h)}$$

and

$$\frac{1}{1+\pi}\|D_y^- W\|_{L_p(\omega^h)} \le \left\|\frac{\partial}{\partial y} T_W\right\|_{L_p(\omega)} \le C_p \|D_y^- W\|_{L_p(\omega^h)};$$

(b)

$$\frac{1}{1+\pi}\|D_x^+ D_x^- W\|_{L_p(\omega^h)} \le \left\|\frac{\partial^2}{\partial x^2} T_W\right\|_{L_p(\omega)} \le C_p \|D_x^+ D_x^- W\|_{L_p(\omega^h)},$$

$$\|D_x^- D_y^- W\|_{L_p(\omega^h)} \le \left\|\frac{\partial^2}{\partial x \partial y} T_W\right\|_{L_p(\omega)} \le C_p \|D_x^- D_y^- W\|_{L_p(\omega^h)}$$

and

$$\frac{1}{1+\pi}\|D_y^+ D_y^- W\|_{L_p(\omega^h)} \le \left\|\frac{\partial^2}{\partial y^2} T_W\right\|_{L_p(\omega)} \le C_p \|D_y^+ D_y^- W\|_{L_p(\omega^h)}.$$

Proof The proof of this result is a straightforward consequence of Lemma 2.58, and Lemmas 2.48 and 2.51 with $n = 1$; Lemma 2.58 is applied in the co-ordinate direction in which differentiation has taken place, and Lemmas 2.48 and 2.51 in the other direction. □

Lemma 2.60 *The norms $\|\cdot\|_{W_p^m(\Omega^h)}$ and $\|\cdot\|_{H_p^m(\Omega^h)}$ are equivalent, uniformly in h, for $m = 0, 1, 2$ and $1 < p < \infty$; i.e. there exist two constants C_1 and C_2, independent of h, such that for all functions V defined on Ω^h (or on $\overline{\Omega}^h$ and equal to zero of Γ^h),*

$$C_1 \|V\|_{W_p^m(\Omega^h)} \le \|V\|_{H_p^m(\Omega^h)} \le C_2 \|V\|_{W_p^m(\Omega^h)}.$$

Proof The statement is obviously true for $m = 0$ with $C_1 = C_2 = 1$. Now for $m = 1, 2$ we shall proceed as follows. Let $\tilde{V}$ denote the odd extension of V to $\omega^h = h\mathbb{I}^2$, where $\mathbb{I} = \{-N+1, \dots, N\}$. Further, let $T_{\tilde{V}}$ denote the trigonometric interpolant of $\tilde{V}$ defined by (2.145) with $n = 2$. By applying Lemma 2.59 with $W = \tilde{V}$, we deduce the existence of two positive constants C_1 and C_2, independent of V and h, (with $C_2 = C_2(p)$ and C_1 independent of p), such that

$$C_1 \|V\|_{W_p^1(\Omega^h)} \le \|T_{\tilde{V}}\|_{W_p^1(\omega)} \le C_2 \|V\|_{W_p^1(\Omega^h)}$$

and

$$C_1\|V\|_{W_p^2(\Omega^h)} \le \|T_{\tilde V}\|_{W_p^2(\omega)} \le C_2\|V\|_{W_p^2(\Omega^h)}.$$

For $p \in (1,\infty)$ and a nonnegative integer m the Sobolev norm $\|\cdot\|_{W_p^m(\omega)}$ on $\omega = \mathbb{T}^2$, is equivalent to the periodic Bessel-potential norm $\|\cdot\|_{H_p^m(\omega)}$ defined by

$$\|v\|_{H_p^m(\omega)} := \big\|\big(\big(1+|k|^2\big)^{m/2}\hat v\big)^\vee\big\|_{L_p(\omega)}$$

(see, Schmeiser and Triebel [162]), where $\hat{\cdot}$ and $\cdot^\vee$ denote the Fourier transform of a periodic distribution and its inverse, defined in Sect. 1.9.5; therefore,

$$C_1\|V\|_{W_p^m(\Omega^h)} \le \big\|\big(\big(1+|k|^2\big)^{m/2}\widehat{T_{\tilde V}}\big)^\vee\big\|_{L_p(\omega)} \le C_2\|V\|_{W_p^m(\Omega^h)}, \quad m=1,2.$$

Finally, since $((1+|k|^2)^{m/2}\widehat{T_{\tilde V}})^\vee = T_{\mathcal{F}^{-1}((1+|k|^2)^{m/2}\mathcal{F}\tilde V)}$ on ω, we have by Lemmas 2.48 and 2.51 that

$$C_1\|V\|_{W_p^m(\Omega^h)} \le \big\|\mathcal{F}^{-1}\big(\big(1+|k|^2\big)^{m/2}\mathcal{F}\tilde V\big)\big\|_{L_p(\omega^h)} \le C_2\|V\|_{W_p^m(\Omega^h)}, \quad m=1,2,$$

from which the result follows by noting that

$$\begin{aligned}\big\|\mathcal{F}^{-1}\big(\big(1+|k|^2\big)^{m/2}\mathcal{F}\tilde V\big)\big\|_{L_p(\omega^h)} &= 4^{1/p}\big\|\mathcal{F}_\sigma^{-1}\big(\big(1+|k|^2\big)^{m/2}\mathcal{F}_\sigma V\big)\big\|_{L_p(\Omega^h)}\\ &= 4^{1/p}\|V\|_{H_p^m(\Omega^h)}, \quad m=1,2. \qquad \square\end{aligned}$$

We shall now use function space interpolation to obtain scales of error bounds in fractional-order discrete Bessel-potential norms. We start with a generalization of an interpolation inequality of Mokin (cf. Theorem 5 in [141]).

Lemma 2.61 *Let α and β be two nonnegative real numbers such that $\alpha < \beta$ and suppose that $1 < p < \infty$. There exists a positive constant C, independent of h, such that for any real number r, $\alpha \le r \le \beta$,*

$$\|V\|_{H_p^r(\Omega^h)} \le C\|V\|^{1-\mu}_{H_p^\alpha(\Omega^h)}\|V\|^{\mu}_{H_p^\beta(\Omega^h)} \quad \forall V \in H_p^\beta\big(\Omega^h\big),$$

where $\mu = (r-\alpha)/(\beta-\alpha)$.

Proof Let us first prove the result for $\alpha = 0$. We define $W := I_{-r,h}V$; then

$$\begin{aligned}\|V\|_{H_p^r(\Omega^h)} &= \|W\|_{L_p(\Omega^h)} = 4^{-1/p}\|\tilde W\|_{L_p(\omega^h)}\\ &\le 4^{-1/p}(1+\pi)^2\|T_{\tilde W}\|_{L_p(\omega)} = 4^{-1/p}(1+\pi)^2\|T_{\tilde V}\|_{H_p^r(\omega)}.\end{aligned}$$

Also,

$$\|T_{\tilde V}\|_{H_p^r(\omega)} \le C\|T_{\tilde V}\|^{1-(r/\beta)}_{L_p(\omega)}\|T_{\tilde V}\|^{r/\beta}_{H_p^\beta(\omega)},$$

(see, Nikol'skiĭ [144], p. 310) where $C = C(p, r, s)$ is a positive constant, and by Lemma 2.51 we have that

$$\|T_{\tilde{V}}\|_{L_p(\omega)} \le 4^{1/p} C \|V\|_{L_p(\Omega^h)} \quad \text{and} \quad \|T_{\tilde{V}}\|_{H_p^\beta(\omega)} \le 4^{1/p} C \|V\|_{H_p^\beta(\Omega^h)}.$$

Combining the last four inequalities, we deduce the statement of the lemma in the case of $\alpha = 0$.

For $\alpha > 0$, let us define $W := I_{-\alpha,h} V$. Then,

$$\|W\|_{H_p^{\beta-\alpha}(\Omega^h)} = \|V\|_{H_p^\beta(\Omega^h)} \quad \text{and} \quad \|W\|_{H_p^{r-\alpha}(\Omega^h)} = \|V\|_{H_p^r(\Omega^h)};$$

moreover, as $0 \le r - \alpha \le \beta - \alpha$, it follows from the case of $\alpha = 0$ above that

$$\|W\|_{H_p^{r-\alpha}(\Omega^h)} \le C \|W\|_{L_p(\Omega^h)}^{1-\mu} \|W\|_{H_p^{\beta-\alpha}(\Omega^h)}^{\mu},$$

and hence the desired inequality. □

Lemma 2.61 will play a key role in the proof of the next theorem, concerned with optimal error bounds in fractional-order discrete Bessel-potential norms.

Theorem 2.62 *Let u be the weak solution of the boundary-value problem (2.149), (2.150), let U be the solution of the finite difference scheme (2.153), (2.154). If u belongs to $W_p^s(\Omega)$, $2/p < s \le 2$ and $0 \le r \le 2$, or $2/p < s \le 3$ and $1 \le r \le 2$, with $1 < p < \infty$ and $r \le s$, then we have that*

$$\|u - U\|_{H_p^r(\Omega^h)} \le C h^{s-r} |u|_{W_p^s(\Omega)},$$

with a positive constant C, dependent on p, r and s, but independent of h.

Proof Let us suppose that u belongs to $W_p^s(\Omega)$, $2/p < s \le 2$, $1 < p < \infty$ and $0 \le r \le 2$. We apply Lemma 2.61 with $\alpha = 0$, $\beta = 2$ and Theorem 2.56 to obtain the error bound.

Similarly, if u belongs to $W_p^s(\Omega)$, $2/p < s \le 3$, $1 < p < \infty$ and $1 \le r \le 2$, then we take $\alpha = 1$ and $\beta = 2$ in Lemma 2.61 in combination with Theorem 2.56 to deduce the error bound. □

By invoking Lemma 2.61 with $\alpha = 0$ and $\beta = 1$, we obtain from Theorem 2.57, using function space interpolation, the following scale of error bounds in fractional-order discrete Bessel-potential norms.

Theorem 2.63 *Let u be the weak solution of the boundary-value problem (2.149), (2.150), let U be the solution of the finite difference scheme (2.151), (2.152). If u belongs to $W_p^s(\Omega)$, $0 \le s < 1 + 1/p$, $s \ne 1/p$, $1 < p < \infty$, $0 \le r \le 1$ and $r \le s$, then*

$$\left\|T_h^{11} u - U\right\|_{H_p^r(\Omega^h)} \le C h^{s-r} |u|_{W_p^s(\Omega)},$$

with a positive constant C, dependent on p, r and s, but independent of h.

The error bounds stated in Theorems 2.56, 2.57, 2.62, and 2.63 cover the range of possible Sobolev indices, $s \in [0, 4]$, for which the solution U of the difference scheme (2.151), (2.152) converges to the weak solution u (or its mollification $T_h^{11}u$) of the boundary-value problem (2.149), (2.150), provided that $u \in W_p^s(\Omega)$. To conclude, we note that to derive these results it is not essential that u is *weak* solution: indeed, if we assume that $u \in W_p^s(\Omega)$ with $s > 1/p$ is a solution of the boundary-value problem in the sense of distributions and that it satisfies a homogeneous Dirichlet boundary condition in the sense of the trace theorem, the error bounds obtained above still hold.

2.6 Approximation of Second-Order Elliptic Equations with Variable Coefficients

Hitherto we have been concerned with the construction and error analysis of finite difference schemes for second-order linear elliptic equations of the form $-\Delta u + c(x, y)u = f(x, y)$. In particular, we derived optimal-order error bounds under minimal smoothness requirements on the solution. Here we shall extend these results to elliptic equations with variable coefficients in the principal part of the differential operator, under minimal regularity hypotheses on the solution and the coefficients.

In Sect. 2.6.1 we consider the Dirichlet problem for a second-order elliptic equation with variable coefficients in the principal part of the operator. The finite difference approximation of this problem is shown to be convergent, with optimal order, in the discrete W_2^1 norm. In Sects. 2.6.2 and 2.6.3 similar results are proved in the discrete W_2^2 norm and in the discrete L_2 norm; then, using function space interpolation, these bounds are extended to fractional-order discrete W_2^r norms, with $r \in [0, 2]$, in Sect. 2.6.4. In Sect. 2.6.5 we focus on elliptic equations with separated variables and derive optimal bounds in the discrete L_2 norm, which are compatible with our hypotheses on the smoothness of the data.

2.6.1 Convergence in the Discrete $\mathbf{W_2^1}$ Norm

As a model problem, we shall consider the following homogeneous Dirichlet boundary-value problem for a second-order linear elliptic equation with variable coefficients on the open unit square $\Omega = (0, 1)^2$:

$$\begin{aligned} \mathcal{L}u := -\sum_{i,j=1}^{2} \partial_i(a_{ij}\partial_j u) + au = f \quad \text{in } \Omega, \\ u = 0 \quad \text{on } \Gamma = \partial\Omega. \end{aligned} \tag{2.166}$$

For the sake of notational simplicity, we have denoted the two independent variables here by x_1 and x_2 instead of x and y.

We shall suppose that (2.166) has a solution in $W_2^s(\Omega)$, which satisfies the partial differential equation in the sense of distributions and the boundary condition in the sense of the trace theorem, with the right-hand side f being an element of $W_2^{s-2}(\Omega)$. In order for the solution of this problem to have a well-defined trace on $\partial\Omega$ it is necessary to assume that $s > 1/2$. It is then natural to require that the coefficients a_{ij} and a belong to appropriate spaces of multipliers; that is,

$$a_{ij} \in M\big(W_2^{s-1}(\Omega)\big), \quad a \in M\big(W_2^s(\Omega) \to W_2^{s-2}(\Omega)\big).$$

According to the results in Sect. 1.8 the following conditions are sufficient in order to ensure that this is the case:

(a) if $|s-1| > 1$, then

$$a_{ij} \in W_2^{|s-1|}(\Omega), \quad a \in W_2^{|s-1|-1}(\Omega);$$

(b) if $0 \leq |s-1| \leq 1$, then

$$a_{ij} \in W_p^{|s-1|+\delta}(\Omega), \quad a = a_0 + \sum_{i=1}^{2} \partial_i a_i,$$

$$a_0 \in L_{2+\varepsilon}(\Omega), \quad a_i \in W_p^{|s-1|+\delta}(\Omega),$$

where $\varepsilon > 0$; and $\delta > 0$, $p \geq 2/|s-1|$ for $0 < |s-1| < 1$; $\delta = 0$, $p > 2$ for $s = 0$; $\delta = 0$, $p = \infty$ when $s = 1$.

In addition to these assumptions on the smoothness of the data, we shall adopt the following structural hypotheses on the coefficients a_{ij} and a:

- there exists a $c_0 > 0$ such that

$$\sum_{i,j=1}^{2} a_{ij}(x)\xi_i\xi_j \geq c_0 \sum_{i=1}^{2} \xi_i^2 \quad \forall x \in \overline{\Omega},\ \forall \xi = (\xi_1, \xi_2) \in \mathbb{R}^2;$$

- the matrix $(a_{ij}) \in \mathbb{R}^{2\times 2}$ is symmetric, i.e.

$$a_{ij} = a_{ji}, \quad i, j = 1, 2;$$

- the coefficient a is nonnegative in the sense of distributions; i.e.

$$\langle a\varphi, \varphi\rangle_{\mathcal{D}'\times\mathcal{D}} \geq 0 \quad \forall \varphi \in \mathcal{D}(\Omega).$$

We shall construct a finite difference approximation of this boundary-value problem on the uniform mesh $\overline{\Omega}^h := \Omega^h \cup \Gamma^h$ of mesh-size $h := 1/N$, with $N \geq 2$, defined in Sect. 2.2.4.

When $s \leq 3$, our hypotheses on the smoothness of the data do not guarantee that the forcing function f and the coefficient a are continuous on Ω: it is therefore necessary to mollify them so as to ensure that they have well-defined values at the mesh-points.

These observations lead us to consider the following finite difference approximation of the boundary-value problem:

$$\begin{aligned} \mathcal{L}_h U &= T_h^{22} f \quad \text{on } \Omega^h, \\ U &= 0 \quad \text{on } \Gamma^h, \end{aligned} \tag{2.167}$$

with

$$\mathcal{L}_h U := -\frac{1}{2} \sum_{i,j=1}^{2} \left[D_{x_i}^{+}\left(a_{ij} D_{x_j}^{-} U\right) + D_{x_i}^{-}\left(a_{ij} D_{x_j}^{+} U\right)\right] + \left(T_h^{22} a\right) U,$$

where $D_{x_i}^{\pm} V$, $i = 1, 2$, are the divided difference operators in the x_i co-ordinate direction defined in Sect. 2.2.4, and T_h^{22} is the mollifier with mesh-size h defined in Sect. 1.9.2.

It is helpful to note that the two-dimensional mollifier T_h^{22} can be expressed in terms of the one-dimensional mollifiers $T_1 = T_{1,h}$ and $T_2 = T_{2,h}$, acting in the x_1 and x_2 co-ordinate direction, respectively, as

$$T_h^{22} = T_1^2 T_2^2.$$

For a locally integrable function w defined on Ω,

$$T_1 w(x_1, x_2) := \frac{1}{h} \int_{x_1 - h/2}^{x_1 + h/2} w(\xi_1, x_2)\, \mathrm{d}\xi_1,$$

$$T_1^2 w(x_1, x_2) := \frac{1}{h} \int_{x_1 - h}^{x_1 + h} \left(1 - \left|\frac{x_1 - \xi_1}{h}\right|\right) w(\xi_1, x_2)\, \mathrm{d}\xi_1;$$

$T_2 w$ and $T_2^2 w$ can be represented analogously. When w is a distribution, T_i and T_i^2 are defined as convolutions of w with the scaled univariate B-splines θ_h^1 and θ_h^2, respectively, as explained in Sect. 1.9.

We note that (2.167) is the standard symmetric seven-point difference scheme with mollified right-hand side and mollified coefficient a.

With the notations from Sect. 2.2.4, we consider the discrete L_2 inner product $(V, W)_h$ (see (2.48)) in the linear space S_0^h of real-valued mesh-functions defined on $\overline{\Omega}^h$ that vanish on Γ^h, the associated discrete L_2 norm $\|V\|_{L_2(\Omega^h)}$, and the discrete Sobolev norms $\|V\|_{W_2^1(\Omega^h)}$ and $\|V\|_{W_2^2(\Omega^h)}$.

The error bounds stated in the next theorem are compatible with the smoothness hypotheses (a) and (b) formulated above, for the coefficients appearing in the partial differential equation.

Theorem 2.64 *The difference scheme* (2.167) *satisfies the following error bounds in the* $W_2^1(\Omega^h)$ *norm*:

$$\|u-U\|_{W_2^1(\Omega^h)} \le Ch^{s-1}\Big(\max_{i,j}\|a_{ij}\|_{W_2^{s-1}(\Omega)} + \|a\|_{W_2^{s-2}(\Omega)}\Big)\|u\|_{W_2^s(\Omega)}, \quad \text{for } 2 < s \le 3, \tag{2.168}$$

and

$$\|u-U\|_{W_2^1(\Omega^h)} \le Ch^{s-1}\Big(\max_{i,j}\|a_{ij}\|_{W_p^{s-1+\delta}(\Omega)} + \max_{i}\|a_i\|_{W_p^{s-1+\delta}(\Omega)} + \|a_0\|_{L_{2+\varepsilon}(\Omega)}\Big)\|u\|_{W_2^s(\Omega)}, \quad \text{for } 1 < s \le 2, \tag{2.169}$$

where p, δ *and* ε *are as in condition* (*b*) *above*, *and* C *is a positive constant*, *independent of* h.

Before embarking on the proofs of these error bounds we shall make some preliminary observations. Let u denote the solution of the boundary-value problem (2.166) and let U be the solution of the finite difference scheme (2.167). When $s > 1$, as in Theorem 2.64, the function u is continuous on $\overline{\Omega}$ and therefore the *global error* $e := u - U$ is correctly defined on the uniform mesh $\overline{\Omega}^h$. In addition, it is easily seen that

$$\begin{aligned} \mathcal{L}_h e &= \sum_{i,j=1}^{2} D_{x_i}^- \eta_{ij} + \eta \quad \text{on } \Omega^h, \\ e &= 0 \quad \text{on } \Gamma^h, \end{aligned} \tag{2.170}$$

where

$$\eta_{ij} := T_i^+ T_{3-i}^2(a_{ij}\partial_j u) - \frac{1}{2}\big(a_{ij} D_{x_j}^+ u + a_{ij}^{+i} D_{x_j}^- u^{+i}\big), \quad i = 1, 2,$$

and

$$\eta := \big(T_1^2 T_2^2 a\big)u - T_1^2 T_2^2 (au).$$

Here, for a locally integrable function w defined on Ω, we have used the asymmetric mollifiers $T_i^{\pm} w$, defined at $x = (x_1, x_2)$ by

$$\big(T_i^{\pm} w\big)(x) := (T_i w)\left(x \pm \frac{1}{2} h \mathrm{e}_i\right), \quad \text{with } \mathrm{e}_i := (\delta_{i1}, \delta_{i2}),\ i = 1, 2.$$

By taking the $(\cdot,\cdot)_h$ inner product of $L_h e$ with e and performing summations by parts in the leading terms on the left- and right-hand sides, in exactly the same manner as in the argument that led to the estimate (2.83) stated in Lemma 2.24, we arrive at the following result.

Lemma 2.65 *The difference scheme* (2.170) *is stable*, *in the sense that*

$$\|e\|_{W_2^1(\Omega^h)} \leq C\left(\sum_{i,j=1}^{2} \|\eta_{ij}\|_{L_2(\Omega_i^h)} + \|\eta\|_{L_2(\Omega^h)}\right), \tag{2.171}$$

where C is a positive constant, *independent of h*.

The error analysis of the finite difference scheme (2.167) is thereby reduced to estimating the right-hand side in the inequality (2.171). To this end, we decompose η_{ij} as follows:

$$\eta_{ij} = \eta_{ij1} + \eta_{ij2} + \eta_{ij3} + \eta_{ij4},$$

where

$$\eta_{ij1} := T_i^+ T_{3-i}^2 (a_{ij} \partial_j u) - \big(T_i^+ T_{3-i}^2 a_{ij}\big)\big(T_i^+ T_{3-i}^2 \partial_j u\big),$$

$$\eta_{ij2} := \left[T_i^+ T_{3-i}^2 a_{ij} - \frac{1}{2}\big(a_{ij} + a_{ij}^{+i}\big)\right]\big(T_i^+ T_{3-i}^2 \partial_j u\big),$$

$$\eta_{ij3} := \frac{1}{2}\big(a_{ij} + a_{ij}^{+i}\big)\left[T_i^+ T_{3-i}^2 \partial_j u - \frac{1}{2}\big(D_{x_j}^+ u + D_{x_j}^- u^{+i}\big)\right]$$

and

$$\eta_{ij4} := -\frac{1}{4}\big(a_{ij} - a_{ij}^{+i}\big)\big(D_{x_j}^+ u - D_{x_j}^- u^{+i}\big).$$

We shall also perform a decomposition of η, but the form of this decomposition will depend on whether $1 < s \leq 2$ or $2 < s \leq 3$.

When $1 < s \leq 2$, we shall write

$$\eta = \eta_0 + \eta_1 + \eta_2,$$

where

$$\eta_0 := \big(T_1^2 T_2^2 a_0\big)u - T_1^2 T_2^2 (a_0 u)$$

and

$$\eta_i := \big(T_1^2 T_2^2 \partial_i a_i\big)u - T_1^2 T_2^2 (u \partial_i a_i), \quad i = 1, 2.$$

Whereas if $2 < s \leq 3$, we shall use the decomposition

$$\eta = \eta_3 + \eta_4,$$

where

$$\eta_3 := \big(T_1^2 T_2^2 a\big)\big(u - T_1^2 T_2^2 u\big)$$

and

$$\eta_4 := \left(T_1^2 T_2^2 a\right)\left(T_1^2 T_2^2 u\right) - T_1^2 T_2^2 (au).$$

Proof of Theorem 2.64 We introduce the 'elementary rectangles'

$$K^0 = K^0(x) := \left\{y = (y_1, y_2) : |y_j - x_j| < h,\, j = 1, 2\right\}$$

and

$$K^i = K^i(x) := \left\{y : x_i < y_i < x_i + h,\, |y_{3-i} - x_{3-i}| < h\right\}, \quad i = 1, 2.$$

The linear transformation $y = x + h\tilde{x}$ defines a bijective mapping of the 'canonical rectangles'

$$\tilde{K}^0 := \left\{\tilde{x} = (\tilde{x}_1, \tilde{x}_2) : |\tilde{x}_j| < 1,\, j = 1, 2\right\}$$

and

$$\tilde{K}^i := \{\tilde{x} : 0 < \tilde{x}_i < 1,\, |\tilde{x}_{3-i}| < 1\}, \quad i = 1, 2,$$

onto K^0 and K^i, respectively. Further, we define

$$\tilde{a}_{ij}(\tilde{x}) := a_{ij}(x + h\tilde{x}), \qquad \tilde{u}(\tilde{x}) := u(x + h\tilde{x}),$$

and so on. The value of η_{ij1} at a mesh-point $x \in \Omega_i^h$ can be expressed as

$$\begin{aligned}\eta_{ij1}(x) = \frac{1}{h}\Bigg[&\int_{\tilde{K}^i} \left(1 - |\tilde{x}_{3-i}|\right)\tilde{a}_{ij}(\tilde{x}) \frac{\partial \tilde{u}}{\partial \tilde{x}_j}\,\mathrm{d}\tilde{x} \\ &- \int_{\tilde{K}^i} \left(1 - |\tilde{x}_{3-i}|\right)\tilde{a}_{ij}(\tilde{x})\,\mathrm{d}\tilde{x} \times \int_{\tilde{K}^i} \left(1 - |\tilde{x}_{3-i}|\right)\frac{\partial \tilde{u}}{\partial \tilde{x}_j}\,\mathrm{d}\tilde{x}\Bigg].\end{aligned}$$

Hence we deduce that $\eta_{ij1}(x)$ is a bounded bilinear functional of the argument

$$(\tilde{a}_{ij}, \tilde{u}) \in W_q^\lambda\left(\tilde{K}^i\right) \times W_{2q/(q-2)}^\mu\left(\tilde{K}^i\right),$$

where $\lambda \geq 0$, $\mu \geq 1$ and $q > 2$. Furthermore, $\eta_{ij1} = 0$ whenever $\tilde{a}_{ij}$ is a constant function or $\tilde{u}$ is a polynomial of degree 1. By applying the bilinear version of the Bramble–Hilbert lemma (cf. Lemma 2.30 with $m = 2$), we deduce that

$$\left|\eta_{ij1}(x)\right| \leq \frac{C}{h} |\tilde{a}_{ij}|_{W_q^\lambda(\tilde{K}^i)} |\tilde{u}|_{W_{2q/(q-2)}^\mu(\tilde{K}^i)}, \quad 0 \leq \lambda \leq 1,\ 1 \leq \mu \leq 2.$$

Returning from the canonical variables $(\tilde{x}_1, \tilde{x}_2)$ to the original variables (x_1, x_2) we obtain

$$|\tilde{a}_{ij}|_{W_q^\lambda(\tilde{K}^i)} = h^{\lambda - 2/q} |a_{ij}|_{W_q^\lambda(K^i)}$$

and

$$|\tilde{u}|_{W^{\mu}_{2q/(q-2)}(\tilde{K}^i)} = h^{\mu-(q-2)/q}|u|_{W^{\mu}_{2q/(q-2)}(K^i)}.$$

Therefore,

$$\left|\eta_{ij1}(x)\right| \leq Ch^{\lambda+\mu-2}|a_{ij}|_{W^{\lambda}_q(K^i)}|u|_{W^{\mu}_{2q/(q-2)}(K^i)}, \quad 0 \leq \lambda \leq 1, 1 \leq \mu \leq 2.$$

By summing through the mesh-points in Ω_i^h and applying Hölder's inequality we then deduce, for $0 \leq \lambda \leq 1$ and $1 \leq \mu \leq 2$, the bound

$$\|\eta_{ij1}\|_{L_2(\Omega_i^h)} \leq Ch^{\lambda+\mu-1}|a_{ij}|_{W^{\lambda}_q(\Omega)}|u|_{W^{\mu}_{2q/(q-2)}(\Omega)}. \tag{2.172}$$

Let us choose $\lambda = s - 1$, $\mu = 1$ and $q = p$. Thanks to the Sobolev embedding theorem (cf. Theorem 1.34),

$$W^{s-1+\delta}_p(\Omega) \hookrightarrow W^{s-1}_p(\Omega) \quad \text{and} \quad W^s_2(\Omega) \hookrightarrow W^1_{2p/(p-2)}(\Omega), \quad 1 < s \leq 2.$$

Thus, (2.172) yields

$$\|\eta_{ij1}\|_{L_2(\Omega_i^h)} \leq Ch^{s-1}\|a_{ij}\|_{W^{s-1+\delta}_p(\Omega)}\|u\|_{W^s_2(\Omega)}, \quad 1 < s \leq 2. \tag{2.173}$$

Analogous bounds hold for η_{ij2}, η_{ij4}, η_1 and η_2. Now suppose that $q > 2$; then, the following Sobolev embeddings hold:

$$W^{\lambda+\mu-1}_2(\Omega) \hookrightarrow W^{\lambda}_q(\Omega) \quad \text{for } \mu > 2 - 2/q$$

and

$$W^{\lambda+\mu}_2(\Omega) \hookrightarrow W^{\mu}_{2q/(q-2)}(\Omega) \quad \text{for } \lambda > 2/q.$$

Setting $\lambda + \mu = s$ in (2.172) yields

$$\|\eta_{ij1}\|_{L_2(\Omega_i^h)} \leq Ch^{s-1}\|a_{ij}\|_{W^{s-1}_2(\Omega)}\|u\|_{W^s_2(\Omega)}, \quad 2 < s \leq 3. \tag{2.174}$$

The functional η_{ij4} is bounded in a similar fashion.

For $s > 2$, η_{ij2} is a bilinear functional of the argument

$$(a_{ij}, u) \in W^{s-1}_2\left(K^i\right) \times W^1_{\infty}\left(K^i\right)$$

and $\eta_{ij2} = 0$ whenever a_{ij} is a polynomial of degree 1 or if u is a constant function. By applying Lemma 2.65 and the embedding $W^s_2(\Omega) \hookrightarrow W^1_{\infty}(\Omega)$ we obtain a bound on η_{ij2}, which is of the form (2.174).

By a similar argument, $\eta_{ij3}(x)$ is a bounded bilinear functional of the argument

$$(a_{ij}, u) \in C\left(\overline{K}^i\right) \times W^s_2\left(K^i\right)$$

for $s > 1$ and it vanishes whenever u is a polynomial of degree 2. By noting the embeddings

$$W_p^{s-1+\delta}(\Omega) \hookrightarrow C(\overline{\Omega}) \quad \text{for } 1 < s \leq 2$$

and

$$W_2^{s-1}(\Omega) \hookrightarrow C(\overline{\Omega}) \quad \text{for } s > 2,$$

we obtain bounds of the form (2.173) and (2.174) for η_{ij3}.

Let $2 < q < 2/(3-s)$. When $2 < s \leq 3$, $\eta_3(x)$ is a bounded bilinear functional of the argument

$$(a, u) \in L_q\big(K^0\big) \times W_{2q/(q-2)}^{s-1}\big(K^0\big).$$

Moreover, $\eta_3 = 0$ when u is a polynomial of degree 1. By noting the Bramble–Hilbert lemma and the Sobolev embeddings

$$W_2^{s-2}(\Omega) \hookrightarrow L_q(\Omega) \quad \text{and} \quad W_2^s(\Omega) \hookrightarrow W_{2q/(q-2)}^{s-1}(\Omega)$$

we obtain

$$\|\eta_3\|_{L_2(\Omega^h)} \leq Ch^{s-1}\|a\|_{W_2^{s-2}(\Omega)}\|u\|_{W_2^s(\Omega)}, \quad 2 < s \leq 3. \tag{2.175}$$

When $2 < s \leq 3$, η_4 is a bounded bilinear functional of

$$(a, u) \in W_2^{s-2}\big(K^0\big) \times W_\infty^1\big(K^0\big)$$

and $\eta_4 = 0$ whenever a or u is a constant function. Using the same technique as before, together with the embedding

$$W_2^s(\Omega) \hookrightarrow W_\infty^1(\Omega),$$

we obtain a bound of the form (2.175) for η_4.

Finally, let $2 < q < \min\{2+\varepsilon, 2/(2-s)\}$. Then, for $1 < s \leq 2$, $\eta_0(x)$ is a bounded bilinear functional of the argument

$$(a_0, u) \in L_q\big(K^0\big) \times W_{2q/(q-2)}^{s-1}\big(K^0\big)$$

and it vanishes when u is a constant function. By noting the embeddings

$$L_{2+\varepsilon}(\Omega) \hookrightarrow L_q(\Omega) \quad \text{and} \quad W_2^s(\Omega) \hookrightarrow W_{2q/(q-2)}^{s-1}(\Omega)$$

we obtain

$$\|\eta_0\|_{L_2(\Omega^h)} \leq Ch^{s-1}\|a_0\|_{L_{2+\varepsilon}(\Omega)}\|u\|_{W_2^s(\Omega)}, \quad 1 < s \leq 2. \tag{2.176}$$

Finally, by combining (2.171) with (2.172)–(2.176) we deduce the desired bounds on the global error. $\square$

2.6.2 Convergence in the Discrete W_2^2 Norm

In this section we consider the error analysis of the scheme (2.167) in the discrete W_2^2 norm (2.50).

From the error bound (2.168) in the $W_2^1(\Omega^h)$ norm derived in the previous section for the difference scheme (2.167) and the *inverse inequality*

$$|V|_{W_2^2(\Omega^h)} \leq \frac{\sqrt{6}}{h}|V|_{W_2^1(\Omega^h)} \quad \forall V \in S_0^h, \tag{2.177}$$

we immediately deduce, with $V = e$, the following error bound in the $W_2^2(\Omega^h)$ norm

$$\|u - U\|_{W_2^2(\Omega^h)} \leq Ch^{s-2}\Big(\max_{i,j}\|a_{ij}\|_{W_2^{s-1}(\Omega)} + \|a\|_{W_2^{s-2}(\Omega)}\Big)\|u\|_{W_2^s(\Omega)}$$
$$\text{for } 2 < s \leq 3. \tag{2.178}$$

In order to derive an analogous error bound when $3 < s \leq 4$ it is necessary to establish the discrete counterpart of the elliptic regularity result

$$\|v\|_{W_2^2(\Omega)} \leq C\|\mathcal{L}v\|_{L_2(\Omega)} \quad \forall v \in W_2^2(\Omega) \cap \mathring{W}_2^1(\Omega),$$

called the *second fundamental inequality*, following the terminology of Ladyzhenskaya and Ural'tseva [118]. A result of this kind was proved for the finite difference operator $\mathcal{L}_h$ by D'yakonov [39]; it states that

$$|V|_{W_2^2(\Omega^h)} \leq C\|\mathcal{L}_h V\|_{L_2(\Omega^h)} \quad \forall V \in S_0^h, \tag{2.179}$$

where

$$C := C(a_{11}, a_{12}, a_{22}, a) = C_0\big(1 + \big\|T_h^{22}a\big\|_{L_q(\Omega^h)}\big)\Big(1 + \max_{i,j}\|a_{ij}\|_{W_q^1(\Omega^h)}^{q/(q-2)}\Big),$$

with $2 < q \leq \infty$; here $\|\cdot\|_{L_q(\Omega^h)}$ and $\|\cdot\|_{W_q^1(\Omega^h)}$ are mesh-dependent norms defined, for $q < \infty$, by

$$\|V\|_{L_q(\Omega^h)} := \Big(h^2 \sum_{x\in\Omega^h}\big|V(x)\big|^q\Big)^{1/q},$$

$$\|V\|_{W_q^1(\Omega^h)} := \Bigg(\|V\|_{L_q(\Omega^h)}^q + \sum_{i=1}^{2}\big\|D_{x_i}^+ V\big\|_{L_q(\Omega_i^h)}^q\Bigg)^{1/q},$$

where $\|\cdot\|_{L_q(\Omega_i^h)}$ is defined in the same way as $\|\cdot\|_{L_q(\Omega^h)}$, except that the sum is taken over mesh-points in Ω_i^h instead of Ω^h. When $q = \infty$,

$$\|V\|_{L_\infty(\Omega^h)} = \|V\|_{\infty,h} := \max_{x\in\Omega^h}\big|V(x)\big|,$$

with an analogous definition of $\|V\|_{W^1_\infty(\Omega^h)}$.

By applying the Bramble–Hilbert lemma it is easily shown that

$$\|a_{ij}\|_{W^1_q(\Omega^h)} \le C_1 \|a_{ij}\|_{W^1_q(\Omega)} \quad \text{and} \quad \left\|T_1^2 T_2^2 a\right\|_{L_q(\Omega^h)} \le C_2 \|a\|_{L_q(\Omega)},$$

where C_1 and C_2 are independent of h. Thus we can assume in (2.179) that

$$\begin{aligned} C &= C(a_{11}, a_{12}, a_{22}, a) \\ &:= C_3(1 + \|a\|_{L_q(\Omega)})\left(1 + \max_{i,j} \|a_{ij}\|^{q/(q-2)}_{W^1_q(\Omega)}\right), \quad 2 < q \le \infty. \end{aligned}$$

Following the terminology of Ladyzhenskaya and Ural'tseva again, we note that the discrete version of the *first fundamental inequality* is

$$c_0 |V|^2_{W^1_2(\Omega^h)} \le (\mathcal{L}_h V, V)_h \quad \forall V \in S^h_0. \tag{2.180}$$

For the difference operator $\mathcal{L}_h$ appearing in (2.167) the first fundamental inequality is easily shown using summation by parts, in the same way as in the case of the result stated in Lemma 2.65.

Now we are ready to consider the error analysis of the difference scheme (2.167) in the norm $W^2_2(\Omega^h)$ for $u \in W^s_2(\Omega)$ when $3 < s \le 4$.

It follows from (2.170) and (2.177) that

$$\|e\|_{W^2_2(\Omega^h)} \le C\left(\sum_{i,j=1}^{2} \left\|D^-_{x_i}\eta_{ij}\right\|_{L_2(\Omega^h_i)} + \|\eta\|_{L_2(\Omega^h)}\right), \tag{2.181}$$

where C is a positive constant, independent of h. By bounding $D^-_{x_i}\eta_{ij}$ and η analogously as in the previous section, we obtain the error bound (2.178) for $3 < s \le 4$. Thus we deduce that (2.178) holds for $2 < s \le 4$ (see also Berikelashvili [10]).

2.6.3 Convergence in the Discrete L_2 Norm

The derivation of an optimal error bound in the $L_2(\Omega^h)$ norm is based on a technique that is usually referred to as a *duality argument*: it uses the adjoint of the difference operator $\mathcal{L}_h$ and the second fundamental inequality for the adjoint of the difference operator $\mathcal{L}_h$. Since in our case the difference operator $\mathcal{L}_h$ is symmetric and, more specifically, selfadjoint on the finite-dimensional space S^h_0 of real-valued mesh-functions defined on $\overline{\Omega}^h$ that vanish on Γ^h, equipped with the inner product of $L_2(\Omega^h)$, the second fundamental inequality for the adjoint of $\mathcal{L}_h$ is, in fact, identical to the second fundamental inequality for $\mathcal{L}_h$, stated in (2.179).

For the sake of simplicity, we shall restrict ourselves to the case when $a(x) \equiv 0$; the boundary-value problem (2.166) then becomes

$$-\sum_{i,j=1}^{2} \partial_i (a_{ij} \partial_j u) = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \Gamma = \partial\Omega, \tag{2.182}$$

and the corresponding finite difference scheme is

$$\begin{aligned} \mathcal{L}_h U := -\frac{1}{2} \sum_{i,j=1}^{2} \left[D_{x_i}^+ \left(a_{ij} D_{x_j}^- U \right) + D_{x_i}^- \left(a_{ij} D_{x_j}^+ U \right) \right] = T_h^{22} f \quad \text{in } \Omega^h, \\ U = 0 \quad \text{on } \Gamma^h. \end{aligned} \tag{2.183}$$

The error analysis of this scheme in the $L_2(\Omega^h)$ norm is based on the observation that the global error $e := u - U$ is the solution of the difference scheme

$$\mathcal{L}_h e = \sum_{i,j=1}^{2} D_{x_i}^- \eta_{ij} \quad \text{in } \Omega^h, \qquad e = 0 \quad \text{on } \Gamma^h, \tag{2.184}$$

where the η_{ij} are the same as in (2.170). The right-hand side can be rewritten as follows:

$$\sum_{i,j=1}^{2} D_{x_i}^- \eta_{ij} = \sum_{i=1}^{2} \left(\mathcal{L}_{ii} \xi_{ii} + \mathcal{K}_i \chi_i + \sum_{j=1}^{2} D_{x_i}^- \upsilon_{ij} \right), \tag{2.185}$$

where

$$\mathcal{L}_{ii} V := -D_{x_i}^- \left[\left(T_i^+ T_{3-i}^2 a_{ii} \right) D_{x_i}^+ V \right], \qquad \mathcal{K}_i V := D_{x_i}^- \left[\left(T_i^+ T_{3-i}^2 a_{i,3-i} \right) D_{x_{3-i}}^+ V \right],$$

and

$$\begin{aligned} \xi_{ij} &:= u - \frac{1}{2} \left(T_{3-i}^- T_{3-j}^+ u + T_{3-i}^+ T_{3-j}^- u \right), \\ \chi_i &:= \varrho_i - \frac{1}{2} \left(\xi_{i,3-i} + \xi_{i,3-i}^{+i,-(3-i)} \right), \\ \varrho_i &:= \frac{1}{4} \left[\left(T_{3-i}^- T_i^+ u - T_{3-i}^+ T_i^- u \right) - \left(T_{3-i}^- T_i^+ u - T_{3-i}^+ T_i^- u \right)^{+i,-(3-i)} \right], \\ \upsilon_{ij} &:= T_i^+ T_{3-i}^2 (a_{ij} \partial_j u) - \left(T_i^+ T_{3-i}^2 a_{ij} \right) \left(T_i^+ T_{3-i}^2 \partial_j u \right) \\ &\quad + \frac{1}{2} \left[\left(T_i^+ T_{3-i}^2 a_{ij} \right) \left(D_{x_j}^+ u + D_{x_j}^- u^{+i} \right) - a_{ij} D_{x_j}^+ u - a_{ij}^{+i} D_{x_j}^- u^{+i} \right]. \end{aligned}$$

Here we have assumed that the solution $u \in W_2^s(\Omega) \cap \mathring{W}_2^1(\Omega)$, $0 \le s \le 2$, has been extended, preserving its Sobolev class, to the square $(-h_0, 1 + h_0)^2$ where h_0 is a fixed positive constant such that $h < h_0$.

Lemma 2.66 *Suppose that $a_{ij} \in W^1_q(\Omega)$, $q > 2$. The solution of the finite difference scheme* (2.184) *then satisfies the bound*

$$\|e\|_{L_2(\Omega^h)} \le C \sum_{i=1}^{2} \Bigg(\|\xi_{ii}\|_{L_2(\Omega^h)} + \|\xi_{i,3-i}\|_{L_2(\overline{\Omega}^h)} + \|\varrho_i\|_{L_2(\Omega^h_{i-1,2-i})} + \sum_{j=1}^{2} \|\upsilon_{ij}\|_{L_2(\Omega^h_i)} \Bigg), \tag{2.186}$$

where C is a positive constant, independent of h.

Proof The proof is based on a duality argument. Let us consider the auxiliary function W, defined as the solution of the finite difference scheme

$$\mathcal{L}_h W = e \quad \text{in } \Omega^h, \qquad W = 0 \quad \text{on } \Gamma^h.$$

We note in passing that in general one would have written $(\mathcal{L}_h)^*$, the adjoint of $\mathcal{L}_h$, on the left-hand side instead of $\mathcal{L}_h$; however, in our case $\mathcal{L}_h$ is selfadjoint. Thus, crucially, $(e, \mathcal{L}_h W)_h = (\mathcal{L}_h e, W)_h$. It then follows from (2.184) and (2.185) that

$$\begin{aligned}
\|e\|^2_{L_2(\Omega^h)} &= (e, \mathcal{L}_h W)_h = (\mathcal{L}_h e, W)_h \\
&= \sum_{i=1}^{2} \Bigg[(\mathcal{L}_{ii}\xi_{ii}, W)_h + (\mathcal{K}_i \chi_i, W)_h + \sum_{j=1}^{2} \big(D^-_{x_i} \upsilon_{ij}, W\big)_h \Bigg] \\
&= \sum_{i=1}^{2} \Bigg[(\xi_{ii}, \mathcal{L}_{ii} W)_h + \big(\chi_i, \mathcal{K}^*_i W\big)_{i-1,2-i,h} - \sum_{j=1}^{2} \big(\upsilon_{ij}, D^+_{x_i} W\big)_{i,h} \Bigg] \\
&\le \sum_{i=1}^{2} \Bigg(\|\xi_{ii}\|_{L_2(\Omega^h)} \|\mathcal{L}_{ii} W\|_{L_2(\Omega^h)} + \|\chi_i\|_{L_2(\Omega^h_{i-1,2-i})} \big\|\mathcal{K}^*_i W\big\|_{L_2(\Omega^h_{i-1,2-i})} \\
&\qquad + \sum_{j=1}^{2} \|\upsilon_{ij}\|_{L_2(\Omega^h_i)} \big\|D^+_{x_i} W\big\|_{L_2(\Omega^h_i)} \Bigg),
\end{aligned}$$

where

$$\mathcal{K}^*_i W = D^-_{x_{3-i}} \big[\big(T^+_i T^2_{3-i} a_{i,3-i} \big) D^+_{x_i} W \big].$$

The second fundamental inequality (2.179) implies that

$$\|\mathcal{L}_{ii} W\|_{L_2(\Omega^h)}, \qquad \big\|\mathcal{K}^*_i W\big\|_{L_2(\Omega^h_{i-1,2-i})}, \qquad \big\|D^+_{x_i} W\big\|_{L_2(\Omega^h_i)}$$

are all bounded by

$$C \|\mathcal{L}_h W\|_{L_2(\Omega^h)} = C \|e\|_{L_2(\Omega^h)},$$

and hence, after substitution of the defining expression for χ_i, we deduce the inequality (2.186). □

We observe that for the second fundamental inequality to hold it is necessary that $a_{ij} \in W_q^1(\Omega)$, $q > 2$; thus we can only expect a sharp error bound when $s = 2$. Let us assume that this is indeed the case, and we proceed to estimate the terms that appear on the right-hand side of the inequality (2.186).

We begin by noting that ξ_{ij} and ϱ_i are bounded linear functionals on $W_2^2(\Omega)$ that vanish on all polynomials of degree 1. By the Bramble–Hilbert lemma,

$$\|\xi_{ii}\|_{L_2(\Omega^h)}, \quad \|\xi_{i,3-i}\|_{L_2(\overline{\Omega}^h)}, \quad \|\varrho_i\|_{L_2(\Omega^h_{i-1,2-i})} \le Ch^2\|u\|_{W_2^2(\Omega)}. \tag{2.187}$$

Arguing in the same way as in the previous section, υ_{ij} is decomposed into three terms that are bounded by means of the Bramble–Hilbert lemma to obtain:

$$\|\upsilon_{ij}\|_{L_2(\Omega_i^h)} \le Ch^2\big(\|a_{ij}\|_{W^1_\infty(\Omega)}\|u\|_{W_2^2(\Omega)} + \|a_{ij}\|_{W^2_\infty(\Omega)}\|u\|_{W_2^1(\Omega)}\big). \tag{2.188}$$

From (2.186)–(2.188) we deduce the following error bound for the difference scheme (2.183):

$$\|u - U\|_{L_2(\Omega^h)} \le Ch^2 \max_{i,j} \|a_{ij}\|_{W^2_\infty(\Omega)}\|u\|_{W_2^2(\Omega)}. \tag{2.189}$$

While the power of h in the error bound (2.189) is optimal in the sense that it is compatible with the smoothness of u, the bound is not entirely satisfactory as the coefficients a_{ij} are required to belong to $W^2_\infty(\Omega)$, which, in the light of the hypotheses (a) and (b) from the beginning of Sect. 2.6.1, can be seen as an excessively strong assumption on the regularity of the coefficients a_{ij}. The requirement for the additional smoothness of the coefficients a_{ij} can be attributed to our crude bound on $D^-_{x_i}\upsilon_{ij}$ in (2.186).

An improved estimate can be obtained by considering an alternative scheme where the coefficients a_{ij} have been mollified:

$$\begin{aligned} \hat{\mathcal{L}}_h U := \sum_{i,j=1}^{2} \mathcal{L}_{ij} U = T_h^{22} f \quad &\text{in } \Omega^h, \\ U = 0 \quad &\text{on } \Gamma^h, \end{aligned} \tag{2.190}$$

where

$$\mathcal{L}_{ij} U := -\frac{1}{2} D^-_{x_i}\big[\big(T_i^+ T_{3-i}^2 a_{ij}\big) D^+_{x_j}\big(U + U^{+i,-j}\big)\big].$$

For this scheme the global error $e := u - U$ satisfies

$$\hat{\mathcal{L}}_h e = \sum_{i=1}^{2}\left(\mathcal{L}_{ii}\xi_{ii} + \mathcal{K}_i\chi_i + \sum_{j=1}^{2} D^-_{x_i}\eta_{ij1}\right) \quad \text{in } \Omega^h, \qquad z = 0 \quad \text{on } \Gamma^h,$$

where ξ_{ii}, χ_i and η_{ij1} are as before. Assuming that $a_{ij} \in W_q^1(\Omega)$, $q > 2$, and proceeding in the same manner as in the case of our previous scheme where the coefficients a_{ij} were not mollified, we obtain the bound

$$\|e\|_{L_2(\Omega^h)} \leq C \sum_{i=1}^{2} \Biggl(\|\xi_{ii}\|_{L_2(\Omega^h)} + \|\xi_{i,3-i}\|_{L_2(\overline{\Omega}^h)} + \|\varrho_i\|_{L_2(\Omega^h_{i-1,2-i})} + \sum_{j=1}^{2} \|\eta_{ij1}\|_{L_2(\Omega_i^h)} \Biggr).$$

Using the estimates (2.187) and (2.172) derived earlier and slightly strengthening the smoothness requirements on the a_{ij} by demanding that $a_{ij} \in W_\infty^1(\Omega)$, we arrive at the error bound

$$\|u - U\|_{L_2(\Omega^h)} \leq Ch^2 \max_{i,j} \|a_{ij}\|_{W_\infty^1(\Omega)} \|u\|_{W_2^2(\Omega)}, \tag{2.191}$$

which is now *almost compatible* with the smoothness of the data in the sense that we assumed $a_{ij} \in W_\infty^1(\Omega)$ instead of the minimal smoothness requirement $a_{ij} \in W_q^1(\Omega)$, $q > 2$.

Let us now discuss the case when u belongs to the fractional-order Sobolev space $W_2^s(\Omega)$, $1 < s \leq 2$. Allowing some incompatibility between the smoothness of the coefficients and the corresponding solution by assuming instead of our initial hypothesis

$$u \in W_2^s(\Omega), \qquad a_{ij} \in W_p^{|s-1|+\delta}(\Omega), \quad 1 < s \leq 2,$$

that

$$u \in W_2^s(\Omega), \quad 1 < s \leq 2; \qquad a_{ij} \in W_\infty^1(\Omega)$$

and arguing as above, instead of (2.191) we arrive at the error bound

$$\|u - U\|_{L_2(\Omega^h)} \leq Ch^s \max_{i,j} \|a_{ij}\|_{W_\infty^1(\Omega)} \|u\|_{W_2^s(\Omega)}, \, 1 < s \leq 2.$$

This error bound is again incompatible with the smoothness of the data, except in the case of $s = 2$ when it coincides with (2.191).

2.6.4 Convergence in Discrete Fractional-Order Norms

By noting our error bounds in integer-order discrete Sobolev norms and the interpolation inequalities (2.54) we can obtain new bounds in fractional-order discrete

Sobolev norms. Thus, for example, for the scheme (2.167) from (2.168) and (2.178), we have that

$$\|u-U\|_{W_2^r(\Omega^h)} \le Ch^{s-r}\Big(\max_{i,j}\|a_{ij}\|_{W_2^{s-1}(\Omega)} + \|a\|_{W_2^{s-2}(\Omega)}\Big)\|u\|_{W_2^s(\Omega)},$$
$$\text{for } 1 \le r \le 2 < s \le 3.$$

From (2.169), (2.177) and (2.54) we deduce that

$$\|u-U\|_{W_2^r(\Omega^h)} \le Ch^{s-r}\Big(\max_{i,j}\|a_{ij}\|_{W_p^{s-1+\delta}(\Omega)} + \max_{i}\|a_i\|_{W_p^{s-1+\delta}(\Omega)}$$
$$+ \|a_0\|_{L_{2+\varepsilon}(\Omega)}\Big)\|u\|_{W_2^s(\Omega)}, \quad \text{for } 1 \le r < s \le 2.$$

Similarly, from (2.191), (2.54), the *inverse inequality*

$$|V|_{W_2^1(\Omega^h)} \le \frac{2\sqrt{2}}{h}\|V\|_{L_2(\Omega^h)} \quad \forall V \in S_0^h$$

with $V = e$ and (2.177) we obtain the following error bound for the difference scheme (2.190):

$$\|u-U\|_{W_2^r(\Omega^h)} \le Ch^{2-r}\max_{i,j}\|a_{ij}\|_{W_\infty^1(\Omega)}\|u\|_{W_2^2(\Omega)}, \quad 0 \le r \le 2.$$

In the next section we shall further sharpen these error bounds in the special case of an equation where the off-diagonal entries in the coefficient matrix (a_{ij}) are identically zero.

2.6.5 Convergence in the Discrete L_2 Norm: Separated Variables

In Sect. 2.6.3 we saw that the derivation of optimal error bounds in the $L_2(\Omega^h)$ norm under minimal smoothness requirements on the coefficients a_{ij} is associated with technical difficulties. The error bounds that we obtained are satisfactory in this respect only when $s = 2$, while for $s < 2$ they are incompatible with the natural minimal regularity requirements on the coefficients. These results can be improved in the case of a differential equation that separates the two variables; that is, when

$$\begin{aligned} -\sum_{i=1}^{2}\partial_i(a_i\partial_i u) = f \quad &\text{in } \Omega, \\ u = 0 \quad &\text{on } \Gamma = \partial\Omega, \end{aligned} \tag{2.192}$$

where

$$a_i = a_i(x_i), \quad i = 1, 2,$$

are such that there exist positive constants c_0 and c_1 with

$$0 < c_0 \le a_i(x_i) \le c_1 \quad \text{for all } x_i \in (0,1),\; i = 1,2.$$

In order to ensure that the a_i belong to the function space of multipliers $M(W_2^{s-1}(\Omega))$, we shall suppose that

$$a_i \in W_p^{|s-1|+\delta}(0,1),$$

where the real numbers s, p and δ are assumed to satisfy the following conditions:

$$\begin{aligned}
&p = 2, \qquad \delta = 0 \quad \text{when } |s-1| > 1/2,\\
&p > 2, \qquad \delta > 0 \quad \text{when } s = 1/2 \text{ or } s = 3/2,\\
&p \ge 1/|s-1|, \qquad \delta > 0 \quad \text{when } 0 < |s-1| < 1/2,\\
&p = \infty, \qquad \delta = 0 \quad \text{when } s = 1.
\end{aligned} \tag{2.193}$$

Let us introduce the following univariate mollifiers:

$$(S_i f)(x) := \frac{1}{h}\int_{x_i-h}^{x_i+h} \kappa_i(t) f\big(x + (t - x_i)\mathrm{e}_i\big)\,\mathrm{d}t, \quad i = 1,2,$$

where

$$\kappa_i(t) := \begin{cases} \int_{x_i-h}^{t} \frac{\mathrm{d}\tau}{a_i(\tau)} \Big/ \int_{x_i-h}^{x_i} \frac{\mathrm{d}\tau}{a_i(\tau)}, & t \in (x_i - h, x_i),\\ \int_{t}^{x_i+h} \frac{\mathrm{d}\tau}{a_i(\tau)} \Big/ \int_{x_i}^{x_i+h} \frac{\mathrm{d}\tau}{a_i(\tau)}, & t \in (x_i, x_i + h). \end{cases}$$

These operators satisfy the identity

$$S_i\big(\partial_i(a_i \partial_i u)\big) = D_{x_i}^-\big(\hat{a}_i D_{x_i}^+ u\big),$$

where $\hat{a}_i$ is the harmonic average of a_i, defined by

$$\hat{a}_i(x_i) := \left(\frac{1}{h}\int_{x_i}^{x_i+h} \frac{\mathrm{d}\tau}{a_i(\tau)}\right)^{-1}, \quad i = 1,2.$$

In particular when $a_i(x_i) \equiv 1$, we have that

$$S_i = T_i^2 = T_i^+ T_i^-.$$

We approximate the boundary-value problem (2.192) by the following finite difference scheme:

$$-\sum_{i=1}^{2} b_{3-i} D_{x_i}^-\big(\hat{a}_i D_{x_i}^+ U\big) = S_1 S_2 f \quad \text{in } \Omega^h, \tag{2.194}$$

$$U = 0 \quad \text{on } \Gamma^h, \tag{2.195}$$

where $b_i := S_i(1)$, $i = 1, 2$. We define the global error by

$$e := \overline{u} - U, \quad \text{where } \overline{u} := \begin{cases} T_h^{11} u, & \text{if } 0 < s \le 1, \\ u, & \text{if } 1 < s \le 2. \end{cases}$$

Then, e is easily seen to be a solution of the following finite difference scheme on the mesh $\overline{\Omega}^h$:

$$-\sum_{i=1}^{2} b_{3-i} D_{x_i}^{-}\big(\hat{a}_i D_{x_i}^{+} e\big) = \sum_{i=1}^{2} D_{x_i}^{-}\big(\hat{a}_i D_{x_i}^{+} \psi_i\big) \quad \text{in } \Omega^h,$$
$$e = 0 \quad \text{on } \Gamma^h,$$

where $\psi_i = S_{3-i}(u) - b_{3-i}\overline{u}$, $i = 1, 2$. It is easy to show by a duality argument (cf. the proof of Lemma 2.66) that

$$\|e\|_{L_2(\Omega^h)} \le C\big(\|\psi_1\|_{L_2(\Omega^h)} + \|\psi_2\|_{L_2(\Omega^h)}\big). \tag{2.196}$$

The task of deriving an error bound for the difference scheme (2.194) has thus been reduced to estimating the expression on the right-hand side of (2.196). We shall discuss the cases $1/2 < s \le 1$ and $1 < s \le 2$ separately.

First suppose that $1/2 < s \le 1$. Clearly, the value of ψ_i at a node $x \in \Omega^h$ is a bounded linear functional of $u \in W_2^s(K^0)$, $s > 1/2$, where

$$K^0 = K^0(x) = \big\{y = (y_1, y_2) : |y_j - x_j| < h,\ j = 1, 2\big\}.$$

Moreover, $\psi_i = 0$ when u is a constant function. By applying the Bramble–Hilbert lemma we deduce that

$$|\psi_i| \le C h^{s-1} |u|_{W_2^s(K^0)}, \quad 1/2 < s \le 1.$$

Summing over the nodes of the mesh Ω^h we obtain, for $i = 1, 2$, that

$$\|\psi_i\|_{L_2(\Omega^h)} \le C h^s |u|_{W_2^s(\Omega)}, \quad 1/2 < s \le 1. \tag{2.197}$$

Now let us consider the case $1 < s \le 2$. The key difficulty in obtaining an error bound is that ψ_{3-i} represents a nonlinear functional of a_i, $i = 1, 2$; nevertheless ψ_{3-i}, $i = 1, 2$, may be conveniently decomposed and, thereby, the nonlinear terms can be directly estimated. Let us write

$$\psi_{3-i} = \psi_{3-i,1} + \psi_{3-i,2} + \psi_{3-i,3},$$

where

$$\psi_{3-i,1} := \int_0^1 \big[u(x + h\tau \mathrm{e}_i) - 2u(x) + u(x - h\tau \mathrm{e}_i)\big]\bigg(\int_{x_i - h}^{x_i - h\tau} \frac{\mathrm{d}\sigma}{a_i(\sigma)}\bigg)$$

$$\times \left(\int_{x_i-h}^{x_i} \frac{\mathrm{d}\sigma}{a_i(\sigma)} \right)^{-1} \mathrm{d}\tau,$$

$$\psi_{3-i,2} := \int_0^1 \big[u(x+h\tau \mathrm{e}_i) - u(x) \big] \left(\int_{x_i}^{x_i+h} \frac{\mathrm{d}\sigma}{a_i(\sigma)} \int_{x_i-h}^{x_i} \frac{\mathrm{d}\sigma}{a_i(\sigma)} \right)^{-1}$$

$$\times \left(\int_{x_i+h\tau}^{x_i+h} \frac{\mathrm{d}\sigma}{a_i(\sigma)} \right) h^{-1} \int_{x_i-h}^{x_i} \int_{x_i}^{x_i+h} \frac{a_i(t) - a_i(t')}{a_i(t) a_i(t')} \, \mathrm{d}t \, \mathrm{d}t' \, \mathrm{d}\tau,$$

$$\psi_{3-i,3} := \int_0^1 \big[u(x+h\tau \mathrm{e}_i) - u(x) \big] \left(\int_{x_i-h}^{x_i} \frac{\mathrm{d}\sigma}{a_i(\sigma)} \right)^{-1}$$

$$\times h^{-1} (1-\tau)^{-1} \int_{x_i-h}^{x_i-h\tau} \int_{x_i+h\tau}^{x_i+h} \frac{a_i(t) - a_i(t')}{a_i(t) a_i(t')} \, \mathrm{d}t \, \mathrm{d}t' \, \mathrm{d}\tau.$$

The value of $\psi_{3-i,1}$ at $x \in \Omega^h$ is a bounded linear functional of $u \in W_2^s(K^0)$, $s > 1$, which vanishes whenever u is a polynomial of degree 1. Using the Bramble–Hilbert lemma we obtain

$$\|\psi_{3-i,1}\|_{L_2(\Omega^h)} \le C h^s |u|_{W_2^s(\Omega)}, \quad 1 < s \le 2. \tag{2.198}$$

For $3/2 < s \le 2$, $\psi_{3-i,2}$ is a bounded linear functional of $u \in W_2^s(K^0)$:

$$|\psi_{3-i,2}| \le C h^{\lambda - 1/2} \big(h^{-1} \|u\|_{L_2(K^0)} + |u|_{W_2^1(K^0)}$$

$$+ h^{s-1} |u|_{W_2^s(K^0)} \big) |a_i|_{W_2^\lambda(I^0)}, \lambda > 0,$$

where $I^0 = I^0(x_i) := (x_i - h, x_i + h)$. Moreover, $\psi_{3-i,2} = 0$ when u is a constant function, and therefore the term $h^{-1}\|u\|_{L_2(K^0)}$ on the right-hand side can be eliminated by applying the Bramble–Hilbert lemma. Summing over the nodes in the mesh Ω^h yields

$$\|\psi_{3-i,2}\|_{L_2(\Omega^h)} \le C h^{\lambda+1/2} \Big(\max_{x_i} |u|_{W_2^1(\Omega_{h,i})} + h^{s-1} |u|_{W_2^s(\Omega)} \Big) |a_i|_{W_2^\lambda(0,1)},$$

where

$$\Omega_{h,i} = \Omega_{h,i}(x) := \big\{ y \in \mathbb{R}^2 : \; x_i - h < y_i < x_i + h, \; 0 < y_{3-i} < 1 \big\}.$$

Choosing $\lambda = s - 1$ and invoking the boundary-layer estimate (see Oganesyan and Rukhovets [148], Chap. I, §8)

$$\|v\|_{L_2(0,\varepsilon)} \le C F(\varepsilon) \|v\|_{W_2^s(0,1)}, \quad 0 < \varepsilon < 1, \; 0 \le s \le 1, \tag{2.199}$$

where

$$F(\varepsilon) := \begin{cases} \varepsilon^s & 0 \le s < 1/2, \\ \varepsilon^{1/2} |\log \varepsilon| & s = 1/2, \\ \varepsilon^{1/2} & 1/2 < s \le 1, \end{cases}$$

which implies that

$$|u|_{W_2^1(\Omega_{h,i})} \le Ch^{1/2}\|u\|_{W_2^s(\Omega)}, \quad s > 3/2,$$

we thus obtain the bound

$$\|\psi_{3-i,2}\|_{L_2(\Omega^h)} \le Ch^s \|a_i\|_{W_2^{s-1}(0,1)}\|u\|_{W_2^s(\Omega)}, \quad 3/2 < s \le 2. \tag{2.200}$$

Similarly,

$$\|\psi_{3-i,2}\|_{L_2(\Omega^h)} \le Ch^s \|a_i\|_{W_p^{s-1+\delta}(0,1)}\|u\|_{W_2^s(\Omega)}, \quad 1 < s \le 3/2, \tag{2.201}$$

with p as in (2.193). An analogous bound holds for $\psi_{3-i,3}$. Combining (2.196) with (2.197), (2.198), (2.200) and (2.201) we thus obtain the following result.

Theorem 2.67 *Suppose that $u \in W_2^s(\Omega)$ and $a_i \in W_p^{|s-1|}(\Omega)$, $i = 1, 2$, with $1/2 < s \le 2$ and p as in (2.193). Then, the finite difference scheme (2.194) satisfies the error bound*

$$\|u - U\|_{L_2(\Omega^h)} \le Ch^s \max_i \|a_i\|_{W_p^{|s-1|+\delta}(0,1)}\|u\|_{W_2^s(\Omega)}, \tag{2.202}$$

where C is a positive constant, independent of h.

Unlike our earlier optimal error bounds in the $L_2(\Omega^h)$ norm, (2.202) is now also compatible with the smoothness of the coefficients.

We note that for $0 < s \le 1/2$ the function $S_1 S_2 f$, with $f \in W_2^{s-2}(\Omega)$, is not necessarily continuous on $\overline{\Omega}$; in this case the right-hand side of the difference scheme (2.194) is not defined at the mesh-points. A more fundamental difficulty is that $u \in W_2^s(\Omega)$ does not have a trace on $\Gamma = \partial\Omega$ when $s \le 1/2$, and it makes no sense, therefore, to demand that it satisfies a homogeneous Dirichlet boundary condition on Γ.

2.7 Fourth-Order Elliptic Equations

This section is devoted to boundary-value problems for fourth-order elliptic equations with variable coefficients of the form

$$\mathcal{L}u := \partial_1^2 M_1(u) + 2\partial_1\partial_2 M_3(u) + \partial_2^2 M_2(u) = f(x), \quad x \in \Omega, \tag{2.203}$$

where $\Omega = (0, 1)^2$ and

$$M_1(u) := a_1 \partial_1^2 u + a_0 \partial_2^2 u,$$

$$M_2(u) := a_0 \partial_1^2 u + a_2 \partial_2^2 u,$$

$$M_3(u) := a_3\partial_1\partial_2 u.$$

We shall assume that

$$\begin{aligned} &a_i \ge c_0 > 0, \quad i = 1, 2, 3, \qquad a_1 a_2 - a_0^2 \ge c_1 > 0, \quad x \in \Omega, \\ &u \in W_2^s(\Omega), \qquad f \in W_2^{s-4}(\Omega), \quad 2 < s \le 4. \end{aligned} \tag{2.204}$$

In order for (2.204) to hold it is necessary that the coefficients a_i belong to the multiplier space $M(W_2^{s-2}(\Omega))$. According to the results in Sect. 1.8, the following conditions are sufficient for that to be the case:

$$a_i \in W_p^{s-2+\varepsilon}(\Omega), \quad i = 0, 1, 2, 3, \tag{2.205}$$

where

$$\begin{aligned} &p = 2, \qquad \varepsilon = 0 \quad \text{when } 3 < s \le 4, \\ &p > 2, \qquad \varepsilon = 0 \quad \text{when } s = 3, \\ &p \ge 2/(s-2), \qquad \varepsilon > 0 \quad \text{when } 2 < s < 3. \end{aligned}$$

We begin by considering the partial differential equation (2.203) subject to the boundary conditions

$$\begin{aligned} u = 0 \quad &\text{on } \Gamma = \partial\Omega; \\ \partial_i^2 u = 0 \quad &\text{on } \Gamma_{i0} \cup \Gamma_{i1}, i = 1, 2, \end{aligned} \tag{2.206}$$

where

$$\Gamma_{ik} := \{x \in \Gamma : x_i = k, 0 < x_{3-i} < 1\}, \quad i, k = 0, 1.$$

By adopting the same notation as in Sects. 2.2.4 and 2.7 we approximate the boundary-value problem (2.203), (2.206) by the finite difference scheme

$$\mathcal{L}_h U = T_h^{22} f, \quad \text{on } \Omega^h, \tag{2.207}$$

$$\begin{aligned} U = 0, \quad &\text{on } \Gamma^h, \\ D_{x_i}^+ D_{x_i}^- U = 0, \quad &\text{on } \Gamma_{i0}^h \cup \Gamma_{i1}^h, \ i = 1, 2, \end{aligned} \tag{2.208}$$

where $\Gamma_{ik}^h = \Gamma_{ik} \cap \Gamma^h$,

$$\mathcal{L}_h U := D_{x_1}^+ D_{x_1}^- m_1(U) + 2D_{x_1}^- D_{x_2}^- m_3(U) + D_{x_2}^+ D_{x_2}^- m_2(U),$$

and

$$\begin{aligned} m_1(U) &:= a_1 D_{x_1}^+ D_{x_1}^- U + a_0 D_{x_2}^+ D_{x_2}^- U, \\ m_2(U) &:= a_0 D_{x_1}^+ D_{x_1}^- U + a_2 D_{x_2}^+ D_{x_2}^- U, \end{aligned}$$

$$m_3(U) := \hat{a}_3 D^+_{x_1} D^+_{x_2} U,$$

with

$$\hat{a}_3(x) := a_3\left(x_1 + \frac{1}{2}h, x_2 + \frac{1}{2}h\right).$$

Let us note that the difference scheme also involves mesh-points in $h\mathbb{Z}^2$ that are contained in $[-h, 1+h]^2$. Thus we shall suppose that the solution u and the coefficients a_i have been extended onto the larger square $(-h_0, 1+h_0)^2$ preserving their Sobolev class; here h_0 is a positive constant, $h_0 > h$.

Next we develop the error analysis of this finite difference scheme. The global error $e := u - U$ is easily seen to satisfy the following difference scheme:

$$\mathcal{L}_h e = D^+_{x_1} D^-_{x_1} \varphi_1 + 2 D^-_{x_1} D^-_{x_2} \varphi_3 + D^+_{x_2} D^-_{x_2} \varphi_2, \quad x \in \Omega^h, \tag{2.209}$$

$$\begin{aligned} e &= 0, \quad x \in \Gamma^h, \\ D^+_{x_i} D^-_{x_i} e &= D^+_{x_i} D^-_{x_i} u, \quad x \in \Gamma^h_{i0} \cup \Gamma^h_{i1}, \ i = 1, 2, \end{aligned} \tag{2.210}$$

where

$$\varphi_i := m_i(u) - T^2_{3-i} M_i(u), \quad i = 1, 2; \qquad \varphi_3 := m_3(u) - T^+_1 T^+_2 M_3(u).$$

Thus (2.206), (2.208) and (2.210) imply that

$$m_i(e) = \varphi_i, \quad x \in \Gamma^h_{i0} \cup \Gamma^h_{i1}, \ i = 1, 2.$$

By taking the inner product of (2.209) with e, performing summations by parts and applying the Cauchy–Schwarz inequality we get

$$\|e\|^2_{W^2_2(\Omega^h)} \le C\big(\|\varphi_1\|^2_{L_2(\Omega^h)} + \|\varphi_2\|^2_{L_2(\Omega^h)} + \|\varphi_3\|^2_{L_2(\Omega^h_{00})}\big). \tag{2.211}$$

Theorem 2.68 *Assuming that the data and the corresponding solution of the boundary-value problem* (2.203), (2.206) *obey the conditions* (2.204) *and* (2.205), *the difference scheme* (2.207), (2.208) *satisfies the error bound*

$$\|u - U\|_{W^2_2(\Omega^h)} \le C h^{s-2} \max_i \|a_i\|_{W^{s-2+\varepsilon}_p(\Omega)} \|u\|_{W^s_2(\Omega)}, \quad 5/2 < s \le 4. \tag{2.212}$$

Proof In order to prove the error bound (2.212) it suffices to bound the terms on the right-hand side of the inequality (2.211). Let us begin by representing φ_1 as the sum

$$\varphi_1 = \sum_{j=1}^{8} \varphi_{1,j},$$

where

$$\begin{aligned}
\varphi_{1,k} &:= a_{2-k}\big(D^+_{x_k}D^-_{x_k}u - T_1^2T_2^2\partial_k^2 u\big),\\
\varphi_{1,k+2} &:= \big(a_{2-k} - T_1^2T_2^2 a_{2-k}\big)\big(T_1^2T_2^2\partial_k^2 u\big),\\
\varphi_{1,k+4} &:= \big(T_1^2T_2^2 a_{2-k}\big)\big(T_1^2T_2^2\partial_k^2 u\big) - T_1^2T_2^2\big(a_{2-k}\partial_k^2 u\big),\\
\varphi_{1,k+6} &:= T_1^2T_2^2\big(a_{2-k}\partial_k^2 u\big) - T_2^2\big(a_{2-k}\partial_k^2 u\big), \quad k = 1, 2,
\end{aligned}$$

with an analogous representation of φ_2. Further, let

$$\varphi_3 = \varphi_{3,1} + \varphi_{3,2},$$

where

$$\begin{aligned}
\varphi_{3,1} &:= \big(\hat{a}_3 - T_1^+T_2^+a_3\big)D^+_{x_1}D^+_{x_2}u,\\
\varphi_{3,2} &:= \big(T_1^+T_2^+a_3\big)D^+_{x_1}D^+_{x_2}u - T_1^+T_2^+(a_3\partial_1\partial_2 u).
\end{aligned}$$

When $s \ge 2$, the value of $\varphi_{1,1}$ at a mesh-point $x \in \Omega^h$ is a bounded linear functional of $u \in W_2^s(K^0)$:

$$|\varphi_{1,1}| \le C(h)\|a_1\|_{C(\overline{\Omega})}\|u\|_{W_2^s(K^0)}.$$

Moreover, $\varphi_{1,1} = 0$ when u is a polynomial of degree 3. By the Bramble–Hilbert lemma,

$$|\varphi_{1,1}| \le Ch^{s-3}\|a_1\|_{C(\overline{\Omega})}|u|_{W_2^s(K^0)}, \quad 2 \le s \le 4.$$

By noting the Sobolev embedding $W_p^{s-2+\varepsilon}(K^0) \hookrightarrow C(\overline{K^0})$, $s > 2$, and summing over the mesh-points in Ω^h we thus obtain

$$\|\varphi_{1,1}\|_{L_2(\Omega^h)} \le Ch^{s-2}\|a_1\|_{W_p^{s-2+\varepsilon}(\Omega)}|u|_{W_2^s(\Omega)}, \quad 2 \le s \le 4. \tag{2.213}$$

The term $\varphi_{1,2}$ is bounded in the same way. Next $\varphi_{1,3}(x)$, $x \in \Omega^h$, is a bounded bilinear functional of $(a_1, u) \in W_p^\lambda(K^0) \times W_q^2(K^0)$, with $\lambda p > 2$; $q = \infty$ when $p = 2$; and $q = 2p/(p-2)$ when $p > 2$. Moreover, $\varphi_{1,3} = 0$ when either a_1 or u is a polynomial of degree 1. From the bilinear version of the Bramble–Hilbert lemma (cf. Lemma 2.30 with $m = 2$) we deduce that

$$|\varphi_{1,3}| \le Ch^{\lambda-1}\|a_1\|_{W_p^\lambda(K^0)}|u|_{W_q^2(K^0)}, \quad 2/p < \lambda \le 2,$$

and thereby

$$\|\varphi_{1,3}\|_{L_2(\Omega^h)} \le Ch^\lambda\|a_1\|_{W_p^\lambda(\Omega)}\|u\|_{W_q^2(\Omega)}.$$

By choosing $\lambda = s - 2 + \varepsilon$ and noting the Sobolev embeddings

$$W_2^s(\Omega) \hookrightarrow W_\infty^2(\Omega), \quad s > 3,$$

and

$$W_2^s(\Omega) \hookrightarrow W_{2p/(p-2)}^2(\Omega), \quad 2 < s \leq 3,$$

we obtain

$$\|\varphi_{1,3}\|_{L_2(\Omega^h)} \leq Ch^{s-2} \|a_1\|_{W_p^{s-2+\varepsilon}(\Omega)} \|u\|_{W_2^s(\Omega)}, \quad 2 < s \leq 4. \tag{2.214}$$

The terms $\varphi_{1,4}$ and $\varphi_{3,1}$ are bounded in the same way.

For $\lambda \geq 0$, $\mu \geq 2$ and $q > 2$ the value of $\varphi_{1,5}(x)$ at $x \in \Omega^h$ is a bounded bilinear functional of $(a_1, u) \in W_q^\lambda(K^0) \times W_{2q/(q-2)}^\mu(K^0)$. Furthermore, $\varphi_{1,5} = 0$ when a_1 is a constant function or when u is a polynomial of degree 2. By the bilinear version of the Bramble–Hilbert lemma,

$$\|\varphi_{1,5}\|_{L_2(\Omega^h)} \leq Ch^{\lambda+\mu-2} \|a_1\|_{W_q^\lambda(\Omega)} \|u\|_{W_{2q/(q-2)}^\mu(\Omega)},$$

where $0 \leq \lambda \leq 1$ and $2 \leq \mu \leq 3$. Now let $\lambda + \mu = s$. When $\lambda + \mu > 3$, there exists a $q = q(\lambda, \mu)$ such that $\lambda \geq 2/q \geq 3 - \mu$; then,

$$W_p^{s-2+\varepsilon}(\Omega) = W_2^{\lambda+\mu-2+\varepsilon}(\Omega) \hookrightarrow W_q^\lambda(\Omega)$$

and

$$W_2^s(\Omega) = W_2^{\lambda+\mu}(\Omega) \hookrightarrow W_{2q/(q-2)}^\mu(\Omega).$$

Analogously, when $2 < \lambda + \mu \leq 3$, there exists a real number q such that $\lambda \geq 2/q \geq 2/p - (\mu - 2)$. In this case,

$$W_p^{s-2+\varepsilon}(\Omega) = W_p^{\lambda+\mu-2+\varepsilon}(\Omega) \hookrightarrow W_q^\lambda(\Omega)$$

and

$$W_2^s(\Omega) = W_2^{\lambda+\mu}(\Omega) \hookrightarrow W_{2q/(q-2)}^\mu(\Omega).$$

It follows from these embeddings that

$$\|\varphi_{1,5}\|_{L_2(\Omega^h)} \leq Ch^{s-2} \|a_1\|_{W_p^{s-2+\varepsilon}(\Omega)} \|u\|_{W_2^s(\Omega)}, \quad 2 < s \leq 4. \tag{2.215}$$

The terms $\varphi_{1,6}$ and $\varphi_{3,2}$ are bounded in the same way.

When $\lambda > 1/2$, the value of $\varphi_{1,7}(x)$ at $x \in \Omega^h$ is a bounded linear functional of $a_1 \partial_1^2 u \in W_2^\lambda(K^0)$, which vanishes on all polynomials of degree 1. By the Bramble–Hilbert lemma, we have that

$$\|\varphi_{1,7}\|_{L_2(\Omega^h)} \leq Ch^\lambda |a_1 \partial_1^2 u|_{W_2^\lambda(\Omega)}, \quad 1/2 < \lambda \leq 2.$$

By choosing $\lambda = s - 2$, the inequality

$$\left|a_1 \partial_1^2 u\right|_{W_2^\lambda(\Omega)} \le C \|a_1\|_{W_p^{\lambda+\varepsilon}(\Omega)} \left\|\partial_1^2 u\right\|_{W_2^\lambda(\Omega)}$$

implies that

$$\|\varphi_{1,7}\|_{L_2(\Omega^h)} \le C h^{s-2} \|a_1\|_{W_p^{s-2+\varepsilon}(\Omega)} \|u\|_{W_2^s(\Omega)}, \quad 5/2 < s \le 4. \tag{2.216}$$

The term $\varphi_{1,8}$ is bounded in the same way. Finally (2.213)–(2.216) and (2.211) yield the desired error bound (2.212). □

We note that for $2 < s \le 5/2$ the function $T_h^{22} f$ is not necessarily continuous on $\overline{\Omega}$ and therefore the right-hand side in the difference equation (2.207) is not defined for this range of values of the Sobolev index s. In fact, for $s \le 5/2$, the second-normal derivative of $u \in W_2^s(\Omega)$ does not have a trace on $\Gamma_{i0} \cup \Gamma_{i1}$ and therefore the boundary-value problem (2.203)–(2.206) is not meaningful as stated for this range of s.

Now let us consider the partial differential equation (2.203) subject to the homogeneous Dirichlet boundary conditions

$$\begin{aligned} u &= 0 \quad \text{on } \Gamma, \\ \partial_i u &= 0 \quad \text{on } \Gamma_{i0} \cup \Gamma_{i1},\ i = 1, 2. \end{aligned} \tag{2.217}$$

With the notational conventions from Sects. 2.2.4 and 2.7 equation (2.203) is again approximated by (2.207), and the boundary conditions (2.217) are discretized as follows:

$$\begin{aligned} U &= 0 \quad \text{on } \Gamma^h, \\ D_{x_i}^0 U &= 0 \quad \text{on } \Gamma_{i0}^h \cup \Gamma_{i1}^h,\ i = 1, 2. \end{aligned} \tag{2.218}$$

The error $e := u - U$ satisfies (2.209) and the boundary conditions

$$\begin{aligned} e &= 0 \quad \text{on } \Gamma^h, \\ D_{x_i}^0 e &= D_{x_i}^0 u \quad \text{on } \Gamma_{i0}^h \cup \Gamma_{i1}^h,\ i = 1, 2. \end{aligned} \tag{2.219}$$

Defining $\zeta_i = \zeta_i(x)$ by

$$\zeta_i := \left(D_{x_i}^0 u - \partial_i u\right)/h, \quad i = 1, 2,$$

the derivative boundary condition in (2.219) can be rewritten as

$$D_{x_i}^0 e = h\zeta_i, \quad x \in \Gamma_{i0}^h \cup \Gamma_{i1}^h,\ i = 1, 2.$$

Theorem 2.69 *The following bound holds on the global error $e := u - U$ between the analytical solution u and its finite difference approximation U:*

$$\|u - U\|_{W_2^2(\Omega^h)} \le Ch^{\min\{s-2,3/2\}} |\log h|^{1-|\mathrm{sgn}(s-7/2)|} \times \max_i \|a_i\|_{W_p^{s-2+\varepsilon}(\Omega)} \|u\|_{W_2^s(\Omega)}, \quad 5/2 < s < 4. \tag{2.220}$$

Proof We begin by noting that

$$\|e\|^2_{W_2^2(\Omega^h)} \le C\Bigg(\|\varphi_1\|^2_{L_2(\Omega_1^h \cup \Gamma_{11}^h)} + \|\varphi_2\|^2_{L_2(\Omega_2^h \cup \Gamma_{21}^h)} + \|\varphi_3\|^2_{L_2(\Omega_{00}^h)} + \sum_{i=1}^{2} h^2 \sum_{x \in \Gamma_{i0}^h \cup \Gamma_{i1}^h} \zeta_i^2(x) \Bigg). \tag{2.221}$$

The first three terms on the right-hand side of (2.221) are bounded in the same way as in the case of the boundary-value problem (2.203), (2.204) considered earlier. The only new ingredient in the analysis is the estimation of the last term in (2.221), which we discuss below.

When $s > 2$, ζ_i represents a bounded linear functional of $u \in W_2^s(K^0)$, which vanishes on all polynomials of degree 2. By applying the Bramble–Hilbert lemma we obtain

$$\Bigg(h^2 \sum_{x \in \Gamma_{i0}^h} \zeta_i^2(x) \Bigg)^{1/2} \le Ch^{s-2} |u|_{W_2^s(\Omega_{i0})}, \quad 2 < s \le 3, \tag{2.222}$$

where

$$\Omega_{i0} = \Omega_{hi}(0) := \{x : -h < x_i < h,\ 0 < x_{3-i} < 1\}.$$

By noting the boundary-layer estimate (2.199), we deduce from (2.222) that

$$\Bigg(h^2 \sum_{x \in \Gamma_{i0}^h} \zeta_i^2 \Bigg)^{1/2} \le Ch^{\min\{s-2,3/2\}} |\log h|^{1-|\mathrm{sgn}(s-7/2)|} \|u\|_{W_2^s(\Omega)}, \quad 2 < s \le 4. \tag{2.223}$$

For $x \in \Gamma_{i1}^h$ the terms ζ_i, $i = 1, 2$, are bounded analogously. From (2.221), (2.223) and our earlier bounds on φ_1, φ_2 and φ_3 we obtain the desired error bound (2.220) for the difference scheme (2.207), (2.218).

For $s < 7/2$ the solution of (2.203), (2.217) has an even extension (i.e. an extension as an even function) across Γ that preserves the Sobolev class W_2^s. With such an even extension of u, $\zeta_i = 0$ on $\Gamma_{i0}^h \cup \Gamma_{i1}^h$, and (2.220) is then a direct consequence of (2.207)–(2.216) and (2.221). $\square$

Finally, we consider the partial differential equation (2.203) subject to the *natural boundary conditions*

$$\begin{aligned} &M_i(u) = 0 \quad \text{and} \quad \partial_i M_i(u) + 2\partial_{3-i} M_3(u) = 0 \quad \text{on } \Gamma_{i0} \cup \Gamma_{i1},\ i = 1, 2; \\ &M_3(u) = 0 \quad \text{on } \Gamma_* = \big\{(0,0), (0,1), (1,0), (1,1)\big\}. \end{aligned} \tag{2.224}$$

The solution of problem (2.203), (2.224) is unique, up to the addition of a polynomial of degree 1. In order to ensure that we have a unique solution, we shall assume that, in addition to (2.224), the values of u at three vertices of Ω have been fixed; that is,

$$u(0,0) = c_{00}, \qquad u(0,1) = c_{01}, \qquad u(1,0) = c_{10}. \tag{2.225}$$

With the notational conventions from Sects. 2.2.4 and 2.7, the conditions (2.224), (2.225) are approximated by

$$\begin{aligned} m_i(U) = 0, \quad D^0_{x_i} m_i(U) + D^-_{x_{3-i}}\big[m_3(U) + m_3(U)^{-i}\big] = 0, \\ \text{on } \overline{\Gamma}^h_{i0} \cup \overline{\Gamma}^h_{i1},\ i = 1, 2; \end{aligned} \tag{2.226}$$

$$\begin{aligned} &m_3(U) + m_3(U)^{-1} + m_3(U)^{-2} + m_3(U)^{-1,-2} = 0 \quad \text{on} \Gamma_*; \\ &U(0,0) = c_{00}, \qquad U(0,1) = c_{01}, \qquad U(1,0) = c_{10}, \end{aligned} \tag{2.227}$$

where $\overline{\Gamma}^h_{ik} := \overline{\Gamma}_{ik} \cap \Gamma^h$. Let us observe that the difference scheme also involves points exterior to $\overline{\Omega}$ that are at a distance $\le 2h$ from Γ; therefore (2.203), (2.226), (2.227) has fewer equations than unknowns. In order to account for the missing equations, we also discretize the partial differential equation at the boundary mesh-points. Let us introduce the asymmetric mollifiers

$$T_i^{2\pm} f := 2\int_0^1 (1-t) f(x \pm th\mathrm{e}_i)\,\mathrm{d}t, \quad i = 1, 2,$$

and the additional equations

$$\mathcal{L}_h U = \begin{cases} T_i^{2+} T_{3-i}^2 f & \text{for } x \in \Gamma^h_{i0} \\ T_i^{2-} T_{3-i}^2 f & \text{for } x \in \Gamma^h_{i1}, \\ T_1^{2+} T_2^{2+} f & \text{for } x = (0,0), \\ \text{and analogously} & \text{for } x = (0,1), (1,0), (1,1). \end{cases} \tag{2.228}$$

Theorem 2.70 *The difference scheme* (2.203), (2.226), (2.228) *satisfies the error bound*

$$\begin{aligned} [\![u - U]\!]_{W_2^2(\Omega^h)} \le C h^{\min\{s-2, 3/2\}} |\log h|^{1-|\mathrm{sgn}(s-7/2)|} \\ \times \max_i \|a_i\|_{W_p^{s-2+\varepsilon}(\Omega)} \|u\|_{W_2^s(\Omega)}, \quad 3 < s \le 4, \end{aligned}$$

where C is a positive constant, independent of h.

Proof The global error $e := u - U$ satisfies the inequality

$$\begin{aligned}\|e\|^2_{W_2^2(\Omega^h)} \le C\big(&\|\varphi_1\|^2_{L_2(\Omega^h)} + \|\varphi_2\|^2_{L_2(\Omega^h)} + \|\varphi_3\|^2_{L_2(\Omega^h_{00})} \\ &+ \|\phi_1\|^2_{L_2(\Omega^h)} + \|\phi_2\|^2_{L_2(\Omega^h)}\big),\end{aligned} \tag{2.229}$$

where, for $i = 1, 2$,

$$\phi_i := \begin{cases} T^2_{3-i} M_i(u) - T^{2+}_{3-i} M_i(u) & \text{on } \overline{\Gamma}^h_{i0}, \\ T^2_{3-i} M_i(u) - T^{2-}_{3-i} M_i(u) & \text{on } \overline{\Gamma}^h_{i1}, \\ 0 \quad \text{at the remaining mesh-points.} \end{cases}$$

The terms φ_1, φ_2 and φ_3 are estimated in the same way as before. Finally, ϕ_i is a bounded linear functional of $M_i(u) \in W_2^\lambda(\Omega)$, $\lambda > 1/2$, which vanishes on all constant functions. Using the Bramble–Hilbert lemma and the boundary-layer estimate (2.199) we obtain

$$\begin{aligned}\|\phi_i\|_{L_2(\Omega^h)} \le C h^{\min\{s-2,3/2\}} |\log h|^{1-|\mathrm{sgn}(s-7/2)|} \\ \times \max_j \|a_j\|_{W_p^{s-2+\varepsilon}(\Omega)} \|u\|_{W_2^s(\Omega)}, \quad 3 < s \le 4.\end{aligned} \tag{2.230}$$

The desired error bound follows from (2.229), (2.230) and our earlier bounds on the terms φ_1, φ_2 and φ_3. □

2.8 An Elliptic Interface Problem

The technique of convergence analysis introduced in earlier sections of this chapter can be extended to finite difference schemes for more general boundary-value problems. As an example, we consider here a model partial differential equation with a singular coefficient. Problems of the kind discussed here are usually referred to as *interface problems* or *transmission problems*. For further details we refer the reader to Jovanović and Vulkov [101].

Let $\Omega = (0, 1)^2$ and $\Gamma = \partial\Omega$. A typical point in $\overline{\Omega}$ will be denoted by $x = (x_1, x_2)$. Let further Σ be the intersection of the line segment $x_2 = \xi$, $0 < \xi < 1$, with $\overline{\Omega}$. We consider the Dirichlet boundary-value problem

$$\mathcal{L}u + k(x)\delta_\Sigma(x)u = f(x) \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \Gamma, \tag{2.231}$$

where $\delta_\Sigma(x) = \delta(x_2 - \xi)$ is the Dirac distribution concentrated on Σ, $k(x) = k(x_1)$ and $\mathcal{L}$ is the symmetric elliptic operator introduced in (2.166); i.e.

$$\mathcal{L}u := -\sum_{i,j=1}^{2} \partial_i(a_{ij}\partial_j u) + au.$$

The Dirac distribution δ_Σ belongs to the Sobolev space $W_2^{-\lambda}(\Omega)$, with $\lambda > 1/2$. Equation (2.231) must be therefore understood in a weak sense: we seek $u \in \mathring{W}_2^1(\Omega)$

such that

$$\langle \mathcal{L}u, v\rangle + (k\delta_\Sigma)(uv) = \langle f, v\rangle \quad \forall v \in \mathring{W}_2^1(\Omega), \tag{2.232}$$

where $\langle f, v\rangle$ denotes the duality pairing between the spaces $W_2^{-1}(\Omega)$ and $\mathring{W}_2^1(\Omega)$, and

$$(k\delta_\Sigma)(w) := \int_\Sigma kw\bigg|_\Sigma \,\mathrm{d}\Sigma, \quad w \in W_1^1(\Omega),$$

where $w|_\Sigma \in L_1(\Sigma)$ denotes the trace of $w \in W_1^1(\Omega)$ on Σ, and $k \in L_\infty(\Sigma)$.

Alternatively, problem (2.232) can be restated as follows: *find* $u \in \mathring{W}_2^1(\Omega)$ *such that*

$$a(u, v) = \langle f, v\rangle \quad \forall v \in \mathring{W}_2^1(\Omega), \tag{2.233}$$

where

$$a(u, v) = \int_\Omega \left(\sum_{i,j=1}^{2} a_{ij}\partial_j u \partial_i v + auv\right) \mathrm{d}x + \int_\Sigma k(uv)\bigg|_\Sigma \,\mathrm{d}\Sigma. \tag{2.234}$$

Thus, (2.233) can be seen as the weak formulation of the boundary-value problem (2.231). A relevant point in this respect is that for the domain $\Omega = (0, 1)^2 \subset \mathbb{R}^2$ the product uv of $u, v \in \mathring{W}_2^1(\Omega)$ belongs to $\mathring{W}_p^1(\Omega)$ for all $p \in [1, 2)$ and thus by Theorem 1.42 (see also Theorem 1.5.1.3 on p. 38 of Grisvard [62] for $p \in (1, 2)$ and Theorem 2.10 on p. 37 of Giusti [54] for $p = 1$), the boundary integral term in (2.234) is meaningful. The following assertion concerning the existence of a unique weak solution is an immediate consequence of the Lax–Milgram theorem and the trace theorem for $W_2^1(\Omega)$.

Lemma 2.71 *Suppose that*

$$f \in W_2^{-1}(\Omega), \quad a_{ij}, a \in L_\infty(\Omega), \quad k \in L_\infty(\Sigma), \quad a_{ij} = a_{ji}, \quad a \geq 0, \quad k \geq 0,$$

$$\exists c_0 > 0\ \forall \xi = (\xi_1, \xi_2) \in \mathbb{R}^2\ \forall x \in \Omega: \quad \sum_{i,j=1}^{2} a_{ij}(x)\xi_i\xi_j \geq c_0 \sum_{i=1}^{2} \xi_i^2.$$

Then, there exists a unique weak solution $u \in \mathring{W}_2^1(\Omega)$ *to the boundary-value problem* (2.233), (2.234), *and*

$$\|u\|_{W_2^1(\Omega)} \leq C\|f\|_{W_2^{-1}(\Omega)}.$$

Let us now assume that the coefficients a_{ij}, $i, j = 1, 2$, and a of the differential operator $\mathcal{L}$ belong to the Hölder space $C^{0,\lambda}(\overline{\Omega})$, with $\lambda > |\theta|$ and $|\theta| < 1/2$. The bilinear functional $a(\cdot, \cdot)$ can then be continuously extended to $\mathring{W}_2^{1-\theta}(\Omega) \times \mathring{W}_2^{1+\theta}(\Omega)$. The following assertion can be proved by applying Theorem 3.3 in Nečas [143].

Lemma 2.72 *Suppose that*

$$f \in W_2^{\theta-1}(\Omega), \quad |\theta| < 1/2, \quad a_{ij}, a \in C^{0,\lambda}(\overline{\Omega}), \quad \lambda > |\theta|, \quad k \in L_\infty(\Sigma),$$
$$a_{ij} = a_{ji}, \quad a \ge 0, \quad k \ge 0,$$
$$\exists c_0 > 0 \, \forall \xi = (\xi_1, \xi_2) \in \mathbb{R}^2 \, \forall x \in \Omega : \quad \sum_{i,j=1}^{2} a_{ij}(x)\xi_i\xi_j \ge c_0 \sum_{i=1}^{2} \xi_i^2.$$

Then, there exists a unique solution $u \in \mathring{W}_2^{1+\theta}(\Omega)$ to the boundary-value problem (2.233), (2.234).

In the case when f does not contain a concentrated singularity on Σ, such as δ_Σ, problem (2.233), (2.234) can be shown to be the weak formulation of the following boundary-value problem with transmission (conjugation) conditions on the interface Σ:

$$\begin{gathered} \mathcal{L}u = f \quad \text{in } \Omega^- \cup \Omega^+, \qquad u = 0 \quad \text{on } \Gamma, \\ [u]_\Sigma = 0, \quad \left[\sum_{j=1}^{2} a_{2j}\partial_j u\right]_\Sigma = ku|_\Sigma, \end{gathered} \tag{2.235}$$

where $\Omega^- := (0, 1) \times (0, \xi)$, $\Omega^+ := (0, 1) \times (\xi, 1)$, and

$$[u]_\Sigma := u(x_1, \xi + 0) - u(x_1, \xi - 0).$$

In this sense, the boundary-value problems (2.231) and (2.235) are equivalent.

Higher regularity of the solution can be proved under additional assumptions on the data. For $s \ge 2$ we define the subspace $\widehat{W}_2^s(\Omega)$ of $\mathring{W}_2^1(\Omega)$, consisting of all $u \in \mathring{W}_2^1(\Omega)$ such that

$$\begin{gathered} \partial_1^i u \in L_2(\Omega), \quad i = 0, 1, \ldots, s, \\ \partial_1^{i-1}\partial_2 u \in L_2(\Omega), \quad i = 1, 2, \ldots, s, \\ \partial_1^{i-j}\partial_2^j u \in L_2(\Omega^-) \cap L_2(\Omega^+), \quad i = j, j+1, \ldots, s, \; j = 2, 3, \ldots, s, \end{gathered}$$

with the norm $\|\cdot\|_{\widehat{W}_2^s(\Omega)}$ defined by

$$\begin{aligned} \|u\|^2_{\widehat{W}_2^s(\Omega)} := {} & \sum_{i=0}^{s} \|\partial_1^i u\|^2_{L_2(\Omega)} + \sum_{i=1}^{s} \|\partial_1^{i-1}\partial_2 u\|^2_{L_2(\Omega)} \\ & + \sum_{j=2}^{s}\sum_{i=j}^{s} \left(\|\partial_1^{i-j}\partial_2^j u\|^2_{L_2(\Omega^-)} + \|\partial_1^{i-j}\partial_2^j u\|^2_{L_2(\Omega^+)} \right). \end{aligned}$$

Obviously,

$$\widehat{W}_2^s(\Omega) \subset \widetilde{W}_2^s(\Omega) := \mathring{W}_2^1(\Omega) \cap W_2^s(\Omega^-) \cap W_2^s(\Omega^+).$$

Lemma 2.73 *Suppose that in addition to the assumptions of Lemma 2.71 we have that*

$$f \in L_2(\Omega), \qquad a_{ij} \in W^1_\infty(\Omega), \qquad k \in W^1_\infty(\Sigma);$$

then, $u \in \widehat{W}^2_2(\Omega)$. *If, in addition,*

$$\partial_1 f \in L_2(\Omega), \qquad \partial_2 f \in L_2\big(\Omega^\pm\big), \qquad a_{ij} \in W^2_\infty(\Omega),$$
$$a \in W^1_\infty(\Omega), \qquad k \in W^2_\infty(\Sigma)$$

and

$$f = a_{12} = \partial_1 a_{11} = 0 \quad \textit{for } x_1 = 0 \textit{ and } x_1 = 1,$$

then $u \in \widehat{W}^3_2(\Omega)$.

Proof For $x \in \Omega^- \cup \Omega^+$ (2.235) can be written as

$$a_{11}\partial_1^2 u + 2a_{12}\partial_1\partial_2 u + a_{22}\partial_2^2 u = -\sum_{i,j=1}^{2} \partial_i a_{ij}\partial_j u + au - f. \tag{2.236}$$

Multiplying (2.236) by $\partial_1^2 u$, integrating over Ω and performing partial integration we obtain

$$\int_\Omega \big[a_{11}\big(\partial_1^2 u\big)^2 + 2a_{12}\partial_1^2 u\,\partial_1\partial_2 u + a_{22}(\partial_1\partial_2 u)^2\big]\,\mathrm{d}x + \int_\Sigma k(\partial_1 u)^2\Big|_\Sigma \,\mathrm{d}\Sigma$$
$$= I_1 + I_2 + I_3,$$

where

$$I_1 := -\int_\Omega \left(\sum_{i,j=1}^{2} \partial_i a_{ij}\partial_j u - au + f\right)\partial_1^2 u\,\mathrm{d}x,$$

$$I_2 := \int_\Omega \big(\partial_2 a_{22}\partial_2 u\,\partial_1^2 u - \partial_1 a_{22}\partial_2 u\,\partial_1\partial_2 u\big)\,\mathrm{d}x,$$

$$I_3 := -\int_\Sigma k' u\,\partial_1 u\,\mathrm{d}\Sigma.$$

Further,

$$\int_\Omega \big[a_{11}\big(\partial_1^2 u\big)^2 + 2a_{12}\partial_1^2 u\,\partial_1\partial_2 u + a_{22}(\partial_1\partial_2 u)^2\big]\,\mathrm{d}x + \int_\Sigma k(\partial_1 u)^2\,\mathrm{d}\Sigma$$
$$\geq c_0\big(\big\|\partial_1^2 u\big\|^2_{L_2(\Omega)} + \|\partial_1\partial_2 u\|^2_{L_2(\Omega)}\big).$$

The integrals I_1, I_2 and I_3 can be bounded by applying the Cauchy–Schwarz inequality with $\varepsilon \in (0,1)$ as follows:

$$|I_1| \le \varepsilon \big\| \partial_1^2 u \big\|^2_{L_2(\Omega)} + \frac{C}{\varepsilon}\big(\|u\|^2_{W_2^1(\Omega)} + \|f\|^2_{L_2(\Omega)}\big).$$

Similarly,

$$|I_2| \le \varepsilon\big(\big\| \partial_1^2 u \big\|^2_{L_2(\Omega)} + \|\partial_1\partial_2 u\|^2_{L_2(\Omega)}\big) + \frac{C}{\varepsilon}\|\partial_2 u\|^2_{L_2(\Omega)}$$

and

$$\begin{aligned}|I_3| &\le \varepsilon \|\partial_1 u\|^2_{L_2(\Sigma)} + \frac{C}{\varepsilon}\|u\|^2_{L_2(\Sigma)} \\ &\le C_1\varepsilon\big(\big\| \partial_1^2 u \big\|^2_{L_2(\Omega)} + \|\partial_1\partial_2 u\|^2_{L_2(\Omega)}\big) + \frac{C}{\varepsilon}\|u\|^2_{W_2^1(\Omega)}.\end{aligned}$$

Hence, by selecting a sufficiently small $\varepsilon > 0$, we obtain the bound

$$\big\| \partial_1^2 u \big\|^2_{L_2(\Omega)} + \|\partial_1\partial_2 u\|^2_{L_2(\Omega)} \le C\|f\|^2_{L_2(\Omega)}.$$

From (2.236) we immediately have that

$$\big\| \partial_2^2 u \big\|_{L_2(\Omega^\pm)} \le C\big(\big\| \partial_1^2 u \big\|_{L_2(\Omega)} + \|\partial_1\partial_2 u\|_{L_2(\Omega)} + \|u\|_{W_2^1(\Omega)} + \|f\|_{L_2(\Omega)}\big),$$

which proves the first part of the lemma.

When the assumptions of the second part of the lemma are satisfied, we deduce from (2.231) that

$$\partial_1^2 u = 0 \quad \text{on } \Gamma.$$

By differentiating (2.231) one obtains

$$\mathcal{L}\partial_1^2 u + k(x)\delta_\Sigma(x)\partial_1^2 u = f_1(x), \quad x \in \Omega,$$

where

$$\begin{aligned} f_1 := \partial_1^2 f &+ \sum_{i,j=1}^{2} \partial_i\big(2\partial_1 a_{ij}\partial_1\partial_j u + \partial_1^2 a_{ij}\partial_j u\big) \\ &- 2\partial_1 a\partial_1 u - \partial_1^2 a u - 2k'\delta_\Sigma\partial_1 u - k''\delta_\Sigma u \in W_2^{-1}(\Omega).\end{aligned}$$

By applying Lemma 2.71 we then deduce the regularity result stated in the second part of the lemma. □

For further details regarding the analysis of elliptic boundary-value problems in domains with corners we refer to Grisvard [62] and Dauge [28].

2.8.1 Finite Difference Approximation

In the sequel we shall assume that the weak solution of the boundary-value problem (2.231) belongs to $\widetilde{W}_2^s(\Omega)$, $s > 2$, and that the coefficients of the equation satisfy the following regularity hypotheses:

$$a_{ij} \in W_2^{s-1}(\Omega^-) \cap W_2^{s-1}(\Omega^+) \cap C(\overline{\Omega}), \qquad a \in W_2^{s-2}(\Omega^-) \cap W_2^{s-2}(\Omega^+)$$

and

$$k \in W_2^{s-1}(\Sigma).$$

We define

$$\|u\|_{\widetilde{W}_2^s(\Omega)} := \left(\|u\|^2_{W_2^1(\Omega)} + \|u\|^2_{W_2^s(\Omega^-)} + \|u\|^2_{W_2^s(\Omega^+)}\right)^{1/2}.$$

In particular, for $s = 0$ we set

$$\|u\|_{\widetilde{W}_2^0(\Omega)} = \|u\|_{\widetilde{L}_2(\Omega)} := \left(\|u\|^2_{L_2(\Omega)} + \|u\|^2_{L_2(\Sigma)}\right)^{1/2}.$$

For the sake of simplicity we shall also assume that ξ is a rational number. Let $\overline{\Omega}^h$ be a uniform square mesh on $\overline{\Omega}$ with mesh-size $h := 1/N$, where N is an integer such that ξN is also an integer. We shall use the notations from Sect. 2.2 and define

$$\Sigma^h := \Omega^h \cap \Sigma \quad \text{ and } \quad \Sigma^h_- := \Sigma^h \cup \{(0, \xi)\}.$$

Let us approximate the boundary-value problem (2.231) on the mesh $\overline{\Omega}^h$ by the following finite difference scheme with mollified right-hand side:

$$\mathcal{L}_h U + k\delta_{\Sigma^h} U = T_1^2 T_2^2 f \quad \text{in } \Omega^h, \qquad U = 0 \quad \text{on } \Gamma^h, \tag{2.237}$$

where

$$\mathcal{L}_h U := -\frac{1}{2}\sum_{i,j=1}^{2}\left[D^+_{x_i}\left(a_{ij} D^-_{x_j} U\right) + D^-_{x_i}\left(a_{ij} D^+_{x_j} U\right)\right] + \left(T_1^2 T_2^2 a\right) U$$

and

$$\delta_{\Sigma^h}(x) = \delta_h(x_2 - \xi) := \begin{cases} 0 & \text{for } x \in \Omega^h \setminus \Sigma^h, \\ 1/h & \text{for } x \in \Sigma^h \end{cases}$$

is the discrete Dirac delta-function.

Further, we define the asymmetric mollifiers T_2^{2-} and T_2^{2+} by

$$T_2^{2-} f(x_1, x_2) := \frac{2}{h}\int_{x_2-h}^{x_2}\left(1 + \frac{x_2' - x_2}{h}\right) f\left(x_1, x_2'\right) \mathrm{d}x_2',$$

$$T_2^{2+} f(x_1, x_2) := \frac{2}{h} \int_{x_2}^{x_2+h} \left(1 - \frac{x_2' - x_2}{h}\right) f\left(x_1, x_2'\right) \mathrm{d}x_2'.$$

In addition to the discrete inner products and norms defined in Sect. 2.6.1 we introduce

$$(U, V)_{\Sigma^h} := h \sum_{x \in \Sigma^h} U(x)V(x), \qquad \|U\|_{L_2(\Sigma^h)} := (U, U)_{\Sigma^h}^{1/2},$$
$$|U|_{W_2^{1/2}(\Sigma^h)} := \left(h^2 \sum_{x \in \Sigma_-^h} \sum_{x' \in \Sigma_-^h, x' \neq x} \frac{|U(x) - U(x')|^2}{|x_1 - x_1'|^2}\right)^{1/2}.$$

The following lemma holds.

Lemma 2.74 *Let $U \in S_0^h$ and let V be a mesh-function defined on Σ_-^h. Then,*

$$\left|\left(D_{x_1}^- V, U\right)_{\Sigma^h}\right| \leq C\|U\|_{W_2^1(\Omega^h)} |V|_{W_2^{1/2}(\Sigma^h)}.$$

Proof Similarly as in the proof of Lemma 2 in Jovanović and Popović [92], we expand U and V in the following Fourier sums:

$$U(x_1, x_2) = \sum_{k=1}^{N-1} \sum_{l=1}^{N-1} b_{kl} \sin k\pi x_1 \sin l\pi x_2 = \sum_{k=1}^{N-1} B_k(x_2) \sin k\pi x_1, \tag{2.238}$$

$$V(x_1) = \sum_{k=1}^{N-1} a_k \cos k\pi \left(x_1 + \frac{h}{2}\right). \tag{2.239}$$

Hence we have that

$$D_{x_1}^- V(x_1) = -\sum_{k=1}^{N-1} \sqrt{\lambda_k} a_k \sin k\pi x_1, \quad \text{where } \lambda_k := \frac{4}{h^2} \sin^2 \frac{k\pi h}{2}.$$

Using the orthogonality of sine functions, we deduce that

$$\begin{aligned}\left(D_{x_1}^- V, U\right)_{\Sigma^h} &= -\frac{1}{2} \sum_{k=1}^{N-1} \sqrt{\lambda_k} a_k B_k(x_2) \\ &\leq \left(\frac{1}{2} \sum_{k=1}^{N-1} \sqrt{\lambda_k} a_k^2\right)^{1/2} \left(\frac{1}{2} \sum_{k=1}^{N-1} \sqrt{\lambda_k} B_k^2(x_2)\right)^{1/2}.\end{aligned} \tag{2.240}$$

Let us consider the following sum (over mesh-points):

$$N^2(V) := h^2 \sum_{x_1, t_1 = -1, t_1 \neq 0}^{1-h} \left(\frac{V(x_1) - V(x_1 - t_1)}{t}\right)^2, \tag{2.241}$$

where the mesh-function V has been extended outside Σ_-^h by (2.239). Using the periodicity and orthogonality of cosine functions, we then deduce that

$$N^2(V) = h^2 \sum_{x_1=-1}^{1-h} \sum_{0\neq t_1=-1}^{1-h} \frac{-V(x_1+t_1)+2V(x_1)-V(x_1-t_1)}{t_1^2} V(x_1)$$
$$= 4\sum_{k=1}^{N-1} \sqrt{\lambda_k} a_k^2 I_k,$$

where

$$I_k := \frac{\frac{k\pi h}{2}}{\sin\frac{k\pi h}{2}} J_k, \qquad J_k := \frac{k\pi h}{2} \sum_{t_1=h}^{1-h} \left(\frac{\sin\frac{k\pi t_1}{2}}{\frac{k\pi t_1}{2}} \right)^2.$$

We note that

$$1 \leq \frac{\frac{k\pi h}{2}}{\sin\frac{k\pi h}{2}} \leq \frac{\pi}{2}$$

and that J_k is a Riemann sum for $\int_0^{k\pi/2} \left(\frac{\sin\tau}{\tau}\right)^2 \mathrm{d}\tau$, which therefore satisfies the following two-sided bound:

$$\frac{1}{\pi} \leq J_k \leq \frac{\pi}{2} + \frac{2}{\pi}.$$

Hence,

$$\frac{4}{\pi} \sum_{k=1}^{N-1} \sqrt{\lambda_k} a_k^2 \leq N^2(V) \leq \left(\pi^2+4\right) \sum_{k=1}^{N-1} \sqrt{\lambda_k} a_k^2.$$

From (2.241), using the periodicity of the cosine function, we also have that

$$N^2(V) = h^2 \sum_{x_1, x_1'=-1, x_1\neq x_1'}^{1-h} \left(\frac{V(x_1)-V(x_1')}{x_1-x_1'} \right)^2 \leq 4|V|^2_{W_2^{1/2}(\Sigma^h)},$$

whereby

$$\sum_{k=1}^{N-1} \sqrt{\lambda_k} a_k^2 \leq \pi |V|^2_{W_2^{1/2}(\Sigma^h)}. \tag{2.242}$$

On the other hand, since $B_k(0)=0$, we obtain

$$B_k^2(x_2) = h\sum_{x_2'=0}^{x_2-h} D_{x_2}^+ \left(B_k^2(x_2')\right) = h\sum_{x_2'=0}^{x_2-h} \left(D_{x_2}^+ B_k(x_2')\right)\left(B_k(x_2'+h)+B_k(x_2')\right)$$

$$\leq \varepsilon_k h \sum_{x_2'=h}^{1-h} B_k^2\big(x_2'\big) + \frac{1}{\varepsilon_k} h \sum_{x_2'=0}^{1-h} \big(D_{x_2}^+ B_k\big(x_2'\big)\big)^2,$$

with $\varepsilon_k > 0$, $k = 1, \ldots, N-1$, to be chosen.

Selecting $\varepsilon_k = \sqrt{\lambda_k}$ for $k = 1, \ldots, N-1$, and using the discrete Parseval identities (2.21) and (2.22), we have that

$$\frac{1}{2} \sum_{k=1}^{N-1} \sqrt{\lambda_k} B_k^2(x_2) \leq \big\| D_{x_1}^+ U \big\|_{L_2(\Omega_1^h)}^2 + \big\| D_{x_2}^+ U \big\|_{L_2(\Omega_2^h)}^2 \leq \|U\|_{W_2^1(\Omega^h)}^2. \tag{2.243}$$

Finally, the assertion follows from the inequalities (2.240), (2.242) and (2.243) with $C = \sqrt{\pi/2}$. That completes the proof. □

2.8.2 Convergence in the Discrete W_2^1 Norm

Let u be the solution of the boundary-value problem (2.231) and let U denote the solution of the finite difference scheme (2.237). The global error $e := u - U$ then satisfies the finite difference scheme

$$\mathcal{L}_h e + k\delta_{\Sigma^h} e = \varphi \text{ in } \Omega^h, \quad e = 0 \text{ on } \Gamma^h, \tag{2.244}$$

where

$$\begin{aligned}
\varphi &:= \sum_{i,j=1}^{2} D_{x_i}^- \eta_{ij} + \eta + \delta_{\Sigma^h} \mu, \\
\eta_{ij} &:= T_i^+ T_{3-i}^2 (a_{ij} \partial_j u) - \frac{1}{2}\big(a_{ij} D_{x_j}^+ u + a_{ij}^{+i} D_{x_j}^- u^{+i}\big), \\
\eta &:= \big(T_1^2 T_2^2 a\big) u - T_1^2 T_2^2 (au), \\
\mu &:= ku - T_1^2 (ku).
\end{aligned}$$

Let us decompose η_{1j} and η as follows:

$$\eta_{1j} = \tilde{\eta}_{1j} + \delta_{\Sigma^h} \hat{\eta}_{1j} \quad \text{and} \quad \eta = \tilde{\eta} + \delta_{\Sigma^h} \hat{\eta},$$

where

$$\begin{aligned}
\hat{\eta}_{11} &:= \frac{h^2}{6} T_1^+ \big([a_{11} \partial_1 \partial_2 u + \partial_2 a_{11} \partial_1 u]_\Sigma\big), \\
\hat{\eta}_{12} &:= \frac{h^2}{6} T_1^+ \big(\big[a_{12} \partial_2^2 u + \partial_2 a_{12} \partial_2 u\big]_\Sigma\big) - \frac{h^2}{4} T_1^+ \big(\big[\partial_1 (a_{12} \partial_2 u)\big]_\Sigma\big),
\end{aligned}$$

$$\hat{\eta} := -\frac{h^2}{6}\big[\big(T_1^2 a\big)\big(T_1^2 \partial_2 u\big)\big]_\Sigma.$$

By performing summations by parts and applying Lemma 2.74 we deduce the following bound:

$$\|z\|_{W_2^1(\Omega^h)} \le C\Bigg[\sum_{j=1}^{2}\big(\|\eta_{2j}\|_{L_2(\Omega_2^h)} + \|\tilde{\eta}_{1j}\|_{L_2(\Omega_1^h)} + |\hat{\eta}_{1j}|_{W_2^{1/2}(\Sigma^h)}\big)$$
$$+ \|\tilde{\eta}\|_{L_2(\Omega^h)} + \|\hat{\eta}\|_{L_2(\Sigma^h)} + \|\mu\|_{L_2(\Sigma^h)}\Bigg]. \tag{2.245}$$

Hence, in order to estimate the convergence rate of the finite difference scheme (2.244), it suffices to bound the terms on the right-hand side of (2.245).

The terms η_{2j}, $j = 1, 2$, have been bounded in Sect. 2.6.1. After summation over the mesh Ω_2^h we obtain

$$\|\eta_{2j}\|_{L_2(\Omega_2^h)} \le Ch^{s-1}\big(\|a_{2j}\|_{W_2^{s-1}(\Omega^-)}\|u\|_{W_2^s(\Omega^-)}$$
$$+ \|a_{2j}\|_{W_2^{s-1}(\Omega^+)}\|u\|_{W_2^s(\Omega^+)}\big), \quad 2 < s \le 3. \tag{2.246}$$

The terms $\tilde{\eta}_{1j}$ for $x \in \Omega_1^h \setminus \Sigma_-^h$ can be bounded in the same way. For $x \in \Sigma_-^h$ we set

$$\tilde{\eta}_{11} := \sum_{k=1}^{3}\big(\eta_{11,k}^- + \eta_{11,k}^+\big),$$
$$\tilde{\eta}_{12} := \sum_{k=1}^{4}\big(\eta_{12,k}^- + \eta_{12,k}^+\big),$$

where

$$\eta_{11,1}^\pm := \frac{1}{2}T_1^+T_2^{2\pm}(a_{11}\partial_1 u) - \frac{1}{2}\big(T_1^+T_2^{2\pm}a_{11}\big)\big(T_1^+T_2^{2\pm}\partial_1 u\big)$$
$$\pm\frac{h}{6}\big(T_1^+\partial_2 a_{11}\big)\big[\big(T_1^+T_2^{2\pm}\partial_1 u\big) - \big(T_1^+\partial_1 u\big)\big]$$
$$\pm\frac{h}{6}\bigg[\frac{a_{11}+a_{11}^{+1}}{2}\big(T_1^+\partial_1\partial_2 u\big) - T_1^+(a_{11}\partial_1\partial_2 u)\bigg]$$
$$\pm\frac{h}{6}\big[\big(T_1^+\partial_2 a_{11}\big)\big(T_1^+\partial_1 u\big) - T_1^+(\partial_2 a_{11}\partial_1 u)\big]\big|_{x_2=\xi\pm 0},$$

$$\eta_{11,2}^\pm := \frac{1}{2}\bigg[\big(T_1^+T_2^{2\pm}a_{11}\big) - \frac{a_{11}+a_{11}^{+1}}{2} \mp \frac{h}{3}\big(T_1^+\partial_2 a_{11}\big)\bigg]\big(T_1^+T_2^{2\pm}\partial_1 u\big)\big|_{x_2=\xi\pm 0},$$

$$\eta_{11,3}^{\pm} := \frac{a_{11}+a_{11}^{+1}}{4}\Big[\big(T_1^+T_2^{2\pm}\partial_1 u\big) - u_{x_1} \mp \frac{h}{3}\big(T_1^+\partial_1\partial_2 u\big)\Big]\Big|_{x_2=\xi\pm 0},$$

$$\begin{aligned}
\eta_{12,1}^{\pm} := {}& \frac{1}{2}T_1^+T_2^{2\pm}(a_{12}\partial_2 u) - \frac{1}{2}\big(T_1^+T_2^{2\pm}a_{12}\big)\big(T_1^+T_2^{2\pm}\partial_2 u\big)\\
&\pm \frac{h}{6}\big(T_1^+\partial_2 a_{12}\big)\big[\big(T_1^+T_2^{2\pm}\partial_2 u\big) - \big(T_1^+\partial_2 u\big)\big]\\
&\pm \frac{h}{6}\Big[\frac{a_{12}+a_{12}^{+1}}{2}\big(T_1^+\partial_2^2 u\big) - T_1^+\big(a_{12}\partial_2^2 u\big)\Big]\\
&\pm \frac{h}{6}\big[\big(T_1^+\partial_2 a_{12}\big)\big(T_1^+\partial_2 u\big) - T_1^+(\partial_2 a_{12}\partial_2 u)\big]\\
&\pm \frac{h}{4}T_1^+\big(\partial_1 a_{12}\big(T_2^{\pm}\partial_2 u - \partial_2 u\big)\big)\big|_{x_2=\xi\pm 0},
\end{aligned}$$

$$\eta_{12,2}^{\pm} := \frac{1}{2}\Big[\big(T_1^+T_2^{2\pm}a_{12}\big) - \frac{a_{12}+a_{12}^{+1}}{2} \mp \frac{h}{3}\big(T_1^+\partial_2 a_{12}\big)\Big]\big(T_1^+T_2^{2\pm}\partial_2 u\big)\big|_{x_2=\xi\pm 0},$$

$$\begin{aligned}
\eta_{12,3}^{+} := {}& \frac{a_{12}+a_{12}^{+1}}{4}\Big[\big(T_1^+T_2^{2+}\partial_2 u\big) - \frac{D_{x_2}^+u + D_{x_2}^+u^{+1}}{2} - \frac{h}{3}\big(T_1^+\partial_2^2 u\big)\Big]\\
&+ \frac{h}{4}T_1^+\big(a_{12}\big(T_2^+\partial_1\partial_2 u - \partial_1\partial_2 u\big)\big)\big|_{x_2=\xi+0},
\end{aligned}$$

$$\begin{aligned}
\eta_{12,3}^{-} := {}& \frac{a_{12}+a_{12}^{+1}}{4}\Big[\big(T_1^+T_2^{2-}\partial_2 u\big) - \frac{D_{x_2}^+u + D_{x_2}^+u^{+1}}{2} - \frac{h}{3}\big(T_1^+\partial_2^2 u\big)\Big]\\
&+ \frac{h}{4}T_1^+\big(a_{12}\big(T_2^-\partial_1\partial_2 u - \partial_1\partial_2 u\big)\big)\big|_{x_2=\xi-0},
\end{aligned}$$

$$\eta_{12,4}^{+} := -\frac{1}{8}\big(a_{12}^{+1} - a_{12}\big)\big(D_{x_2}^+u^{+1} - u_{x_2}\big)\big|_{x_2=\xi+0},$$

$$\eta_{12,4}^{-} := -\frac{1}{8}\big(a_{12}^{+1} - a_{12}\big)\big(D_{x_2}^-u^{+1} - u_{\bar{x}_2}\big)\big|_{x_2=\xi-0}.$$

The terms $\eta_{1j,k}^{\pm}$ can be bounded analogously to the corresponding terms $\eta_{1j,k}$ considered in Sect. 2.6.1. Thus we obtain:

$$\begin{aligned}
\|\tilde{\eta}_{1j}\|_{L_2(\Omega_1^h)} \le {}& Ch^{s-1}\big(\|a_{1j}\|_{W_2^{s-1}(\Omega^-)}\|u\|_{W_2^s(\Omega^-)}\\
&+ \|a_{1j}\|_{W_2^{s-1}(\Omega^+)}\|u\|_{W_2^s(\Omega^+)}\big), \quad 2.5 < s \le 3. \qquad (2.247)
\end{aligned}$$

For $x \in \Omega^h \setminus \Sigma^h$, the term $\tilde{\eta}$ can be bounded in the same way as the corresponding term η in Sect. 2.6.1. For $x \in \Sigma^h$ we use the following decomposition:

$$\tilde{\eta} = \eta_{(1)}^{+} + \eta_{(1)}^{-} + \eta_{(2)}^{+} + \eta_{(2)}^{-},$$

where

$$\eta^{\pm}_{(1)} := \frac{1}{2}\big(T_1^2 T_2^{2\pm} a\big)\Big[u - \big(T_1^2 T_2^{2\pm} u\big) \pm \frac{h}{3}\big(T_1^2 \partial_2 u\big)\Big]\Big|_{x_2=\xi\pm 0},$$

$$\eta^{\pm}_{(2)} := \frac{1}{2}\Big[\big(T_1^2 T_2^{2\pm} a\big)\big(T_1^2 T_2^{2\pm} u\big) - T_1^2 T_2^{2\pm}(au) \pm \frac{h}{6}\big(\big(T_1^2 a\big) - \big(T_1^2 T_2^{2\pm} a\big)\big)\big(T_1^2 \partial_2 u\big)\Big]\Big|_{x_2=\xi\pm 0}.$$

These terms can be bounded analogously to the terms η_3 and η_4 discussed in Sect. 2.6.1. Hence we deduce that

$$\|\tilde{\eta}\|_{L_2(\Omega^h)} \le Ch^{s-1}\big(\|a\|_{W_2^{s-2}(\Omega^-)}\|u\|_{W_2^s(\Omega^-)} + \|a\|_{W_2^{s-2}(\Omega^+)}\|u\|_{W_2^s(\Omega^+)}\big), \quad 2 < s \le 3. \tag{2.248}$$

The value of μ at the node $(x_1, \xi) \in \Sigma^h$ is a bounded linear functional of $ku \in W_2^{s-1}(\imath)$, $\imath = (x_1 - h, x_1 + h) \times \{\xi\}$, $s > 3/2$, which vanishes on all linear polynomials. Using the Bramble–Hilbert lemma one then obtains that

$$\|\mu\|_{L_2(\Sigma^h)} \le Ch^{s-1}\|ku\|_{W_2^{s-1}(\Sigma)} \le Ch^{s-1}\|k\|_{W_2^{s-1}(\Sigma)}\big(\|u\|_{W_2^s(\Omega^-)} + \|u\|_{W_2^s(\Omega^+)}\big), \quad 1.5 < s \le 3. \tag{2.249}$$

The term $\hat{\eta}$ can be bounded directly:

$$\|\hat{\eta}\|_{L_2(\Sigma^h)} \le Ch^2\big(\|a\|_{L_2(\Sigma^+)}\|\partial_2 u\|_{C(\overline{\Omega}^+)} + \|a\|_{L_2(\Sigma^-)}\|\partial_2 u\|_{C(\overline{\Omega}^-)}\big) \le Ch^2\big(\|a\|_{W_2^{s-2}(\Omega^+)}\|u\|_{W_2^s(\Omega^+)} + \|a\|_{W_2^{s-2}(\Omega^-)}\|u\|_{W_2^s(\Omega^-)}\big), \quad s > 2.5, \tag{2.250}$$

where we have used the following notation:

$$\|a\|_{L_2(\Sigma^\pm)} := \big\|a(\cdot, \xi \pm 0)\big\|_{L_2(0,1)}.$$

For a function $\varphi \in W_2^\lambda(\Sigma)$, $0 < \lambda \le 1/2$, the seminorm $|T_1^+\varphi|_{W_2^{1/2}(\Sigma^h)}$ can be estimated directly:

$$\big|T_1^+\varphi\big|_{W_2^{1/2}(\Sigma^h)} \le 2^{\lambda+1/2}h^{\lambda-1/2}|\varphi|_{W_2^\lambda(\Sigma)} \le Ch^{\lambda-1/2}\|\varphi\|_{W_2^{\lambda+1/2}(\Omega^\pm)}.$$

We thus deduce that

$$|\hat{\eta}_{11}|_{W_2^{1/2}(\Sigma^h)} \le Ch^{s-1}\big(\|a_{11}\partial_1\partial_2 u\|_{W_2^{s-2}(\Omega^+)} + \|a_{11}\partial_1\partial_2 u\|_{W_2^{s-2}(\Omega^-)}$$
$$+ \|\partial_2 a_{11}\partial_1 u\|_{W_2^{s-2}(\Omega^+)} + \|\partial_2 a_{11}\partial_1 u\|_{W_2^{s-2}(\Omega^-)}\big)$$
$$\le Ch^{s-1}\big(\|a_{11}\|_{W_2^{s-1}(\Omega^+)}\|u\|_{W_2^{s}(\Omega^+)} + \|a_{11}\|_{W_2^{s-1}(\Omega^-)}\|u\|_{W_2^{s}(\Omega^-)}\big), \tag{2.251}$$

for $2.5 < s \le 3$, and analogously

$$|\hat{\eta}_{12}|_{W_2^{1/2}(\Sigma^h)} \le Ch^{s-1}\big(\|a_{12}\|_{W_2^{s-1}(\Omega^+)}\|u\|_{W_2^{s}(\Omega^+)}$$
$$+ \|a_{12}\|_{W_2^{s-1}(\Omega^-)}\|u\|_{W_2^{s}(\Omega^-)}\big), \quad \text{for } 2.5 < s \le 3. \tag{2.252}$$

Hence, from (2.245)–(2.252) we obtain the main result of this section.

Theorem 2.75 *Suppose that the solution of the boundary-value problem* (2.231) *belongs to the function space* $\widetilde{W}_2^s(\Omega)$, *and that the coefficients of the equation* (2.231) *satisfy the following regularity hypotheses*:

$$a_{ij} \in W_2^{s-1}(\Omega^+) \cap W_2^{s-1}(\Omega^-) \cap C(\overline{\Omega}),$$
$$a \in W_2^{s-2}(\Omega^+) \cap W_2^{s-2}(\Omega^-), \quad k \in W_2^{s-1}(\Sigma).$$

Then, *the finite difference scheme* (2.244) *converges and the following error bound holds*:

$$\|u - U\|_{W_2^1(\Omega^h)}$$
$$\le Ch^{s-1}\Big(\max_{i,j}\|a_{ij}\|_{W_2^{s-1}(\Omega^+)} + \max_{i,j}\|a_{ij}\|_{W_2^{s-1}(\Omega^-)}$$
$$+ \|a\|_{W_2^{s-2}(\Omega^+)} + \|a\|_{W_2^{s-2}(\Omega^-)} + \|k\|_{W_2^{s-1}(\Sigma)}\Big)\|u\|_{\widetilde{W}_2^s(\Omega)}, \quad 2.5 < s \le 3,$$

where $C = C(s)$ *is a positive constant*, *independent of* h.

2.9 Bibliographical Notes

The principal purpose of this chapter has been to develop a technique for the derivation of error bounds, which are compatible with the smoothness of the data, for finite difference approximations of boundary-value problems for second- and fourth-order linear elliptic partial differential equations. The technique is based on the Bramble–Hilbert lemma and its generalizations (see Bramble and Hilbert [20, 21], Dupont and Scott [37], Dražić [32], Jovanović [79]).

According to the definition of Lazarov, Makarov and Samarskiĭ [125], an error bound of the form

$$\|u - U\|_{W_2^r(\Omega^h)} \le Ch^{s-r}\|u\|_{W_2^s(\Omega)}, \quad s > r, \tag{2.253}$$

is said to be *compatible with the smoothness of the solution* to the boundary–value problem. Similar error bounds, in 'continuous' norms, of the form

$$\|u - u^h\|_{W_2^r(\Omega)} \le Ch^{s-r} \|u\|_{W_2^s(\Omega)}, \quad 0 \le r \le 1 < s \le p+1,$$

are typical for finite elements methods (see e.g. Strang and Fix [169], Ciarlet [26], Brenner and Scott [23]) and are usually referred to as *optimal*; here u^h denotes the finite element approximation of the analytical solution u using continuous piecewise polynomials of degree p.

In the case of equations with variable coefficients the constant C in the error bound (2.253) depends on norms of the coefficients. One of our main objectives in this chapter has therefore been to understand this dependence in the case of various second-order and fourth-order linear elliptic model problems with variable coefficients. Specifically, we proved error bounds that are of the typical form

$$\|u - U\|_{W_2^r(\Omega^h)} \le Ch^{s-r} \Big(\max_{i,j} \|a_{ij}\|_{W_p^{s-1}(\Omega)} + \|a\|_{W_p^{s-2}(\Omega)} \Big) \|u\|_{W_2^s(\Omega)}.$$

To the best of our knowledge, error bounds of the form (2.253) were first derived by Weinelt [195], for $r = 1$ and $s = 2, 3$, in case of Poisson's equation. Subsequently, bounds of the form (2.253) were obtained by Lazarov, Makarov, Samarskiĭ, Weinelt, Jovanović, Ivanović, Süli, Gavrilyuk, Voĭtsekhovskiĭ, Berikelashvili and others, by systematic use of the Bramble–Hilbert lemma.

For example, families of finite difference schemes for Poisson's equation and the generalized Poisson equation with mollified right-hand sides were introduced by Jovanović [111] and Ivanović, Jovanović and Süli [75, 106], and scales of error bounds of the form (2.253) were established in the case of both integer and fractional values of s.

A procedure for determining the constant in the Bramble–Hilbert lemma, using the mapping of elementary rectangles on a canonical rectangle, was proposed by Lazarov [119]; see also [37] and [38] for related issues.

In the papers of Lazarov [119], Lazarov and Makarov [123] and Makarov and Ryzhenko [130, 131], the convergence of various difference schemes was examined for Poisson's equation in cylindrical, polar and spherical coordinates, and error bounds of the form (2.253) were derived under the assumption that the analytical solutions to these problems belong to appropriate weighted Sobolev spaces. Finite difference approximations of Poisson's equation by special classes of finite volume and finite difference schemes on nonuniform meshes were studied by Süli [171] and Jovanović and Matus [73]. In particular, the results in Sects. 2.4 and 2.4.2 are based on the paper [171]. The analysis presented in Sect. 2.4.3 was stimulated by discussions with Professor Rupert Klein, Free University Berlin. For related work, we refer to the paper of Oevermann and Klein [147].

A finite difference scheme with enhanced accuracy for second-order elliptic equations with constant coefficients was derived by Jovanović, Süli and Ivanović [108], and similar results were obtained later by Voĭtsekhovskiĭ and Novichenko [188].

Difference schemes for the biharmonic equation with a nonsmooth source term were considered by Lazarov [120], Gavrilyuk, Lazarov, Makarov and Pirnazarov [50], Ivanović, Jovanović and Süli [76], and for systems of partial differential equations in linear elasticity theory by Kalinin and Makarov [114, 129] and Voĭtsekhovskiĭ and Kalinin [187].

The convergence of the so-called *exact difference schemes* was investigated by Lazarov, Makarov and Samarskiĭ [125].

The error analysis of finite difference schemes for linear partial differential equations with variable coefficients was developed later. The first attempts in this direction were focused on finite difference schemes for the generalized Poisson equation with a variable coefficient in the lowest-order term (Lazarov, Makarov and Weinelt [126, 196], Voĭtsekhovskiĭ, Makarov and Shabliĭ [189]); subsequently, problems with variable coefficients in the principal part of the partial differential operator were considered (Godev and Lazarov [58], Jovanović, Ivanović and Süli [110], Jovanović [83]). Partial differential equations where the coefficient of the lowest-order term belongs to a negative Sobolev space were considered by Voĭtsekhovskiĭ, Makarov and Rybak [192], and Jovanović [83]. Zlotnik [203, 205] obtained different error estimates for discretizations of elliptic problems with variable coefficients.

Fourth-order equations with variable coefficients were studied by Gavrilyuk, Prikazchikov and Khimich [51], and Jovanović [84]. Quasilinear equations in arbitrary domains, solved by a combination of finite difference and fictitious domain methods, were studied by Voĭtsekhovskiĭ and Gavrilyuk [186], Voĭtsekhovskiĭ, Gavrilyuk and Makarov [191] and Jovanović [80, 81].

The technique described above was also used for the solution of eigenvalue problems (Prikazchikov and Khimich [151]), variational inequalities (Voĭtsekhovskiĭ, Gavrilyuk and Sazhenyuk [190], Gavrilyuk and Sazhenyuk [49]) and in the analysis of supraconvergence on nonuniform meshes (Marletta [134]). Berikeshvili systematically used the same technique for the numerical approximation of a general class of elliptic problems, including equations of higher order, systems of elliptic equations, problems with nonlocal boundary conditions, etc.; for further details, we refer to the survey paper [11], which also contains an extensive bibliography. Berikeshvili, Gupta and Mirianashvili [12] investigated the convergence of fourth-order compact difference schemes for three-dimensional convection-diffusion equations. Jovanović and Vulkov [101] studied the finite difference approximation of elliptic interface problems with variable coefficients.

Recently, a group of mathematicians (Barbeiro, Ferreira, Emmrich, Grigorieff et al.) exploited the techniques discussed in this chapter for the analysis of super- and supraconvergence effects in finite-difference and finite-element schemes (see Barbeiro [5], Barbeiro, Ferreira and Grigorieff [6], Emmrich [44], Emmrich and Grigorieff [45] and Ferreira and Grigorieff [47]).

There has also been work on the convergence analysis of finite difference schemes in discrete W_p^k norms, for $p \neq 2$; see, for example, Lazarov and Mokin [124], Lazarov [121], Godev and Lazarov [57], Drenska [33, 34], Süli, Jovanović and Ivanović [173, 174]. In this case, the derivation of a priori estimates is technically more complex—the theory of discrete Fourier multipliers, developed by

Mokin [140], is used instead of standard discrete energy estimates. Error bounds for the difference schemes under consideration are then obtained by combining these a priori estimates with the use of the Bramble–Hilbert lemma, as we have described in this chapter.

An alternative technique for the derivation of error bounds of the form (2.253) in fractional-order norms is based on function space interpolation, and was used by Jovanović [89].

Our goal in the rest of the book is to extend the methodology developed in the present chapter to time-dependent problems. In Chap. 3 we shall be concerned with parabolic partial differential equations, while in Chap. 4 we focus on hyperbolic equations.

http://www.springer.com/978-1-4471-5459-4

Analysis of Finite Difference Schemes
For Linear Partial Differential Equations with
Generalized Solutions
Jovanović, B.S.; Süli, E.
2014, XIII, 408 p. 7 illus. in color., Hardcover
ISBN: 978-1-4471-5459-4