(b) Let $K \subset \Omega$ be a compact set. Since $\Gamma_n \to \Gamma$ there exists a $N(K)$ such that $K \subset \Omega_n$ for $n \geq N(K)$. Thus, the sequence $\{u_n : n \geq N(K)\}$ converges uniformly in the usual sense on K to u so that one can apply Theorem 2.3.11: consequently u is harmonic in K. Since $K \subset \Omega$ may be chosen arbitrarily, it follows that $u \in C^2(\Omega)$. By assumption, we already have $u \in C^0(\overline{\Omega})$. That the boundary value $u = \varphi$ is taken on Γ is deduced from $\varphi_n \to \varphi$ and $\Gamma_n \to \Gamma$. ■

In Theorem 4 one was able to derive the existence of a solution u of (5b) just from $\varphi_n \to \varphi$. This inference is not possible for the case of $\Omega_n \neq \Omega$ as the following example shows.

Let $\Omega_n := K_1(0)\backslash\overline{K_{1/n}(0)} \subset \Omega := K_1(0)\backslash\{0\} \subset \mathbb{R}^2$. The boundaries are $\Gamma_n = \partial K_1(0) \cup \partial K_{1/n}(0)$, and $\Gamma = \partial K_1(0) \cup \{0\}$, and satisfy $\Gamma_n \to \Gamma$. The boundary values

$$\varphi = \varphi_n = 0 \quad \text{on } \partial K_1(0), \qquad \varphi_n = 1 \quad \text{on } \partial K_{1/n}(0), \qquad \varphi(0,0) = 1$$

satisfy the condition $\varphi_n \to \varphi$ (cf. (4a) and Remark 5a). The solutions u_n of (5a) can be given explicitly:

$$u_n(\mathbf{x}) = \log(|\mathbf{x}|)/\log(1/n).$$

Obviously, $u_n(\mathbf{x}) \to u(\mathbf{x}) := 0$ holds pointwise, but $u = 0$ satisfies neither (4b) nor the boundary value problem (5b). Conversely, one infers from Theorem 6a the following result:

Remark 2.4.7. In $\Omega = K_1(0)\backslash\{0\} \subset \mathbb{R}^2$ the potential equation has no solution $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ which assumes the boundary values $u(\mathbf{x}) = 0$ on $\partial K_1(0)$ and $u(\mathbf{x}) = 1$ in $\mathbf{x} = 0$.

3 The Poisson Equation

3.1 Posing the Problem

The Poisson equation reads

$$\Delta u = f \quad \text{in } \Omega \tag{3.1.1a}$$

with given $f \in C^0(\Omega)$. In the physical interpretation f is the source term [for example, the charge density in the case of an electrical potential u]. To determine the solution uniquely one needs a boundary value specification, for example, the Dirichlet condition

$$u = \varphi \quad \text{on } \Gamma. \tag{3.1.1b}$$

Definition 3.1.1. The function u is called the classical solution of the boundary value problem (1a,b) if $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ satisfies the equations (1a,b) pointwise.

Until we introduce weak solutions in Section 7, "solution" will always mean "classical solution".

The solution of the boundary value problem (1a,b) will in general no longer satisfy the mean value property and the maximum principle. But these properties still hold for the differences of two solutions u_1 and u_2 of the Poisson equation, since $\Delta(u_1 - u_2) = f - f = 0$. Thus the uniqueness of the solution of problem (1a,b) immediately follows and Theorem 2.4.3 can be brought over:

Theorem 3.1.2. *Let Ω be bounded.* (a) *The solution of* (1a,b) *is uniquely determined.* (b) *If u^{I} and u^{II} are solutions of the Poisson equation for boundary values φ^{I} and φ^{II}, then we have*

$$\|u^{\mathrm{I}} - u^{\mathrm{II}}\|_\infty \leq \|\varphi^{\mathrm{I}} - \varphi^{\mathrm{II}}\|_\infty. \tag{3.1.2}$$

PROOF. (b) The proof of Theorem 2.4.3 can be repeated verbatim here. (a) If u^{I} and u^{II} are two solutions of (1a,b) then (2) shows that $\|u^{\mathrm{I}} - u^{\mathrm{II}}\|_\infty \leq \|\varphi - \varphi\|_\infty = 0$. Thus $u^{\mathrm{I}} = u^{\mathrm{II}}$. ■

Theorems 2.4.4 and 2.4.6 can be transferred likewise.

3.2 Representation of the Solution by the Green Function

Lemma 3.2.1. *Let the solution of* (1.1a,b) *belong to* $C^2(\overline{\Omega})$, *where* Ω *is a normal domain. Then* u *may be represented as*

$$u(\mathbf{x}) = -\int_\Omega \gamma(\xi,\mathbf{x})f(\xi)\,d\xi + \int_\Gamma \left[\gamma(\xi,\mathbf{x})\frac{\partial}{\partial n}u(\xi) - u(\xi)\frac{\partial}{\partial n_\xi}\gamma(\xi,\mathbf{x})\right] d\Gamma_\xi \tag{3.2.1}$$

for every fundamental solution γ *in* (2.2.9).

The proof is the same as in Theorem 2.2.2 or in Corollary 2.2.3. The term $\int_{\Omega_\epsilon} \gamma\Delta u\,d\xi$ with $\Omega_\epsilon = \Omega\backslash\overline{K_\epsilon(\mathbf{x})}$ becomes $\int_{\Omega_\epsilon}\gamma f\,d\xi$. Since the singularity of $\gamma(\xi,\mathbf{x})$ is integrable at $\xi=\mathbf{x}$, $\int_{\Omega_\epsilon}\gamma f\,d\xi$ converges to $\int_\Omega \gamma f\,d\xi$ as $\epsilon\to 0$.

Exercise 3.2.2. (a) Let $\Omega\subset\mathbb{R}^n$ be bounded, $\mathbf{x}_0\in\Omega$, $f\in C^0(\overline{\Omega}\backslash\{\mathbf{x}_0\})$, and $|f(\mathbf{x})|\le C|\mathbf{x}-\mathbf{x}_0|^{-s}$ for $s<n$. Show that $\int_\Omega f(\mathbf{x})\,d\mathbf{x}$ exists as an improper integral.
(b) Let $\Omega\subset\mathbb{R}^n$ be bounded and let $\mathbf{x}_0(\xi)\in\Omega$ depend continuously on $\xi\in D$, with D compact. Let $f(\mathbf{x},\xi)$ be continuous in $(\mathbf{x},\xi)\in\overline{\Omega}\times D$ with $\mathbf{x}\ne\mathbf{x}_0(\xi)$ and let $|f(\mathbf{x},\xi)|\le C|\mathbf{x}-\mathbf{x}_0(\xi)|^{-s}$, $s<n$. Show that $F(\xi):=\int_\Omega f(\mathbf{x},\xi)\,d\mathbf{x}$ is continuous: $F\in C^0(D)$.

In the boundary integral in (1) one may replace $u(\xi)$ by $\varphi(\xi)$ (cf. (1.1b)). The function $\partial u/\partial n$ on Γ, however, is unknown and cannot be specified arbitrarily either, since the boundary values (1.1b) already determine the solution uniquely (cf. Theorem 3.1.2). To make $\int_\Gamma \gamma\partial u/\partial n\,d\xi$ vanish one must select the fundamental solution so that $\gamma(\xi,\mathbf{x})=0$ for $\xi\in\Gamma$, $\mathbf{x}\in\Omega$.

Definition 3.2.3. A fundamental solution g in (2.2.9) is called a Green function (of the first kind) if $g(\xi,\mathbf{x})=0$ for all $\xi\in\Gamma$, $\mathbf{x}\in\Omega$.

The existence of a Green function is closely related to the solvability of the boundary value problem for the potential equation:

Remark 3.2.4. The Green function exists if and only if for all $\mathbf{x}\in\Omega$ the boundary value problem $\Delta\Phi=0$ in Ω and $\Phi=-s(\cdot,\mathbf{x})$ on Γ has a solution $\Phi\in C^2(\overline{\Omega})$.

The above consideration results in

Theorem 3.2.5. *Let* Ω *be a normal domain. Let the boundary value problem* (1.1a,b) *have a solution* $u\in C^2(\overline{\Omega})$. *Assume the existence of a Green function of the first kind. Then one can express* u *explicitly by*

$$u(\mathbf{x}) = -\int_\Omega g(\xi,\mathbf{x})f(\xi)\,d(\xi) - \int_\Gamma \varphi(\xi)\frac{\partial}{\partial n_\xi}g(\xi,\mathbf{x})\,d\Gamma_\xi. \tag{3.2.2}$$

In the following we reverse the implication. Let the existence of the Green function be assumed. Then, does function u defined by Equation (2) represent the classical solution of the boundary value problem (1.1a,b)? Here it must be proved, in particular, that $u\in C^2(\Omega)$ and $\Delta u=f$. Firstly, it is not even clear yet whether the function $u(\mathbf{x})$ defined by Equation (2) depends continuously on $\mathbf{x}$ since the definition of a fundamental solution $\gamma(\xi,\mathbf{x})$ does not require continuity with respect to the second argument $\mathbf{x}$. Despite that, the Green function is $g(\xi,\mathbf{x})$ with respect to $\mathbf{x}$ in $C^2(\Omega\backslash\{\xi\})$, as the following result shows (cf. Leis [1, p. 67]).

Exercise 3.2.6. Let Ω be a normal domain. Let the Green function exist for Ω, and for fixed $\mathbf{y}\in\Omega$ let $g(\cdot,\mathbf{y})\in C^2(\overline{\Omega}\backslash\{\mathbf{y}\})$ (weaker conditions are possible !). Now prove that

$$g(\mathbf{x},\mathbf{y}) = g(\mathbf{y},\mathbf{x}) \quad \text{for } \mathbf{x},\mathbf{y}\in\Omega. \tag{3.2.3}$$

Hint: Apply the Green formula (2.2.5b) with $\Omega_\epsilon = \Omega\backslash[K_\epsilon(\mathbf{x}')\cup K_\epsilon(\mathbf{x}'')]$, $\mathbf{x}',\mathbf{x}''\in\Omega$, $u(\mathbf{x}):=g(\mathbf{x},\mathbf{x}')$, $v(\mathbf{x}):=g(\mathbf{x},\mathbf{x}'')$, and use (2.2.10).

If one tries to reverse the assertion of Theorem 5, one encounters the surprising difficulty of having to set precise conditions on the source term f. The natural requirement $f\in C^0(\overline{\Omega})$ is necessary for $u\in C^2(\overline{\Omega})$, but it is not sufficient, as the following theorem, whose proof will be appended at the end of this section, shows.

Theorem 3.2.7. *Even if the boundary* Γ *and the boundary values* φ *are sufficiently smooth and if the Green function exists, there are functions* $f\in C^0(\overline{\Omega})$ *to which no solutions* $u\in C^2(\overline{\Omega})$ *correspond.*

Theorem 7 shows that Equation (2) need not represent a classical solution for $f\in C^0(\overline{\Omega})$. However, a sufficient condition for f to do so is Hölder continuity.

Definition 3.2.8. $f\in C^0(\overline{\Omega})$ is said to be Hölder continuous in $\overline{\Omega}$ with the exponent $\lambda\in(0,1)$ if there exists a constant $C=C(f)$ such that

$$|f(\mathbf{x})-f(\mathbf{y})|\le C|\mathbf{x}-\mathbf{y}|^\lambda \quad \text{for all } \mathbf{x},\mathbf{y}\in\overline{\Omega}. \tag{3.2.4a}$$

We write $f\in C^\lambda(\overline{\Omega})$ and define the norm $\|f\|_{C^\lambda(\overline{\Omega})}$ as the smallest constant C which satisfies (4a):

$$\|f\|_{C^\lambda(\overline{\Omega})} := \sup\left\{\frac{|f(\mathbf{x})-f(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^\lambda} : \mathbf{x},\mathbf{y}\in\overline{\Omega}, \mathbf{x}\ne\mathbf{y}\right\}. \tag{3.2.4b}$$

The function $f \in C^k(\overline{\Omega})$ is said to be k-fold Hölder continuously differentiable in $\overline{\Omega}$ (with the exponent λ), if $D^\nu f \in C^\lambda(\overline{\Omega})$ for all $|\nu| \le k$. Here

$$\nu = (\nu_1, \dots, \nu_n) \quad \text{with } \nu_i \in \mathbb{Z}, \quad \nu_i \ge 0, \quad |\nu| = \nu_1 + \cdots + \nu_n \tag{3.2.5a}$$

is a multi-index and

$$D^\nu = \frac{\partial^{|\nu|}}{\partial x_1^{\nu_1} \partial x_2^{\nu_2} \cdots \partial x_n^{\nu_n}} \tag{3.2.5b}$$

a $|\nu|$-fold partial derivative operator. The k-fold Hölder continuously differentiable functions form the linear space $C^{k+\lambda}(\overline{\Omega})$ with the norm

$$\|f\|_{C^{k+\lambda}(\overline{\Omega})} := \max\{\|D^\nu f\|_{C^\lambda(\overline{\Omega})} : |\nu| \le k\}. \tag{3.2.4c}$$

If $s = k + \lambda$ one also writes $C^s(\overline{\Omega})$ for $C^{k+\lambda}(\overline{\Omega})$. The k-fold Lipschitz continuously differentiable functions $f \in C^{k,1}(\overline{\Omega})$ are the result of the choice $\lambda = 1$ in (4a,b). For reasons of completeness let us add that

$$\|f\|_{C^k(\overline{\Omega})} = \max\{\|D^\nu f\|_\infty : |\nu| \le k\} \tag{3.2.4d}$$

is the norm in $C^k(\overline{\Omega})$ for integer $k \ge 0$.

Exercise 3.2.9. (a) f is said to be locally Hölder continuous in Ω if for each $\mathbf{x} \in \Omega$ there exists a neighborhood $K_\epsilon(\mathbf{x})$ such that $f \in C^\lambda(K_\epsilon(\mathbf{x}) \cap \Omega)$. Prove that if $\overline{\Omega}$ is compact then $f \in C^\lambda(\overline{\Omega})$ follows from the local Hölder continuity in $\overline{\Omega}$. Formulate and prove corresponding statements for $C^{k+\lambda}(\overline{\Omega})$ and $C^{k,1}(\overline{\Omega})$.
(b) Let $s > 0$. Show that $|\mathbf{x}|^s \in C^s(\overline{K_R(\mathbf{0})})$, if $s \notin \mathbb{N}$, otherwise $|\mathbf{x}|^s \in C^{s-1,1}(\overline{K_R(\mathbf{0})})$. Hint: $1 - t^s \le (1-t)^s$ for $0 \le t \le 1$, $s \ge 0$.

The function u from Equation (2) can be decomposed into $u_1 + u_2$ where $u_1 = -\int_\Omega g f \, d\xi$ and $u_2 = -\int_\Gamma \varphi \partial g / \partial n \, d\Gamma$; u is the solution of the boundary value problem (1.1a,b) if we are able to show that u_1 and u_2 are solutions of

$$\Delta u_1 = f \text{ in } \Omega, \quad u_1 = 0 \text{ on } \Gamma, \quad \Delta u_2 = 0 \text{ in } \Omega, \quad u_2 = \varphi \text{ on } \Gamma.$$

Theorem 3.2.10. *If the Green function exists and satisfies suitable conditions then*

$$u(\mathbf{x}) = -\int_\Gamma \varphi(\xi) \frac{\partial}{\partial n_\xi} g(\xi, \mathbf{x}) \, d\Gamma_\xi \tag{3.2.6}$$

is a classical solution of $\Delta u = 0$ in Ω, with $u = \varphi$ on Γ.

The proof goes in principle just as for Theorem 2.3.9 (cf. Leis [1, p.69]).

Theorem 3.2.11. *Suppose the Green function $g(\cdot, \mathbf{x}) \in C^2(\overline{\Omega} \backslash \{\mathbf{x}\})$ for $\mathbf{x} \in \Omega$ exists, and let it be $f \in C^\lambda(\overline{\Omega})$. Then*

$$u(\mathbf{x}) = -\int_\Omega f(\xi) g(\xi, \mathbf{x}) \, d\xi \tag{3.2.7}$$

is a classical solution of $\Delta u = f$ in Ω, with $u = 0$ on Γ.

PROOF. The boundary condition $u(\mathbf{x}) = 0$ for $\mathbf{x} \in \Gamma$ follows easily from $g(\mathbf{x}, \xi) = 0$ and (3). The property $u \in C^1(\overline{\Omega})$ and the representation $u_{x_i}(\mathbf{x}) = -\int_\Omega f(\xi) g_{x_i}(\xi, \mathbf{x}) \, d\xi$ result from the

Exercise 3.2.12. Let $\Omega \subset \mathbb{R}^n$ be bounded and $A := \{(\xi, \mathbf{x}) \in \overline{\Omega} \times \overline{\Omega} : \xi \ne \mathbf{x}\}$. For the derivatives of f with respect to $\mathbf{x}$ assume $D_\mathbf{x}^\nu f \in C^0(A)$ and $|D_\mathbf{x}^\nu f(\xi, \mathbf{x})| \le C|\mathbf{x} - \xi|^{-s}$ with $s < n$ for any $|\nu| \le k$. Prove that then $F(\mathbf{x}) := \int_\Omega f(\xi, \mathbf{x}) \, d\xi \in C^k(\overline{\Omega})$ and $D^\nu F(\mathbf{x}) = \int_\Omega D_\mathbf{x}^\nu f(\xi, \mathbf{x}) \, d\xi$, $|\nu| \le k$.

To prove $u \in C^2(\overline{\Omega})$ this step cannot be repeated since $g_{x_i x_j}(\xi, \mathbf{x}) = O(|\xi - \mathbf{x}|^{-n})$ has a singularity which is not integrable. We write the derivative in the form

$$u_{x_i}(\mathbf{x}) = -\int_\Omega [f(\xi) - f(\mathbf{x})] g_{x_i}(\xi, \mathbf{x}) \, d\xi - f(\mathbf{x}) \int_\Omega g_{x_i}(\xi, \mathbf{x}) \, d\xi. \tag{3.2.8a}$$

Let $\partial_j F(\mathbf{x})$ be the difference quotient $(F(\mathbf{x}^\epsilon) - F(\mathbf{x}))/\epsilon$, with $x_j^\epsilon = x_j + \epsilon$ and $x_i^\epsilon = x_i$ for $i \ne j$. The product rule $\partial_j(FG)(\mathbf{x}) = G(\mathbf{x}) \partial_j F(\mathbf{x}) + F(\mathbf{x}^\epsilon) \partial_j G(\mathbf{x})$ applied to Equation (8a) gives

$$\partial_j u_{x_i}(\mathbf{x}) = -\int_\Omega [f(\xi) - f(\mathbf{x}^\epsilon)] \partial_j g_{x_i}(\xi, \mathbf{x}) \, d\xi - f(\mathbf{x}^\epsilon) \partial_j \frac{\partial}{\partial x_i} \int_\Omega g(\xi, \mathbf{x}) \, d\xi.$$

Since $[f(\xi) - f(\mathbf{x})] g_{x_i x_j}(\xi, \mathbf{x}) = O(|\xi - \mathbf{x}|^{\lambda - n})$ is integrable, the limit $\epsilon \to 0$ results in the formula

$$u_{x_i x_j}(\mathbf{x}) = -\int_\Omega [f(\xi) - f(\mathbf{x})] g_{x_i x_j}(\xi, \mathbf{x}) \, d\xi - f(\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j} \int_\Omega g(\xi, \mathbf{x}) \, d\xi. \tag{3.2.8b}$$

Equation (8b) implies

$$\Delta u = \int_\Omega [f(\xi) - f(\mathbf{x})] \Delta g \, d\xi - f(\mathbf{x}) \Delta \int_\Omega g \, d\xi = -f(\mathbf{x}) \Delta \int_\Omega g(\xi, \mathbf{x}) \, d\xi,$$

so that it remains only to show that $\Delta \int_\Omega g \, d\xi = -1$. Choose $K_R(\mathbf{z})$ so that $\mathbf{x} \in K_R(\mathbf{z}) \subset \Omega$. The Green function has the form (2.2.9): $g = s + \Phi$. The first two terms in

$$\int_\Omega g(\xi, \mathbf{x}) \, d\xi = \int_{\Omega \backslash K_R(\mathbf{z})} g(\xi, \mathbf{x}) \, d\xi + \int_{K_R(\mathbf{z})} \Phi(\xi, \mathbf{x}) \, d\xi + \int_{K_R(\mathbf{z})} s(\xi, \mathbf{x}) \, d\xi$$

are harmonic in $K_R(\mathbf{z})$, so that $\Delta \int_{K_R(\mathbf{z})} s(\xi, \mathbf{x}) \, d\xi = -1$ is what has to be proved.

Let $\sigma(r)$ be defined by $s(\xi, \mathbf{x}) = \sigma(|\xi - \mathbf{x}|)$ (cf. (2.2.1)). For fixed $r > 0$ set

$$v(\mathbf{x}) := \frac{1}{\omega_n r^{n-1}} \int_{\partial K_r(\mathbf{z})} s(\xi, \mathbf{z})\, d\Gamma_\xi. \tag{3.2.8c}$$

For all $\mathbf{x} \notin \partial K_R(\mathbf{z})$ (i.e., $|\mathbf{x}-\mathbf{z}| \neq r$) v is harmonic, since $s(\xi, \mathbf{x})$ is nonsingular on $\partial K_r(\mathbf{z})$ and satisfies $\Delta_{\mathbf{x}} s(\xi, \mathbf{x}) = 0$. Since $s(\cdot, \mathbf{x})$ is harmonic in $K_r(\mathbf{z})$ for $r < |\mathbf{z} - \mathbf{x}|$, the mean-value property (2.3.1) holds, which can now be written

$$v(\mathbf{x}) = s(\mathbf{z}, \mathbf{x}) = \sigma(|\mathbf{z} - \mathbf{x}|) \quad \text{for} \quad |\mathbf{z} - \mathbf{x}| > r. \tag{3.2.8d}$$

Using Exercise 3.2.2b we see that $v(\mathbf{x})$ is continuous in $\mathbb{R}^n$, so that we also have

$$v(\mathbf{x}) = \sigma(r) \quad \text{for} \quad |\mathbf{z} - \mathbf{x}| = r. \tag{3.2.8e}$$

Thus v is harmonic in $K_r(\mathbf{x})$ with the constant boundary values (8e). The unique solution is therefore

$$v(\mathbf{x}) = \sigma(r) \quad \text{for} \quad |\mathbf{z} - \mathbf{x}| \le r. \tag{3.2.8f}$$

The equations (8c,d,f) yield

$$\int_{\partial K_r(\mathbf{z})} s(\xi, \mathbf{x})\, d\Gamma_\xi = \omega_n r^{n-1} \sigma(\max\{r, |\mathbf{z} - \mathbf{x}|\})$$

and then, since $0 < |\mathbf{z} - \mathbf{x}| < R$,

$$\int_{K_R(\mathbf{z})} s(\xi, \mathbf{x})\, d\xi = \int_0^R \int_{\partial K_r(\mathbf{z})} s(\xi, \mathbf{x})\, d\Gamma_\xi\, dr$$

$$\omega_n \int_0^{|\mathbf{z}-\mathbf{x}|} r^{n-1} \sigma(|\mathbf{z} - \mathbf{x}|)\, dr + \omega_n \int_{|\mathbf{z}-\mathbf{x}|}^R r^{n-1} \sigma(r)\, dr$$

$$= \omega_n \frac{|\mathbf{z} - \mathbf{x}|^n}{n} \sigma(|\mathbf{z} - \mathbf{x}|) + \omega_n \frac{r^n}{n} \sigma(r) \Big|_{|\mathbf{z}-\mathbf{x}|}^R - \omega_n \int_{|\mathbf{z}-\mathbf{x}|}^R \frac{r^n}{n} \sigma'(r)\, dr$$

$$\frac{\omega_n}{n} R^n \sigma(R) - \omega_n \int_{|\mathbf{z}-\mathbf{x}|}^R \frac{r^n}{n} \left(-\frac{r^{1-n}}{\omega_n} \right) dr = \frac{\omega_n}{n} R^n \sigma(R) + \int_{|\mathbf{z}-\mathbf{x}|}^R \frac{r}{n}\, dr$$

$$= \frac{\omega_n}{n} R^n \sigma(R) + \frac{R^2}{2n} - \frac{|\mathbf{z} - \mathbf{x}|^2}{2n}.$$

From this we see that, independently of R, n, and $\mathbf{z}$, there results

$$\Delta \int_{K_R(\mathbf{z})} s(\xi, \mathbf{x})\, d\xi = -1.$$

■

From Theorems 10 and 11 follows

Theorem 3.2.13. *Under the same assumptions as for Theorems* 10 *and* 11 *Equation* (2) *gives a representation for the classical solution of the boundary value problem* (1.1a,b).

Finally we put forward two inequalities for the Green function as exercises:

Exercise 3.2.14. In Ω, and $\Omega_1 \subset \Omega_2$, respectively, let the Green functions g, g_1 and g_2 exist. Show

(a) $\quad 0 \le g(\mathbf{x}, \mathbf{y}) \le s(\mathbf{x}, \mathbf{y}) \quad$ for $\mathbf{x}, \mathbf{y} \in \Omega \subset \mathbb{R}^n, \quad n \ge 3 \qquad$ (3.2.9)

What is the inequality for n=2?

(b) $\quad g_1(\mathbf{x}, \mathbf{y}) \le g_2(\mathbf{x}, \mathbf{y}) \quad$ for $\mathbf{x}, \mathbf{y} \in \Omega_1 \subset \Omega_2 \qquad$ (3.2.10)

Hint: Exercise 2.3.14.

Supplement: Proof of Theorem 7. If we make use of a later theorem (Theorem 6.1.13) then Theorem 7 follows from

Theorem 3.2.15. *The solution u does not depend continuously on f, if the supremum norm* (2.4.1) *is used as the norm in $C^0(\overline{\Omega})$, and that from* (4d) *is used as the norm in $C^2(\overline{\Omega})$.*

PROOF. Let $\Omega = K_1(0) \subset \mathbb{R}^2$ und $\varphi = 0$. The disk Ω is a normal region for which the Green function is known (cf. Theorem 3.3.1). By Theorem 11 there exist solutions $u^n \in C^2(\overline{\Omega})$ of $\Delta u^n = f_n$ in Ω, $u^n = 0$ on Γ for the functions

$$f_n(\mathbf{x}) = \frac{x_2^2 - x_1^2}{r^2} \rho_n(r), \;\; r := |\mathbf{x}|, \quad \rho_n(r) := \min\left(n \cdot r, \left|\log \frac{r}{2}\right|^{-1}\right),$$

which belong to $C^0(\overline{\Omega})$ and are uniformly bounded: $\|f_n\|_\infty = 1/\log 2$. By Theorem 11 we have $u^n(\mathbf{x}) = -\int_\Omega g(\xi, \mathbf{x}) f_n(\xi)\, d\xi$. Since $|f_n(\xi)| \le n|\xi|$, it follows from Exercise 12 that

$$u^n_{x_1 x_1} = -\int_\Omega g_{x_1 x_1} f_n\, d\xi = -\int_\Omega \Phi_{x_1 x_1} f_n\, d\xi - \int_\Omega s_{x_1 x_1} f_n\, d\xi \text{ at } \mathbf{x} = 0,$$

where $g = \Phi + s$. The first integral is bounded since $\Phi \in C^2(\overline{\Omega})$. The derivative of the singularity function is $s_{x_1 x_1}(\xi, 0) = (\xi_1^2 - \xi_2^2)/|\xi|^4$. The special choice of f_n gives for $\mathbf{x} = \mathbf{0}$

$$u^n_{x_1 x_1}(\mathbf{0}) = -\int_\Omega \Phi_{x_1 x_1}(\xi, \mathbf{0}) f_n(\xi)\, d\xi + \int_\Omega [\xi_1^2 - \xi_2^2]^2 |\xi|^{-6} \rho_n(|\xi|)\, d\xi.$$

The surface integral $K := \int_{\partial K_r(0)} [\xi_1^2 - \xi_2^2]^2 |\xi|^{-5}\, d\xi > 0$ does not depend on $r \in (0, 1]$, so that the second integral takes on the form $I_n := K \int_0^1 r^{-1} \rho_n(r)\, dr$. Since $\int_\epsilon^1 [r|\log(r/2)|]^{-1}\, dr$ diverges as $\epsilon \to 0$, we deduce $I_n \to \infty$ as $n \to \infty$. Since $\|u^n\|_{C^2(\overline{\Omega})} \ge |u^n_{x_1 x_1}(\mathbf{0})|$, it follows that the map $f \mapsto u$ is not bounded, and thus not continuous. ■

3.3 The Green Function for the Ball

Theorem 3.3.1. *The Green function for the ball $K_R(\mathbf{y})$ is given by the function in* (2.2.11a). *For $f \in C^\lambda(\overline{\Omega})$ and $\varphi \in C^{2+\lambda}(\Gamma)$ with $0 < \lambda < 1$, the representation formula* (2.2) *defines a solution $u \in C^{2+\lambda}(\overline{\Omega})$ of the boundary value problem $\Delta u = f$ in Ω, $u = \varphi$ on Γ.*

The proof of the theorem follows from a result of Schauder, which is cited in Theorem 9.1.20.

In the case $n = 2$ the plane $\mathbb{R}^2$ can be identified with $\mathbb{C}$ by the correspondence $(x, y) \leftrightarrow z = x + iy$. The following considerations are based on

Exercise 3.3.2. Let the map $\varphi: z = x + iy \in \Omega \mapsto \zeta = \xi + i\eta = \Phi(z) \in \Omega'$ be holomorphic. Show

$$\Delta_z u(\Phi(z)) = |\Phi'|^2 \Delta_\zeta u(\zeta), \qquad \Phi' = \xi_x + i\eta_x \tag{3.3.1}$$

for $u \in C^2(\Omega')$.

Equation (1) shows, in particular, that a holomorphic transformation of coordinates maps harmonic functions into harmonic functions. An arbitrary simply connected region with at least two boundary points can, by the Riemann mapping theorem, be mapped by a conformal mapping $\Phi_{z_0}: z \in \Omega \mapsto \Phi_{z_0}(z) \in K_1(0)$ onto the unit disk such that $\Phi_{z_0}(z_0) = 0$ for any given $z_0 \in \Omega$. Let $g(\zeta, \zeta')$ be the Green function for $K_1(0)$. One may check that $G(z, z_0) := g(\Phi_{z_0}(z), 0)$ is again a fundamental solution. Now $z \in \partial\Omega$ implies $\Phi_{z_0}(z) \in \partial K_1(0)$, i.e., $G(z, z_0) = 0$. Thus $G(z, z_0)$ is the Green function in Ω. This proves

Theorem 3.3.3. *Let $\Omega \subset \mathbb{R}^2$ be simply connected with at least two boundary points. Then there exists a Green function of the first kind for Ω.*

The explicit forms of various Green functions can be found, for example, in the book by Wloka [1, Exercises 21.1–21.8]. Of numerical interest might be the fact that with conformal mapping one may remove corners which are disturbing (e.g., reentrant corners) (cf. Gladwell-Wait [1, p.70]).

Example 3.3.4. Let Ω be the L-shaped region in Example 2.1.4. Choose $\Phi = z^{2/3}: \Omega \to \Omega'$. Then Φ is conformal in Ω. The sides of the angle $\Gamma_0 \subset \partial\Omega$ (cf. Fig. 2.1.1) are mapped into a single line segment, so that Ω' has no more corners sticking in. The Poisson equation $\Delta u = f$ in Ω corresponds to the equation $\Delta v(\zeta) = \frac{9}{4}|\zeta| f(\zeta)$ in Ω'.

3.4 The Neumann Boundary Value Problem

In (1.1b) and in (2.1.1b) the boundary values $u = \varphi$ were given on Γ. These so-called Dirichlet conditions or "Boundary Conditions of the First Kind" are not the only possibility. An alternative is the Neumann Condition

$$\frac{\partial}{\partial n} u(\mathbf{x}) = \varphi(\mathbf{x}) \qquad \text{on } \Gamma. \tag{3.4.1}$$

In physics this second boundary condition, as it is also called, occurs more frequently than the Dirichlet condition. For example, if u is the velocity potential of a gas, then $\frac{\partial u}{\partial n} = 0$ means that the gas can only move tangentially at the boundary Γ. Except in some unusual cases, the boundary value problem $Lu = f$ in Ω and $\partial u / \partial n = \varphi$ on Γ has a unique solution. An exceptional case does however occur for $L = \Delta$:

Theorem 3.4.1. *Let Ω be a normal region. The Poisson equation $\Delta u = f$ with the Neumann boundary condition* (1) *is only solvable if*

$$\int_\Gamma \varphi(\xi)\, d\Gamma_\xi = \int_\Omega f(\mathbf{x})\, d\mathbf{x}. \tag{3.4.2}$$

If a solution u does exist, then $u + c$, with c any constant, is also a solution.

PROOF. (1) One may repeat the proof of Lemma 2.3.6 for $\Delta u = f$.
(2) Obviously $u + c$ satisfies the same equation. ∎

Later, in Example 7.4.8, we will show that the Neumann boundary value problem for the Poisson equation has a solution if and only if (2) is satisfied, and that two solutions can differ only by a constant.

In the representation (2.1) both the values $u(\xi)$, for $\xi \in \Gamma$, and the normal derivative $\partial u / \partial n$ occur. The Green function of the first kind was chosen in such a way that $g(\xi, \mathbf{x}) = 0$ for $\xi \in \Gamma$. In the case of the second boundary conditions (1) one makes the assumption that $\partial\gamma(\xi, \mathbf{x})/\partial n_\xi = c$ (c: constant), i.e.

$$\partial\Phi(\xi, \mathbf{x})/\partial n_\xi = c - \partial s(\xi, \mathbf{x})/\partial n_\xi$$

for $\Phi = \gamma - s$. The Corollary 2.2.3 with $u \equiv 1$ and $\gamma = s$ gives

$$\int_\Gamma \partial\Phi(\xi, \mathbf{x})/\partial n_\xi\, d\Gamma_\xi = cL + 1, \qquad L := \int_\Gamma d\Gamma.$$

Since Φ must be harmonic (i.e., $f := \Delta\Phi = 0$), from Equation (2) we see that $cL + 1 = 0$ is a necessary condition for the existence of Φ. Thus the condition on the Green Function of the Second Kind for the potential equation is

$$\partial\gamma(\mathbf{x}, \xi)/\partial n_\mathbf{x} = -1 \Big/ \int_\Gamma d\Gamma.$$

Thus the term $\int_\Gamma u\partial\gamma/\partial n\, d\Gamma$ in (2.1) becomes const $\cdot \int_\Gamma u\, d\Gamma$. Since u is only determined up to a constant (cf. Theorem 1), one can fix this constant with the additional condition $\int_\Gamma u\, d\Gamma = 0$. This gives the following result, if we write g for γ:

$$u(\mathbf{x}) = -\int_\Omega f(\xi)g(\xi,\mathbf{x})\,d\xi + \int_\Gamma \varphi(\xi)g(\xi,\mathbf{x})\,d\Gamma_\xi.$$

The Green function of the second kind for the ball $K_R(\mathbf{0}) \subset \mathbb{R}^3$ can be found in Leis [1, p. 79].

3.5 The Integral Equation Method

In the representation (2.1) of the Poisson solution the singularity function s can, in particular, be chosen to be γ. If in addition one imposes the given Neumann data (4.1), one obtains

$$u(x) = \int_\Gamma k(\mathbf{x},\xi)u(\xi)\,d\Gamma_\xi + g(\mathbf{x}) \quad \text{for} \quad \mathbf{x} \in \Omega, \tag{3.5.1}$$

with the kernel function $k(\mathbf{x},\cdot) := -\partial s(\xi,x)/n_\xi$ and the functions

$$g(\mathbf{x}) := g_1(\mathbf{x}) + g_2(\mathbf{x}),$$
$$g_1(\mathbf{x}) := \int_\Gamma s(\xi,x)\varphi(\xi)\,d\Gamma_\xi$$
$$g_2(\mathbf{x}) := \int_\Omega s(\xi,x)f(\xi)\,d\Gamma_\xi.$$

The right-hand side in Equation (1) with the unknown boundary value $u(\xi)$, $\xi \in \Gamma$, can be used as an ansatz solution:

$$\Phi(\mathbf{x}) = \int_\Gamma k(\mathbf{x},\xi)u(\xi)\,d\Gamma_\xi + g(\mathbf{x}). \tag{3.5.2}$$

The first summand on the right of (2) is called the double-layer potential (dipole potential); g_1 is the single-layer potential, while g_2 is a volume potential.

For each $u \in C^0(\Gamma)$ the Φ in (2) is a solution of the Poisson equation (1.1a) in Ω. However, Φ is also defined for an argument $\mathbf{x}$ in the exterior domain $\mathbb{R}^n\backslash\Omega$. A closer look at the kernel function $k(\mathbf{x},\xi)$ shows that it is in fact only weakly singular for the case of smooth boundaries Γ. Thus Φ is also defined for $\mathbf{x} \in \Gamma$. The function Φ which is now defined on all $\mathbb{R}^n$ is not continuous at points of the boundary Γ. At $\mathbf{x}_0 \in \Gamma$ there exists both an interior limit $\Phi_-(\mathbf{x}_0)$ for $\mathbf{x} \to \mathbf{x}_0, \mathbf{x} \in \Omega$ and an exterior limit $\Phi_+(\mathbf{x}_0)$ for $\mathbf{x} \to \mathbf{x}_0, \mathbf{x} \in \mathbb{R}^n\backslash\overline{\Omega}$. In addition we have the third function value $\Phi(\mathbf{x}_0)$ of (2). Their connection is given by the following jump discontinuity relation (cf. Hackbusch [7, Satz 8.2.8]):

$$\Phi_+(\mathbf{x}_0) - \Phi_-(\mathbf{x}_0) = 2u(\mathbf{x}_0), \tag{3.5.3a}$$
$$\Phi_+(\mathbf{x}_0) + \Phi_-(\mathbf{x}_0) = 2\Phi(\mathbf{x}_0), \quad \text{for} \quad x_0 \in \Gamma. \tag{3.5.3b}$$

In order that the ansatz (2) does indeed give the solution u in (1), the boundary value Φ_-, continued from the interior, must agree with the function u which is put in the integral: $\Phi_- = u$. Now one can solve Equation (3a,b) for Φ_-: $\Phi_-(\mathbf{x}_0) = \Phi(\mathbf{x}_0) - u(\mathbf{x}_0)$. The equation $\Phi_- = u$ thus leads to

$$u(\mathbf{x}_0) = \frac{1}{2}\Phi(\mathbf{x}_0) = \int_\Gamma k(\mathbf{x}_0,\xi)u(\xi)\,d\Gamma_\xi + g(\mathbf{x}_0) \text{ for } \mathbf{x}_0 \in \Gamma. \tag{3.5.4}$$

Equation (4) is called a Fredholm integral equation of the second kind for the unknown function $u \in C^0(\Gamma)$. The original Neumann boundary-value problem (1.1a), (4.1) and the integral eqation (4) are equivalent in the following sense: (a) If u is the solution of the Neumann boundary-value problem, then the boundary values, $u(\xi)$, $\xi \in \Gamma$, satisfy the integral equation (4). (b) If $u \in C^0(\Gamma)$ is a solution of the integral equation (4), then the expression (1) gives a solution of (1.1a), (4.1) in the entire domain Ω.

The transformation of a boundary-value problem into an integral equation, and the subsequent solution of the integral equation is referred to as the integral equation method. It allows, for example, a new approach to existence statements, in that one shows the solvability of (4). The integral equation (4) can also be attacked numerically. If methods similar to the finite-element method described in Section 8 are used, then the result is called the boundary-element method (BEM).

One can find references to the integral equation method in, e.g., Hackbusch [7, §§7–9] and Kress[1].

4 Difference Methods for the Poisson Equation

4.1 Introduction: The One-Dimensional Case

Before developing difference methods for the partial differential Poisson equation, let us first recall the discretisation of ordinary differential equations. The equation $a(x)u''(x) + b(x)u'(x) + c(x)u(x) = f(x)$ can be supplemented with initial conditions $u(x_1) = u_1, u'(x_1) = u_1'$ or with boundary conditions $u(x_1) = u_1, u(x_2) = u_2$. The ordinary initial value problems correspond to the hyperbolic and parabolic initial value problems, while an ordinary boundary value problem may be viewed as an elliptic boundary value problem in one variable. In particular one can view

$$-u''(x) = f(x) \quad \text{for} \quad x \in (0,1), \tag{4.1.1a}$$

$$u(0) = \varphi_0, \quad u(1) = \varphi_1 \tag{4.1.1b}$$

as the one-dimensional Poisson equation $-\Delta u = f$ in the domain $\Omega = (0,1)$ with Dirichlet conditions on the boundary $\Gamma = \{0,1\}$.

Difference methods are characterised by the fact that derivatives are replaced by difference quotients (divided differences), in the following called, for short, "differences" . The first derivative $u'(x)$ can be approximated by several (so-called "first") differences, for example, by the forward or right difference:

$$(\partial^+ u)(x) := [u(x+h) - u(x)]/h, \tag{4.1.2a}$$

the backward or left difference

$$(\partial^- u)(x) := [u(x) - u(x-h)]/h, \tag{4.1.2b}$$

or the symmetric difference

$$(\partial^0 u)(x) := [u(x+h) - u(x-h)]/(2h), \tag{4.1.2c}$$

where $h > 0$ is called the step size. An obvious second difference for $u''(x)$ is

$$(-\partial^-\partial^+ u)(x) := [u(x+h) - 2u(x) + u(x-h)]/h^2. \tag{4.1.3}$$

One also calls ∂^+, ∂^-, ∂^0, and $\partial^-\partial^+$ index/ D7/emphdifference operators. The product $\partial^-\partial^+$ may be viewed as $\partial^- \circ \partial^+$ or as $\partial^+ \circ \partial^-$, i.e.. $(\partial^+\partial^-)u(x) = \partial^+(\partial^- u(x))$.

Lemma 4.1.1. *Let $[x-h, x+h] \subset \overline{\Omega}$. Then*

$$\partial^\pm u(x) = u'(x) + hR \quad \text{with } |R| \le \frac{1}{2}\|u\|_{C^2(\overline{\Omega})} \text{ if } u \in C^2(\overline{\Omega}) \tag{4.1.4a}$$

$$\partial^0 u(x) = u'(x) + h^2 R \text{ with } |R| \le \frac{1}{6}\|u\|_{C^3(\overline{\Omega})} \text{ if } u \in C^3(\overline{\Omega}) \tag{4.1.4b}$$

$$\partial^-\partial^+ u(x) = u''(x) + h^2 R \text{ with } |R| \le \frac{1}{12}\|u\|_{C^4(\overline{\Omega})} \text{ if } u \in C^4(\overline{\Omega}). \tag{4.1.4c}$$

PROOF. We give the proof only for (4c). If one applies Taylor's formula

$$u(x \pm h) = u(x) \pm hu'(x) + h^2 u''(x)/2 \pm h^3 u'''(x)/6 + h^4 R_4, \tag{4.1.5a}$$

$$R_4 = h^{-4} \int_x^{x \pm h} u''''(\xi)(x \pm h - \xi)^3/3! d\xi = u''''(x \pm \vartheta h)/4!, \tag{4.1.5b}$$

with $\vartheta \in (0,1)$, to Equation (3), the result is (4c) because $R = [u''''(x + \vartheta_1 h) + u''''(x - \vartheta_2 h)]/24$. ∎

Figure 4.1.1 Grid for $h = 1/8$

We replace $\Omega = (0,1)$ and $\overline{\Omega} = [0,1]$ by the grids

$$\Omega_h = \{h, 2h, \cdots, (n-1)h = 1-h\}, \tag{4.1.6a}$$

$$\overline{\Omega}_h = \{0, h, 2h, \cdots, 1-h, 1\} \tag{4.1.6b}$$

of step size $h = 1/n$. For $x \in \Omega_h$, $\partial^-\partial^+ u(x)$ only contains the values of u at $x, x \pm h \in \overline{\Omega}_h$. Under the assumption that the solution u of Equations (1a,b) belongs to $C^4(\overline{\Omega})$, (4c) yields the equations

$$-\partial^-\partial^+ u(x) = f(x) + O(h^2), \qquad x \in \Omega_h. \tag{4.1.7}$$

If one neglects the remainder term $O(h^2)$ in Equation (7), one obtains

$$-\partial^-\partial^+ u_h(x) \equiv h^{-2}[-u_h(x-h) + 2u_h(x) - u_h(x+h)] = f(x) \quad \text{for} \quad x \in \Omega_h. \tag{4.1.8a}$$

These are $n-1$ equations in $n+1$ unknowns $\{u_n(x), x \in \overline{\Omega}_h\}$. The two missing equations are supplied by boundary conditions (1b):

$$u_h(0) = \varphi_0, \quad u_h(1) = \varphi_1. \tag{4.1.8b}$$

u_h is a grid function defined on $\overline{\Omega}_h$. Its restriction to Ω_h yields the vector

$$u_h = (u_h(h), u_h(2h), \cdots, u_h(1-h))^\mathsf{T}.$$

If in (8a) one eliminates the components $u_h(0)$ and $u_h(1)$ with the aid of Equation (8b), one gets the system of equations

$$L_h u_h = q_h \tag{4.1.9a}$$

with

$$L_h = h^{-2}\begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}, \tag{4.1.9b}$$

$$q_h = (f(h)+h^{-2}\varphi_0, f(2h), f(3h), \cdots, f(1-2h), f(1-h)+h^{-2}\varphi_1)^{\mathsf{T}}. \tag{4.1.9c}$$

4.2 The Five-Point Formula

First we select the unit square

$$\Omega = \{(x,y): 0 < x < 1,\ 0 < y < 1\}$$

as the fundamental domain. More general domains will be discussed in Section 4.8. In the discretisation process Ω is replaced by the grid

$$\Omega_h := \{(x,y) \in \Omega: x/h,\ y/h \in \mathbb{Z}\} \tag{4.2.1a}$$

for step size $h = \frac{1}{n}$ $(n \in \mathbb{N})$. The discrete boundary points form the set

$$\Gamma_h := \{(x,y) \in \Gamma: x/h, y/h \in \mathbb{Z}\}. \tag{4.2.1b}$$

As in (1.1b) we set

$$\overline{\Omega}_h := \Omega_h \cup \Gamma_h = \{(x,y) \in \overline{\Omega}_h: x/h,\ y/h \in \mathbb{Z}\}. \tag{4.2.1c}$$

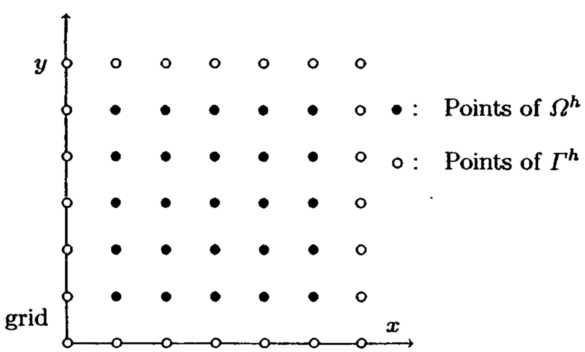


Figure 4.2.1
A two-dimensional grid

In the Poisson equation

$$-\Delta u \equiv -u_{xx} - u_{yy} = f \text{ in } \Omega, \tag{4.2.2a}$$

$$u = \varphi \text{ on } \Gamma, \tag{4.2.2b}$$

the second derivatives u_{xx} and u_{yy} can each be replaced by the respective differences (1.4c) in the x and y directions:

$$\begin{aligned}(\Delta_h u)(x,y) &:= (\partial_x^- \partial_x^+ + \partial_y^- \partial_y^+)u(x,y)\\ &= h^{-2}[u(x-h,y)+u(x+h,y)+u(x,y-h)+u(x,y+h)-4u(x,y)].\end{aligned} \tag{4.2.3}$$

Since on the right side of (3) the function u is evaluated at five points, Δ_h is also called the Five-Point Formula. The discretisation of the boundary value problem (2a,b) using Δ_h leads to the difference equations

$$-\Delta_h u_h(\mathbf{x}) = f(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Omega_h, \tag{4.2.4a}$$

$$u_h(\mathbf{x}) = \varphi(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Gamma_h. \tag{4.2.4b}$$

Through (4a,b) one obtains one equation per grid point $\mathbf{x} = (x,y) \in \overline{\Omega}_h$, and hence one equation per component of the grid function $u_h = (u_h(\mathbf{x}))_{\mathbf{x}\in\overline{\Omega}_h}$. Except in the one-dimensional case, there exists no natural arrangement of grid points, thus one cannot immediately obtain a matrix representation as in (1.9b). The only natural indexing of u_n is that through $\mathbf{x} \in \Omega_h$ or the pair $(i,j) \in \mathbb{Z}^2$ with $\mathbf{x} = (x,y) = (ih, jh)$. Let the matrix elements be given by

$$L_{\mathbf{x}\xi} := \begin{cases} 4h^{-2} & \text{if } \mathbf{x} = \xi \in \Omega_h, \\ -h^{-2} & \text{if } \mathbf{x} \in \Omega_h, \xi \in \overline{\Omega}_h, \mathbf{x} - \xi = \binom{0}{\pm h} \text{ or } \mathbf{x} - \xi = \binom{\pm h}{0}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.2.5}$$

For $\mathbf{x} = \xi$, $L_{\mathbf{xx}}$ is a diagonal element; in the second case, $L_{\mathbf{x}\xi} = -h^{-2}$, we say that $\mathbf{x}$ and ξ are neighbours. If one eliminates the components $u_h(\mathbf{x})$, $\mathbf{x} \in \Gamma_h$, with the aid of Equation (4b), then Equation (4a) assumes the following form:

$$\sum_{\xi\in\Omega_h} L_{\mathbf{x}\xi} u_h(\xi) = q_h(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Omega_h, \tag{4.2.6a}$$

where

$$q_h := f_h + \varphi_h, \quad f_h(\mathbf{x}) := f(\mathbf{x}), \quad \varphi_h(\mathbf{x}) := -\sum_{\xi\in\Gamma_h} L_{\mathbf{x}\xi}\varphi(\xi). \tag{4.2.6b}$$

For the proof split the sum

$$-\Delta_h u_h(\mathbf{x}) = \sum_{\xi\in\overline{\Omega}_h} L_{\mathbf{x}\xi} u_h(\xi)$$

into $\sum_{\xi\in\Omega_h}\dots$ and $\sum_{\xi\in\Gamma_h}\dots$. The second partial sum is $-\varphi_h$ and is moved to the right side of the equation.

Remark 4.2.1. f_h is the restriction of f to the grid Ω_h. For all points far from the boundary we have $\varphi_h(\mathbf{x}) = 0$; here $\mathbf{x} \in \Omega_h$ is said to be far from the boundary if all neighbours $\mathbf{x} \pm (0,h)$, $\mathbf{x} \pm (h,0)$ belong to Ω_h. In the case of homogeneous boundary values $\varphi = 0$ we have $q_h = f_h$.

The system of equations (6a) can be expressed in the form (1.9a):

$$L_h u_h = q_h$$

where the matrix

$$L_h = (L_{\mathbf{x}\xi})_{\mathbf{x},\xi \in \Omega_h}$$

and the grid functions $u_h = (u_h(\mathbf{x}))_{\mathbf{x}\in\Omega_h}$ and $q_h = (q_h(\mathbf{x}))_{\mathbf{x}\in\Omega_h}$ are described by their components (cf. (5), (6b)). Strictly speaking L_h is not a matrix in the usual sense but a linear mapping since the indices $\mathbf{x} \in \Omega_h$ are not ordered.

Exercise 4.2.2. (a) An $N \times N$ matrix $\mathbf{P}$ is called a permutation matrix if $\mathbf{w} := \mathbf{P}\mathbf{v}$, for all $\mathbf{v} \in \mathbb{R}^N$, has the coefficients $w_i = v_{\pi(i)}$, $1 \le i \le N$, where π is a permutation of indices $\{1, \cdots, N\}$. Show that $\mathbf{P}$ is unitary, i.e., $\mathbf{P}^{-1} = \mathbf{P}^{\mathsf{T}}$.
(b) Let I be an index set with N elements. Let the "matrix" coefficients $A_{\alpha\beta}$, $\alpha, \beta \in I$, be given. To the arrangement $\alpha_1, \cdots, \alpha_N$ of indices corresponds the $N \times N$ matrix $\mathbf{A} = (a_{ij})_{i,j=1,\cdots,N}$ with $a_{ij} = A_{\alpha_i\alpha_j}$. Let $\hat{\mathbf{A}}$ be the matrix which belongs to a second arrangement $\hat\alpha_1, \cdots, \hat\alpha_N$ of I. Prove that there exists a permutation matrix $\mathbf{P}$ such that $\hat{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^{\mathsf{T}}$.
(c) Let $A_{\alpha\beta} = A_{\beta\alpha}$ for all $\alpha, \beta \in I$. For each arrangement of the indices the corresponding matrix is symmetric.

A possible linear enumeration of indices $\mathbf{x} \in \Omega_h$ is lexicographical ordering

$$\begin{array}{lllll} (h,h), & (2h,h), & (3h,h), & \cdots, & (1-h,h), \\ (h,2h), & (2h,2h), & (3h,2h), & \cdots, & (1-h,2h), \\ & & \vdots & & \\ (h,1-h), & (2h,1-h), & (3h,1-h), & \cdots, & (1-h,1-h). \end{array} \tag{4.2.7}$$

Generally, the point $\mathbf{x} = (x_1, \cdots, x_d)$ precedes the point $\mathbf{y} = (y_1, \cdots, y_d)$ in lexicographical order, if for a $j \in \{1, \cdots, d\}$ the conditions $x_i = y_i$ $(i > j)$ and $x_j < y_j$ hold. Each line in (7) corresponds to a so-called x-row in the grid Ω_h. A vector u_h whose $(n-1)^2$ components are enumerated in the series (7) thus separates into $n-1$ blocks (so-called x-blocks). The block decomposition of the vectors generates a block decomposition of the matrix L_h which is given in (8).

Exercise 4.2.3. (a) With the lexicographical numbering of grid points the matrix L_h has the form

$$L_h = h^{-2}\begin{bmatrix} T & -I & & & \\ -I & T & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & T & -I \\ & & & -I & T \end{bmatrix}, T = \begin{bmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 4 \end{bmatrix}, \tag{4.2.8}$$

where T is an $(n-1) \times (n-1)$ matrix and L_h contains $(n-1)^2$ blocks. I is the $(n-1) \times (n-1)$ identity matrix.
(b) Let Ω be the rectangle

$$\Omega = (0,a) \times (0,b) = \{(x,y): 0 < x < a,\ 0 < y < b\}.$$

Let the step size h satisfy the conditions $a = nh$ and $b = mh$. Show that the discretisation (4a,b) in the corresponding grid Ω_h leads to a matrix which also has the form (8). But here L_h contains $(m-1)^2$ blocks of the size $(n-1) \times (n-1)$.

Another frequently used arrangement is the chequer-board ordering (or red-black ordering). To this end one chessboard pattern divides Ω_h into "red" and "black" fields:

$$\begin{aligned} \Omega_h^{\mathrm{r}} &:= \{(x,y) \in \Omega_h: (x+y)/h \text{ odd}\}, \\ \Omega_h^{\mathrm{b}} &:= \{(x,y) \in \Omega_h: (x+y)/h \text{ even}\}. \end{aligned} \tag{4.2.9}$$

First one numbers the red squares $(x,y) \in \Omega_h^{\mathrm{r}}$ lexicographically, and then those of Ω_h^{b}. The partition (9) induces a partition of vectors into 2 blocks and a partition of the matrix L_h into $2 \cdot 2 = 4$ blocks.

Exercise 4.2.4. With respect to the chequer-board ordering, the matrix L_h assumes the form

$$L_h = h^{-2}\left[\begin{array}{ccc|ccc} 4 & & & & & \\ & \ddots & & & A & \\ & & 4 & & & \\ \hline & & & 4 & & \\ & A^{\mathsf{T}} & & & \ddots & \\ & & & & & 4 \end{array}\right] \tag{4.2.10}$$

where, in general, $\mathbf{A}$ is a rectangular block matrix because for n even, Ω_h^{r} and Ω_h^{b} contain a different number of points.

The complete $(n-1)^2 \times (n-1)^2$ matrix L_h in (8) or (10) is needed neither for the theoretical investigation of the system of equations $L_h u_h = q_h$ nor for its numerical solution. All properties of L_h considered in the following are invariant with respect to re-numbering of the grid points. Even though numerical methods for the solution of $L_h u_h = q_h$ implicitly use an arrangement of grid points (with the exception of special algorithms for parallel computers), they never employ the complete $(n-1)^2 \times (n-1)^2$ matrix L_h. Every usable algorithm must take into account that L_h is **sparse**, i.e., it has substantially more zero than nonzero elements.

In the following we again return to indexing by $\mathbf{x} \in \Omega_h$. Nevertheless we will continue to refer to the L_h defined by (5) as a matrix.

The difference operator Δ_h is also described by the star

$$\Delta_h = h^{-2} \begin{bmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{bmatrix}. \qquad (4.2.11)$$

The general definition of a **difference star** (with variable coefficients) reads

$$\begin{bmatrix} & & \vdots & & \\ & c_{-1,1}(x,y) & c_{0,1}(x,y) & c_{1,1}(x,y) & \\ \cdots & c_{-1,0}(x,y) & c_{0,0}(x,y) & c_{1,0}(x,y) & \cdots \\ & c_{-1,-1}(x,y) & c_{0,-1}(x,y) & c_{1,-1}(x,y) & \\ & & \vdots & & \end{bmatrix}$$

$$= \sum_{i,j=-\infty}^{\infty} c_{ij}(x,y) u_h(x+ih, y+jh), \qquad (4.2.12)$$

in which the zero coefficients have not been written out.

Attention. The star (11) does not represent a submatrix of L_h! The coefficients of the star appear in each row of L_h.

Remark 4.2.5. Note that the difference operator $-\Delta_h$ cannot be equated with the matrix L_h since Δ_h does not contain information on the type or place of the boundary conditions. We shall say the matrix L_h belongs to a difference star (12) if the system of equations $L_h u_h = q_h$ results from the difference equations in $\mathbf{x} \in \Omega_h$ after elimination of the Dirichlet boundary values $u_h(\mathbf{x}) = \varphi(\mathbf{x})$, $\mathbf{x} \in \Gamma_h$. Even if the vector u_h in $L_h u_h = q_h$ contains only components $u_h(\mathbf{x})$, $\mathbf{x} \in \Omega_h$, one occasionally equates u_h with the grid function on $\overline{\Omega}_h$ which assumes the prescribed boundary values (4b) on Γ_h.

4.3 *M*-matrices, Matrix Norms, Positive Definite Matrices

The elements of the matrix $\mathbf{A}$ are denoted by $a_{\alpha\beta}$, $\alpha, \beta \in I$. Here $\mathbf{A}$ and the index set I assume the places of L_h and Ω_h. We write

$$\mathbf{A} \geq \mathbf{B} \quad \text{if } a_{\alpha\beta} \geq b_{\alpha\beta} \quad \text{for all } \alpha, \beta \in I,$$

and define analogously $\mathbf{A} \leq \mathbf{B}, \mathbf{A} > \mathbf{B}, \mathbf{A} < \mathbf{B}$. The zero matrix is denoted by $\mathbf{0}$.

Definition 4.3.1. $\mathbf{A}$ is called an **M-matrix** if

$$a_{\alpha\alpha} > 0 \quad \text{for all } \alpha \in I, \quad a_{\alpha\beta} \leq 0 \quad \text{for all } \alpha \neq \beta, \qquad (4.3.1a)$$

$$\mathbf{A} \text{ nonsingular and } \mathbf{A}^{-1} \geq \mathbf{0}. \qquad (4.3.1b)$$

The inequalities (1a) can immediately be proved for L_h (cf. (2.5)). However we still need criteria and auxiliary results to prove (1b).

The index $\alpha \in I$ is said to be **directly connected** with $\beta \in I$ if $a_{\alpha\beta} \neq 0$. We say that $\alpha \in I$ is **connected** with $\beta \in I$ if there exists a "connection" (chain of direct connections)

$$\alpha = \alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_k = \beta \quad \text{with } a_{\alpha_{i-1}\alpha_i} \neq 0 \quad (1 \leq i \leq k). \qquad (4.3.2)$$

The index set I together with the direct connections form the **graph of A** (cf. Fig. 1). Frequently $\mathbf{A}$ has a symmetrical structure, i.e., $a_{\alpha\beta} \neq 0$ holds if and only if $a_{\beta\alpha} \neq 0$. In this case α is (directly) connected with β if and only if β is (directly) connected with α.

Definition 4.3.2. A matrix $\mathbf{A}$ is said to be **irreducible** if every $\alpha \in I$ is connected with every $\beta \in I$.

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

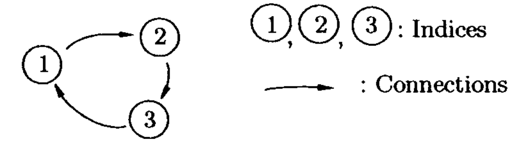


Figure 4.3.1. Graph of the irreducible matrix $\mathbf{A}$

In the case of the matrix $\mathbf{A} = L_h$, two indices $\mathbf{x}, \mathbf{y} \in \Omega_h$ are connected if and only if $\mathbf{y} = \mathbf{x}$ or if $\mathbf{y}$ is a neighbour of $\mathbf{x}$. Arbitrary $\mathbf{x}, \mathbf{y} \in \Omega_h$ can evidently be connected by a chain $\mathbf{x} = \mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(k)} = \mathbf{y}$ of neighbouring points. Thus L_h is irreducible.

Exercise 4.3.3. Prove that $\mathbf{A}$ is irreducible if and only if there is no ordering of the indices such that the resulting matrix has the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix},$$

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are respectively square $n_1 \times n_1$ and $n_2 \times n_2$ matrices $(n_1 \geq 1, n_2 \geq 1)$, and $\mathbf{A}_{12}$ is an $n_1 \times n_2$ submatrix.

The important question as to whether $\mathbf{A} = L_h$ is nonsingular can be treated as a special case of the following statement.

Criterion 4.3.4. *(Gershgorin) Let $K_r(z)$ denote the open disk $\{\zeta \in \mathbb{C}: |z - \zeta| < r\}$, and let $\overline{K_r(z)} := \{\zeta \in \mathbb{C}: |z-\zeta| \le r\}$ denote the closed disk.*
(a) All eigenvalues of A lie in

$$\bigcup_{\alpha\in I} \overline{K_{r_\alpha}(a_{\alpha\alpha})} \quad \text{with } r_\alpha = \sum_{\substack{\beta\neq\alpha\\ \beta\in I}} |a_{\alpha\beta}|.$$

(b) If $\mathbf{A}$ is irreducible, the eigenvalues even lie in

$$\bigcup_{\alpha\in I} K_{r_\alpha}(a_{\alpha\alpha}) \cup (\bigcap_{\alpha\in I} \partial K_{r_\alpha}(a_{\alpha\alpha})).$$

PROOF. (a) Let λ be an eigenvalue of $\mathbf{A}$ and $\mathbf{u}$ a corresponding eigenvector which, without loss of generality, satisfies $||\mathbf{u}||_\infty = 1$, where

$$||\mathbf{u}||_\infty := \max\{|u_\alpha|: \alpha \in I\} \tag{4.3.3}$$

is the maximum norm. There exists (at least) one index $\gamma \in I$ with $|u_\gamma| = 1$.

Assertion 1. $|u_\gamma| = 1$ *implies*

$$|\lambda - a_{\gamma\gamma}| \le \sum_{\substack{\beta\neq\gamma\\ \beta\in I}} |a_{\gamma\beta}||u_\beta| \le \sum_{\substack{\beta\neq\gamma\\ \beta\in I}} |a_{\gamma\beta}| = r_\gamma. \tag{$*$}$$

From $(*)$ follows $\lambda \in \overline{K_{r_\gamma}(a_{\gamma\gamma})}$ and hence the statement. To prove the assertion use the equation from $\mathbf{Au} = \lambda\mathbf{u}$ associated to the index γ:

$$\lambda u_\gamma = \sum_{\beta\in I} a_{\gamma\beta}u_\beta, \quad \text{that is} \quad (\lambda - a_{\gamma\gamma})u_\gamma = \sum_{\substack{\beta\neq\gamma\\ \beta\in I}} a_{\gamma\beta}u_\beta.$$

From $|u_\gamma| = 1$ follows $|\lambda - a_{\gamma\gamma}| = |(\lambda - a_{\gamma\gamma})u_\gamma| \le |\sum_{\beta\neq\gamma} a_{\gamma\beta}u_\beta|$. By taking the modulus into the sum and by using $|u_\beta| \le ||\mathbf{u}||_\infty = 1$, $(*)$ follows.
(b) Let $\mathbf{A}$ be irreducible and let λ be an arbitrary eigenvalue of $\mathbf{A}$ with associated eigenvector $\mathbf{u}$ which in turn is again normalised by $||\mathbf{u}||_\infty = 1$. The case $\lambda \in \bigcup_{\alpha\in I} K_{r_\alpha}(a_{\alpha\alpha})$ immediately leads to the statement. Therefore let $\lambda \notin \bigcup_{\alpha\in I} K_{r_\alpha}(a_{\alpha\alpha})$ be assumed.

Assertion 2. *Let $a_{\alpha\beta} \neq 0$, i.e., γ is directly connected with β; then $|u_\gamma| = 1$ and $|\lambda - a_{\gamma\gamma}| = r_\gamma$ implies $|u_\beta| = 1$ and $|\lambda - a_{\beta\beta}| = r_\beta$.*

Part (a) proves the existence of a $\gamma \in I$ with $|u_\gamma| = 1$ and $|\lambda - a_{\gamma\gamma}| \le r_\gamma$. According to the assumption, $|\lambda - a_{\gamma\gamma}| = r_\gamma$ must hold so that assertion 2 is applicable to γ. Since $\mathbf{A}$ is irreducible, for an arbitrary $\beta \in I$ there exists a connection (2) of γ with β: $\alpha_0 = \gamma, \alpha_1, \cdots, \alpha_k = \beta$, $a_{\alpha_{i-1}\alpha_i} \neq 0$. Assertion 2 shows

$$|u_{\alpha_i}| = 1 \quad \text{and } |\lambda - a_{\alpha_i\alpha_i}| = r_{\alpha_i} \quad \text{for all } i = 0, \cdots, k;$$

in particular, $\lambda \in \partial K_{r_\beta}(a_{\beta\beta})$ for $\beta = \alpha_k$. Since β was chosen arbitrarily, it follows that $\lambda \in \bigcap_{\alpha\in I} \partial K_{r_\alpha}(a_{\alpha\alpha})$, and the statement is proved.
Proof of Assertion 2. Besides the inequality chain $(*)$ there also holds $|\lambda - a_{\gamma\gamma}| = r_\gamma$ so that all the inequalities in $(*)$ become equations. In particular

$$\sum_{\beta\neq\gamma} |a_{\gamma\beta}|\,|u_\beta| = \sum_{\beta\neq\gamma} |a_{\gamma\beta}|$$

must hold. Since $|u_\beta| \le ||\mathbf{u}||_\infty = 1$, the identity $|a_{\gamma\beta}|\,|u_\beta| = |a_{\gamma\beta}|$ must be satisfied for each summand. Hence $a_{\gamma\beta} \neq 0$ implies $|u_\beta| = 1$. The application of Assertion 1 to β yields $|\lambda - a_{\beta\beta}| \le r_\beta$. The assumption $\lambda \notin \bigcup_{\alpha\in I} K_{r_\alpha}(a_{\alpha\alpha})$ proves $|\lambda - a_{\beta\beta}| = r_\beta$. ∎

Exercise 4.3.5. Let $I_\alpha := \{\beta \in I : \text{a connection (2) exists between } \alpha \text{ and } \beta\}$. Show that the eigenvalues of $\mathbf{A}$ lie in

$$\bigcup_{\alpha\in I} \left\{ K_{r_\alpha}(a_{\alpha\alpha}) \cup \bigcap_{\beta\in I_\alpha} \partial K_{r_\beta}(a_{\beta\beta}) \right\}.$$

Definition 4.3.6. (a) $\mathbf{A}$ is said to be diagonally dominant if

$$\sum_{\beta\neq\alpha} |a_{\alpha\beta}| < |a_{\alpha\alpha}| \tag{4.3.4a}$$

for all $\alpha \in I$. (b) $\mathbf{A}$ is said to be irreducibly diagonally dominant if $\mathbf{A}$ is irreducible, the inequality (4a) holds for at least one index $\alpha \in I$ and

$$\sum_{\substack{\beta\neq\alpha\\ \beta\in I}} |a_{\alpha\beta}| \le |a_{\alpha\alpha}| \quad \text{for all } \alpha \in I. \tag{4.3.4b}$$

Note that while an irreducible and diagonally dominant matrix is irreducibly diagonally dominant, the reverse need not hold.

The matrix L_h from Section 4.2, while not diagonally dominant, is irreducibly diagonally dominant, for L_h is irreducible and satisfies (4b). At all points near the boundary—i.e., those $\mathbf{x} \in \Omega_h$, which have a boundary point $\mathbf{y} \in \Gamma_h$ as a neighbour—however, (4a) holds: $\sum_{\substack{\beta\neq\alpha\\ \beta\in I}} |a_{\alpha\beta}| \le 3h^{-2} < 4h^{-2} = a_{\alpha\alpha}$.

The spectral radius $\rho(\mathbf{A})$ of a matrix $\mathbf{A}$ is given by the eigenvalue that is largest in modulus:

$$\rho(\mathbf{A}) := \max\{|\lambda|: \lambda \text{ eigenvalue of } \mathbf{A}\}. \tag{4.3.5}$$

In the following we split $\mathbf{A}$ into

$$\mathbf{A} = \mathbf{D} - \mathbf{B}, \quad \mathbf{D} = \operatorname{diag}\{a_{\alpha\alpha}: \alpha \in I\}, \tag{4.3.6a}$$

where $\mathbf{D}$ is the diagonal part of $\mathbf{A}$:

$$d_{\alpha\alpha} = a_{\alpha\alpha}, \quad d_{\alpha\beta} = 0 \quad \text{for } \alpha \neq \beta. \tag{4.3.6b}$$

$\mathbf{B} := \mathbf{D} - \mathbf{A}$ is the off-diagonal part:

$$b_{\alpha\alpha} = 0, \quad b_{\alpha\beta} = -a_{\alpha\beta} \quad \text{for } \alpha \neq \beta. \tag{4.3.6c}$$

Criterion 4.3.7. *Let* (6a-c) *hold. Sufficient conditions for*

$$\rho(\mathbf{D}^{-1}\mathbf{B}) < 1 \tag{4.3.7}$$

are the diagonal dominance or the irreducible diagonal dominance of $\mathbf{A}$.

PROOF. (a) The coefficients of $\mathbf{C} := \mathbf{D}^{-1}\mathbf{B}$ read

$$c_{\alpha\beta} = -a_{\alpha\beta}/a_{\alpha\alpha} \quad (\alpha \neq \beta), \quad c_{\alpha\alpha} = 0.$$

From the diagonal dominance (4a) follows that $r_\alpha := \sum_{\beta\neq\alpha} |c_{\alpha\beta}| < 1$ for all $\alpha \in I$. By the Gershgorin Criterion 4a all eigenvalues λ of $\mathbf{C}$ lie in $\bigcup_{\alpha\in I} \overline{K_{r_\alpha}(c_{\alpha\alpha})} = \bigcup_{\alpha\in I} \overline{K_{r_\alpha}(0)}$ so that $|\lambda| \leq \max r_\alpha < 1$ and hence $\rho(\mathbf{C}) = \rho(\mathbf{D}^{-1}\mathbf{B}) < 1$ also follows.
(b) If $\mathbf{A}$ is irreducibly diagonally dominant then $r_\beta \leq 1$ for all $\beta \in I$ and $r_\alpha < 1$ for at least one α. According to Criterion 4b all eigenvalues of $\mathbf{C}$ lie in $\bigcup_{\beta\in I} K_{r_\beta}(0) \cup (\bigcap_{\beta\in I} \partial K_{r_\beta}(0))$. This set lies in $K_1(0)$ if $\bigcap_{\beta\in I} \partial K_{r_\beta}(0) \subset K_1(0)$. At first let us assume that all r_β agree: $r_\beta = r$ for all β. Since $r_\alpha < 1$ for one $\alpha \in I$, it follows that $r < 1$ and $\bigcap_\beta \partial K_{r_\beta}(0) = \partial K_r(0) \subset K_1(0)$. But if all r_β are not equal then $\bigcap_\beta \partial K_{r_\beta}(0)$ is empty. Thus in both cases $\lambda \in K_1(0)$ holds and (7) is proved. ■

Exercise 4.3.8. (a) Weaken irreducible diagonal dominance as follows: Let $\mathbf{A}$ satisfy the inequalities (4b) and for all $\beta \in I$ let the connection (2) exist for an index $\alpha \in I$ for which the strict inequality (4a) holds. Prove that even under this assumption $\rho(\mathbf{D}^{-1}\mathbf{B}) < 1$ holds.
Hint. Use Exercise 5.
(b) Show that the geometric series $\mathbf{S} = \sum_{\nu=0}^{\infty} \mathbf{C}^\nu$ converges if and only if $\rho(\mathbf{C}) < 1$. Then the following holds: $\mathbf{S} = (\mathbf{I} - \mathbf{C})^{-1}$. Hint: Represent $\mathbf{C}$ in the form $\mathbf{QRQ}^\mathsf{T}$ ($\mathbf{Q}$ a unitary, and $\mathbf{R}$ an upper triangular matrix) and show $\|\mathbf{C}^\nu\|_\infty \leq K[\rho(\mathbf{C})]^\nu$.
(c) Let $\mathbf{u}$ be a vector. We define $|\mathbf{u}|$ as the vector (!) with the entries $|\mathbf{u}|_\alpha := |u_\alpha|$. For two vectors one writes $\mathbf{v} \leq \mathbf{w}$ if $v_\alpha \leq w_\alpha$ $(\alpha \in I)$. Show that:
(1) $\mathbf{AB} \geq \mathbf{0}$, if $\mathbf{A} \geq \mathbf{0}$, $B \geq 0$; $\mathbf{AB} > \mathbf{0}$, if $\mathbf{A} > \mathbf{0}$, $\mathbf{B} > \mathbf{0}$;
(2) $\mathbf{AD} > \mathbf{0}$ if $\mathbf{A} > \mathbf{0}$, and $\mathbf{D} \geq \mathbf{0}$ is a nonsingular diagonal matrix;
(3) $\mathbf{Av} \leq \mathbf{Aw}$ if $\mathbf{A} \geq \mathbf{0}$ and $\mathbf{v} \leq \mathbf{w}$; $\|\mathbf{v}\|_\infty \leq \|\mathbf{w}\|_\infty$ if $\mathbf{0} \leq \mathbf{v} \leq \mathbf{w}$;

(4) $\mathbf{Au} \leq |\mathbf{Au}| \leq \mathbf{A}|\mathbf{u}|$ if $\mathbf{A} \geq \mathbf{0}$.

The importance of inequality (7) results from

Lemma 4.3.9. *Let* $\mathbf{A}$ *satisfy* (1a). *Let* $\mathbf{D}$ *and* $\mathbf{B}$ *be defined by* (6a-c). $\mathbf{A}$ *is an M-matrix if and only if* $\rho(\mathbf{D}^{-1}\mathbf{B}) < 1$.

PROOF. (a) Let $\mathbf{C} := \mathbf{D}^{-1}\mathbf{B}$ satisfy $\rho(\mathbf{C}) < 1$. Then the geometric series $\mathbf{S} := \sum_{\nu=0}^{\infty} \mathbf{C}^\nu$ converges (cf. Exercise 8b). From $\mathbf{D}^{-1} \geq \mathbf{0}$ and $\mathbf{B} \geq \mathbf{0}$ one infers $\mathbf{C} \geq \mathbf{0}$, $\mathbf{C}^\nu \geq \mathbf{0}$, and $\mathbf{S} \geq \mathbf{0}$. Since $I = \mathbf{S}(\mathbf{I} - \mathbf{C}) = \mathbf{SD}^{-1}\mathbf{D} - \mathbf{B} = (\mathbf{SD})^{-1}\mathbf{A}$, $\mathbf{A}$ has the inverse $\mathbf{A}^{-1} = \mathbf{SD}^{-1}$. $\mathbf{D}^{-1} \geq \mathbf{0}$ and $\mathbf{S} \geq \mathbf{0}$ result in $\mathbf{A}^{-1} \geq \mathbf{0}$. From this (1b) also results, i.e., $\mathbf{A}$ is an M-matrix.
(b) Let $\mathbf{A}$ be an M-matrix. For an eigenvalue λ of $\mathbf{D}^{-1}\mathbf{B}$ select an eigenvector $\mathbf{u} \neq 0$. According to Exercise 8c we have

$$|\lambda|\,|\mathbf{u}| = |\lambda\mathbf{u}| = |\mathbf{D}^{-1}\mathbf{B}\mathbf{u}| \leq \mathbf{D}^{-1}\mathbf{B}|\mathbf{u}|.$$

Because $\mathbf{A}^{-1}\mathbf{D} \geq \mathbf{0}$ (cf. (1a,b)) one obtains $-\mathbf{A}^{-1}\mathbf{D}\mathbf{D}^{-1}\mathbf{B}|\mathbf{u}| \leq -\mathbf{A}^{-1}\mathbf{D}|\lambda|\,|\mathbf{u}|$ so that

$$\begin{aligned}|\mathbf{u}| &= \mathbf{A}^{-1}(\mathbf{D} - \mathbf{B})|\mathbf{u}| = \mathbf{A}^{-1}\mathbf{D}(\mathbf{I} - \mathbf{D}^{-1}\mathbf{B})|u| \leq \mathbf{A}^{-1}\mathbf{D}|\mathbf{u}| - \mathbf{A}^{-1}\mathbf{D}|\lambda|\,|\mathbf{u}| \\ &= (1 - |\lambda|)\mathbf{A}^{-1}\mathbf{D}|\mathbf{u}|\end{aligned}$$

follows. For $|\lambda| \geq 1$ we would get the inequality, $|\mathbf{u}| \leq 0$, i.e., $\mathbf{u} = 0$, in contradiction to the assumption $\mathbf{u} \neq \mathbf{0}$. From this follows $|\lambda| < 1$ for every eigenvalue of $\mathbf{C} = \mathbf{D}^{-1}\mathbf{B}$, thus $\rho(\mathbf{D}^{-1}\mathbf{B}) < 1$. ■

Criterion 7 and Lemma 9 imply

Criterion 4.3.10. *If a matrix* $\mathbf{A}$ *with the property* (1a) *is diagonally dominant or irreducibly diagonally dominant, then* $\mathbf{A}$ *is an M-matrix.*

Theorem 4.3.11. *An irreducible M-matrix* $\mathbf{A}$ *has an element-wise positive inverse:* $\mathbf{A}^{-1} > \mathbf{0}$.

PROOF. Let $\alpha, \beta \in I$ be selected arbitrarily. There exists a connection (2): $\alpha = \alpha_0, \alpha_1, \cdots, \alpha_k = \beta$. Set $\mathbf{C} := \mathbf{D}^{-1}\mathbf{B}$. Since $c_{\alpha_{i-1}\alpha_i} > 0$, it follows that

$$(\mathbf{C}^k)_{\alpha\beta} = \sum_{\gamma_1,\cdots,\gamma_{k-1}\in I} c_{\alpha\gamma_1}c_{\gamma_1\gamma_2}\cdots c_{\gamma_{k-1}\beta} \geq c_{\alpha\alpha_1}c_{\alpha_1\alpha_2}\cdots c_{\alpha_{k-1}\beta} > 0.$$

According to Lemma 9, $\rho(\mathbf{C}) < 1$ holds, so that $\mathbf{S} := \sum_{\nu=0}^{\infty} \mathbf{C}^\nu$ converges. Since $\mathbf{S}_{\alpha\beta} \geq (\mathbf{C}^k)_{\alpha\beta} > 0$ and $\alpha, \beta \in I$ are arbitrary, $\mathbf{S} > \mathbf{0}$ is proved. The assertion results from $\mathbf{A}^{-1} = \mathbf{SD}^{-1} > \mathbf{0}$ (cf. proof of Lemma 9). ■

In the following we derive norm estimates for $\mathbf{A}^{-1}$.

Definition 4.3.12. Let $\mathbf{V}$ be a linear space (vector space) over the field of real numbers ($\mathbf{K} := \mathbb{R}$) or complex numbers ($\mathbf{K} := \mathbb{C}$). The functional $\|\cdot\|$ is called a norm in $\mathbf{V}$ if

$$\|\mathbf{u}\| = 0 \quad \text{only for} \quad \mathbf{u} = \mathbf{0}, \tag{4.3.8a}$$
$$\|\mathbf{u}+\mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\| \quad \text{for all } \mathbf{u},\mathbf{v} \in \mathbf{V}, \tag{4.3.8b}$$
$$\|\lambda\mathbf{u}\| = |\lambda|\ \|\mathbf{u}\| \quad \text{for all } \lambda \in \mathbf{K}, \mathbf{u} \in \mathbf{V}. \tag{4.3.8c}$$

Example. Let $\mathbf{V} = \mathbb{R}^{\#I}$ where $\#I :=$ is the number of elements of the index set I. The maximum norm defined in (3) satisfies the norm axioms (8a-c).

If one views the elements $\mathbf{u} \in \mathbf{V}$ as vectors, $\|\cdot\|$ is called a vector norm. But the matrices also form a linear space. In the latter case one calls $\|\cdot\|$ a matrix norm. A special class of matrix norms is contained in

Definition 4.3.13. Let $\mathbf{V}$ be the vector space with vector norm $\|\cdot\|$. Then one calls

$$|||\mathbf{A}||| := \sup\{\|\mathbf{Au}\|/\|\mathbf{u}\| : 0 \ne \mathbf{u} \in \mathbf{V}\} \tag{4.3.9}$$

the matrix norm associated with the vector norm $\|\cdot\|$.

Exercise 4.3.14. Let $|||\cdot|||$ be defined by (9). Show that: (a) $|||\cdot|||$ is a norm; (b) the following holds:

$$|||\mathbf{AB}||| \le |||\mathbf{A}|||\ |||\mathbf{B}|||, \tag{4.3.10a}$$
$$|||I||| = 1 \quad (I: \text{unit matrix}), \tag{4.3.10b}$$
$$\|\mathbf{Au}\| \le |||\mathbf{A}|||\ \|\mathbf{u}\|, \tag{4.3.10c}$$
$$|||\mathbf{A}||| \ge \rho(\mathbf{A}). \tag{4.3.10a}$$

Example. The matrix norm associated with the maximum norm $\|\cdot\|_\infty$ (cf. (3)) is called the row sum norm and is also denoted by $\|\cdot\|_\infty$. It has the explicit representation

$$\|\mathbf{A}\|_\infty = \max_{\alpha\in I}\{\sum_{\beta\in I}|a_{\alpha\beta}|\}. \tag{4.3.11}$$

Exercise 4.3.15. (a) Prove (11). (b) For matrices $\mathbf{0} \le \mathbf{B} \le \mathbf{C}$ there holds $\|\mathbf{B}\|_\infty \le \|\mathbf{C}\|_\infty$.

In the next theorem we denote by $\mathbb{1}$ the vector having only ones as components:

$$\mathbb{1}_\alpha = 1 \quad \text{for all } \alpha \in I.$$

For the notation $\mathbf{v} \le \mathbf{w}$ see Exercise 8c.

Theorem 4.3.16. *Let* $\mathbf{A}$ *be an M-matrix and let a vector* $\mathbf{w}$ *exist with* $\mathbf{Aw} \ge \mathbb{1}$. *Then* $\|\mathbf{A}^{-1}\|_\infty \le \|\mathbf{w}\|_\infty$.

PROOF. As in the proof of Lemma 9, let $|\mathbf{u}|$ be the vector with the components $|u_\alpha|$. For each $\mathbf{u}$ we have $|\mathbf{u}| \le \|\mathbf{u}\|_\infty\mathbb{1} \le \|\mathbf{u}\|_\infty\mathbf{Aw}$. Since $\mathbf{A}^{-1} \ge \mathbf{0}$, we obtain

$$|\mathbf{A}^{-1}\mathbf{u}| \le \mathbf{A}^{-1}|\mathbf{u}| \le \|\mathbf{u}\|_\infty\mathbf{A}^{-1}\mathbf{Aw} = \|\mathbf{u}\|_\infty\mathbf{w}$$

(cf. Exercise 8c) and $\|\mathbf{A}^{-1}\mathbf{u}\|_\infty/\|\mathbf{u}\|_\infty \le \|\mathbf{w}\|_\infty$. Definition 13 implies that $\|\mathbf{A}^{-1}\|_\infty \le \|\mathbf{w}\|_\infty$. ■

How to estimate with the aid of a majorising matrix is shown in

Theorem 4.3.17. *Let* $\mathbf{A}$ *and* $\mathbf{A}'$ *be M-matrices with* $\mathbf{A}' \ge \mathbf{A}$. *Then the following holds*

$$\mathbf{0} \le \mathbf{A}'^{-1} \le \mathbf{A}^{-1} \quad \text{and} \quad \|\mathbf{A}'^{-1}\|_\infty \le \|\mathbf{A}^{-1}\|_\infty. \tag{4.3.12}$$

PROOF. $\mathbf{A}'^{-1} \le \mathbf{A}^{-1}$ follows from $\mathbf{A}^{-1} - \mathbf{A}'^{-1} = \mathbf{A}^{-1}(\mathbf{A}' - \mathbf{A})\mathbf{A}'^{-1}$ and $\mathbf{A}^{-1} \ge \mathbf{0}$, $\mathbf{A}' - \mathbf{A} \ge \mathbf{0}$, $\mathbf{A}'^{-1} \ge \mathbf{0}$. The remainder follows from Exercise 15b. ■

Exercise 4.3.18. Prove (12) under the following weaker assumptions: $\mathbf{A}$ is an M-matrix, $\mathbf{A}'$ satisfies (1a) and $\mathbf{A}' \ge \mathbf{A}$. Hint: Repeat the considerations from the first part of the proof of Lemma 9 with the matrices $\mathbf{D}'$ and $\mathbf{B}'$ associated to $\mathbf{A}'$.

Exercise 4.3.19. Let $\mathbf{B}$ a principal submatrix of $\mathbf{A}$, that is, there exists a subset $I' \subset I$ such that $\mathbf{B}$ is given by the entries $b_{\alpha\beta} = a_{\alpha\beta}, \alpha, \beta \in I'$. Prove that if $\mathbf{A}$ is an M-matrix, then so is $\mathbf{B}$ and $0 \le (\mathbf{B}^{-1})_{\alpha\beta} \le (\mathbf{A}^{-1})_{\alpha\beta}$ holds for all $\alpha, \beta \in I'$. Hint: Apply Exercise 18 to the following matrix $\mathbf{A}'$: $a'_{\alpha\beta} = a_{\alpha\beta}$ $(\alpha, \beta \in I')$, $a'_{\alpha\alpha} = a_{\alpha\alpha}$ $(\alpha \in I\backslash I')$, $a'_{\alpha\beta} = 0$ otherwise.

Another well-known vector norm is the Euclidean norm

$$\|\mathbf{u}\|_2 := [c\sum_{\alpha\in I}|\mathbf{u}_\alpha|^2]^{1/2} \tag{4.3.13}$$

with fixed scaling constant $c > 0$ (for example, the choice $c = h^2$ in connection with the grid functions from Section 4.2 results in the fact that $c\sum_{\alpha\in I}$ represents an approximation to the integration $\int_\Omega$). The matrix norm associated to $\|\cdot\|_2$ is independent of the factor c. It is called the spectral norm and is also denoted by $\|\cdot\|_2$. The name derives from the following characterisation.

Exercise 4.3.20. Prove : (a) for symmetric matrices there holds $\|\mathbf{A}\|_2 = \rho(\mathbf{A})$ (cf. (5)). (b) For each real matrix holds:

$$\|\mathbf{A}\|_2 = [\rho(\mathbf{A}^\mathsf{T}\mathbf{A})]^{1/2} = [\text{maximal eigenvalue of } \mathbf{A}^\mathsf{T}\mathbf{A}]^{1/2}.$$

(c) For each matrix holds $\|\mathbf{A}\|_2^2 \le \|\mathbf{A}\|_2\|\mathbf{A}^\mathsf{T}\|_2$. Hint: (b) and (10d).

For the proof in the exercise use the scalar product

$$\langle \mathbf{u}, \mathbf{v} \rangle := c \sum_{\alpha \in I} u_\alpha v_\alpha \tag{4.3.14a}$$

(c as in (13)) and its properties

$$\langle \mathbf{u}, \mathbf{u} \rangle = \|\mathbf{u}\|_2^2, \quad \langle \mathbf{A}\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{A}^\mathsf{T}\mathbf{v} \rangle, \quad |\langle \mathbf{u}, \mathbf{v} \rangle| \le \|\mathbf{u}\|_2 \|\mathbf{v}\|_2. \tag{4.3.14b}$$

Here the case $\mathbf{K} = \mathbb{R}$ is always used as basis., i.e., all matrices and vectors are real.

Definition 4.3.21. A matrix $\mathbf{A}$ is said to be positive definite if it is symmetric and

$$\langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle > 0 \quad \text{for all } \mathbf{u} \neq \mathbf{0}. \tag{4.3.15}$$

Exercise 4.3.22. Prove that (a) a symmetric matrix is positive definite if and only if all eigenvalues are positive.
(b) All principal submatrices of a positive definite matrix are positive definite (cf. Exercise 19).
(c) The diagonal elements $a_{\alpha\alpha}$ of a positive definite matrix are positive.
(d) $\mathbf{A}$ is called positive semi-definite if the inequality (15) holds with "$\ge$" instead of "$>$". A positive semidefinite matrix $\mathbf{A}$ has a unique positive semi-definite square root $\mathbf{B} = \mathbf{A}^{1/2}$, which has the property $\mathbf{B}^2 = \mathbf{A}$. If $\mathbf{A}$ is positive definite, then so is $\mathbf{A}^{1/2}$.

A corollary to Exercise 22a is

Lemma 4.3.23. *A positive definite matrix* $\mathbf{A}$ *is nonsingular and has a positive definite inverse.*

The property "$\mathbf{A}^{-1}$ is positive definite" is neither necessary nor sufficient to ensure the property "$\mathbf{A}^{-1} \ge 0$" of an M-matrix. In both cases, however, (irreducible) diagonal dominance is a sufficient criterion (cf. Criterion 10).

Criterion 4.3.24. *If a symmetric matrix with positive diagonal entries is diagonally dominant or irreducibly diagonally dominant then it is positive definite.*

PROOF. Since $r_\alpha < a_{\alpha\alpha}$, resp. $r_\alpha \le a_{\alpha\alpha}$, the Gershgorin circles which occur in Criterion 4 do not intersect the semi-axis $(-\infty, 0]$, so that all the eigenvalues must be positive. By Exercise 22a then $\mathbf{A}$ is positive definite. ∎

Lemma 4.3.25. *Let* $\lambda_{\min}$ *and* $\lambda_{\max}$ *be respectively the smallest and largest eigenvalues of a positive definite matrix* $\mathbf{A}$. *Then there holds*

$$\|\mathbf{A}\|_2 = \lambda_{\max}, \qquad \|\mathbf{A}^{-1}\|_2 = 1/\lambda_{\min}. \tag{4.3.16}$$

PROOF. Exercise 20a shows that $\|\mathbf{A}\|_2 = \rho(\mathbf{A})$ and $\|\mathbf{A}^{-1}\|_2 = \rho(\mathbf{A}^{-1})$. From (5) then result $\rho(\mathbf{A}) = \lambda_{\max}$ and $\rho(\mathbf{A}^{-1}) = 1/\lambda_{\min}$, since $\lambda_{\min} > 0$. ∎

4.4 Properties of the Matrix L_h

Theorem 4.4.1. *The matrix* L_h *(five-point formula) defined in* (2.5) *has the following properties:*

$$L_h \text{ is an } M\text{-matrix}, \tag{4.4.1a}$$

$$L_h \text{ is positive definite}, \tag{4.4.1b}$$

$$\|L_h\|_\infty \le 8h^{-2}, \quad \|L_h^{-1}\|_\infty \le 1/8, \tag{4.4.1c}$$

$$\|L_h\|_2 \le 8h^{-2}\cos^2(\pi h/2) < 8h^{-2}, \tag{4.4.1d}$$

$$\|L_h^{-1}\|_2 \le \frac{1}{8}h^2 \sin^{-2}(\frac{\pi h}{2}) = \frac{1}{2\pi^2} + O(h^2) \le \frac{1}{16}. \tag{4.4.1e}$$

PROOF. (a) In Section 4.3 we already noticed that L_h is irreducibly diagonally dominant and satisfies the inequality (3.1a). By Criterion 4.3.10 then L_h is an M-matrix.
(b) Since L_h is symmetric and irreducibly diagonally dominant, (1b) follows from Criterion 4.3.24.
(c) That $\|L_h\|_\infty \le 8h^{-2}$ can be read from (2.5) and (3.11). To estimate L_h^{-1} one uses Theorem 4.3.16 with $w(x,y) = x(1-x)/2$. Then we have $L_h w \ge \mathbb{1}$ (even that $(L_h w)(x,y) = 1$ unless $y = h$ and $y = 1 - h$) and $\|w\|_\infty \le w(1/2, y) = 1/8$.
(d) The inequalities (1d,e) result from Lemma 4.3.25 and

Lemma 4.4.2. *The* $(n-1)^2$ *eigenvectors of* L_h *are* $u^{\nu\mu}$ $(1 \le \nu, \mu \le n-1)$*:*

$$u^{\nu\mu}(x,y) = \sin(\nu\pi x)\sin(\mu\pi y) \qquad ((x,y) \in \Omega_h). \tag{4.4.2a}$$

The corresponding eigenvalues are

$$\lambda_{\nu\mu} = 4h^{-2}(\sin^2(\nu\pi h/2) + \sin^2(\mu\pi h/2)), \qquad 1 \le \nu, \mu \le n-1. \tag{4.4.2b}$$

PROOF. Let Ω_h^{1D} be the one-dimensional grid (1.6a) and let $u^\nu(x) := \sin(\nu\pi x)$. For each $x \in \Omega_h^{1D}$ there holds

$$\begin{aligned}\partial^-\partial^+ u^\nu(x) &= h^{-2}[\sin(\nu\pi(x-h)) + \sin(\nu\pi(x+h)) - 2\sin(\nu\pi x)] \\ &= 2h^{-2}\sin(\nu\pi x)[\cos(\nu\pi h) - 1]\end{aligned}$$

since $\sin(\nu\pi(x \pm h)) = \sin(\nu\pi x)\cos(\nu\pi h) \pm \cos(\nu\pi x)\sin(\nu\pi h)$. The identity $1 - \cos\xi = 2\sin^2(\xi/2)$ then implies

$$-\partial^-\partial^+u^\nu(x) = 4h^{-2}\sin^2(\nu\pi h/2)u^\nu(x), \qquad x \in \Omega_h^{1D}. \tag{4.4.2c}$$

Let L_h^{1D} be the matrix (1.9b). Note that $(\partial^-\partial^+u)(h)$ also involves the boundary value $u(0)$ which $(L_h^{1D}u)(h)$ does not; similarly $(\partial^-\partial^+u)(1-h)$ depends on $u(1)$. However, since $u(0) = \sin(0) = 0$ and $u(1) = \sin(\nu\pi) = 0$ we have $L_h^{1D}u^\nu = -\partial^-\partial^+u^\nu$, and (2c) can be brought over:

$$L_h^{1D}u^\nu = 4h^{-2}\sin^2(\nu\pi h/2)u^\nu, \qquad 1 \le \nu \le n-1. \tag{4.4.2c'}$$

The two-dimensional grid function $u^{\nu\mu}$ in (2a) can be written as the (tensor) product $u^\nu(x)u^\mu(y)$. Now we have that $(L_hu^{\nu\mu})(x,y)$ is equal to the sum $u^\mu(y)(L_h^{1D}u^\nu)(x) + u^\nu(x)(L_h^{1D}u^\mu)(y)$, so that (2b) follows from (2c'). ■

In the sequel we want to show the analogies between the properties of the Poisson equation (2.2) and the discrete five-point formula (2.4a,b).

The analogue of the mean-value property (2.3.1) is the equation

$$u_h(x,y) = \frac{1}{4}[u_h(x-h,y)+u_h(x+h,y)+u_h(x,y-h)+u_h(x,y+h)]. \tag{4.4.3}$$

From (2.3) and (2.4a) with $f = 0$ we obtain the

Remark 4.4.3. The solution u_h of the discrete potential equation (2.4a) with $f = 0$ satisfies Equation (3) at all grid points $(x,y) \in \Omega_h$.

As in the continuous case the mean-value property (3) implies the maximum-minimum principle.

Remark 4.4.4. Let u_h be a non-constant solution of the discrete potential equation (2.4a) with $f = 0$. The extrema $\max\{u_h(\mathbf{x}): \mathbf{x} \in \overline{\Omega}_h\}$ and $\min\{u_h(\mathbf{x}): \mathbf{x} \in \overline{\Omega}_h\}$ are assumed not on Ω_h but on Γ_h.

PROOF. If u_h were maximal in $(x,y) \in \Omega_h$, then because of Equation (3), all neighbouring points $(x \pm h, y)$ and $(x, y \pm h)$ would have to carry the same values. Since every pair of points can be linked by a chain of neighbouring points, it follows that $u_h = \text{const}$, in contradiction to the assumption. ■

The last proof indirectly uses the fact that L_h is irreducible. The irreducibility of L_h corresponds to the assumption in Theorem 2.3.7 that Ω is a domain, i.e., connected.

The result of carrying over Theorems 2.4.3 and 3.1.2 reads as follows:

Theorem 4.4.5. *Let u_h^1 and u_h^2 be two solutions of* (2.4a): $-\Delta_h u_h^i = f$ *for different boundary values $u_h^i = \varphi^i$ $(i = 1,2)$. Then the following holds:*

$$\|u_h^1 - u_h^2\|_\infty \le \max_{\mathbf{x}\in\Gamma_h} |\varphi^1(\mathbf{x}) - \varphi^2(\mathbf{x})|, \tag{4.4.4a}$$

$$u_h^1 \le u_h^2 \quad \text{in } \Omega_h, \quad \text{if } \varphi^1 \le \varphi^2 \quad \text{on } \Gamma_h. \tag{4.4.4b}$$

PROOF. Let $w_h := u_h^2 - u_h^1$. (1) In the case that $\varphi^1 \le \varphi^2$ one has $w_h \ge 0$ on Γ_h and $-\Delta_h w_h = 0$. Remark 4 proves that $w_h = \text{const} \ge 0$ or $w_h > 0$.
(2) Let M be the right side of (4a). $-M \le w_h \le M$ on Γ_h implies the inequalities $-M \le w_h \le M$ on Ω_h. ■

The discrete analogue of the Green function $g(\mathbf{x},\xi)$ is $h^{-2}L_h^{-1}$. Let δ_ξ be the scaled unit vector

$$\delta_\xi(\mathbf{x}) = \begin{cases} h^{-2} & \text{if } \mathbf{x} = \xi \\ 0 & \text{if } \mathbf{x} \ne \xi \end{cases} \qquad (\mathbf{x}, \xi \in \Omega_h). \tag{4.4.5a}$$

The column of the matrix $h^{-2}L_h^{-1}$ with index $\xi \in \Omega_h$ is given by

$$g_h(\cdot,\xi) := L_h^{-1}\delta_\xi \quad (\xi \in \Omega_h). \tag{4.4.5b}$$

For $\xi \in \Omega_h$ fixed, $g_h(\cdot,\xi)$ is a grid function defined on Ω_h. The domain of definition is extended to $\overline{\Omega}_h \times \overline{\Omega}_h$:

$$\overline{g}_h(\mathbf{x},\xi) = \begin{cases} g_h(\mathbf{x},\xi) & \text{for } \mathbf{x},\xi \in \Omega_h \\ 0 & \text{for } \mathbf{x} \in \Gamma_h \text{ or } \xi \in \Gamma_h \end{cases} \qquad (\mathbf{x},\xi \in \overline{\Omega}_h). \tag{4.4.5c}$$

The $g_h(\mathbf{x},\xi)$ are entries of $h^{-2}L_h^{-1}$: $g_h(\mathbf{x},\xi) = h^{-2}(L_h^{-1})_{\mathbf{x}\xi}$. The symmetry of L_h implies

Remark 4.4.6. $\overline{g}_h(\mathbf{x},\xi) = \overline{g}_h(\xi,\mathbf{x})$ for all $\mathbf{x},\xi \in \overline{\Omega}_h$ (cf. (3.2.3)).

The representation (3.2.7) is recalled by:

Remark 4.4.7. The solution u_h of the system of equations (2.4a) with boundary values $\varphi = 0$ reads

$$u_h(\mathbf{x}) = h^2 \sum_{\xi\in\Omega_h} \overline{g}_h(\xi,\mathbf{x})f(\xi), \quad \mathbf{x} \in \overline{\Omega}_h. \tag{4.4.6}$$

Equation (6) is the component-wise representation of the equation $u_h = L_h^{-1}f_h$. The factor h^2 compensates for h^{-2} in (5a). It was introduced so that the summation $h^2\sum$ in (6) approximates the integral $\int_\Omega$. Since in Section 3.2 we consider the Poisson equation $\Delta u = f$, but here the equation $-\Delta u = f$, the right-hand sides in (3.2.7) and (6) differ in their signs.

The discrete Green function is positive also (cf. (3.2.9)):

Remark 4.4.8. $0 < g_h(\mathbf{x},\xi) \le h^{-2}/8$ for $\mathbf{x},\xi \in \Omega_h$.

PROOF. The upper bound follows from

$$g_h(\mathbf{x},\xi) \le \|g_h(\cdot,\xi)\|_\infty \le \|L_h^{-1}\|_\infty \|\delta_\xi\|_\infty,$$

$\|\delta_\xi\|_\infty = h^{-2}$, and (1c). $g_h > 0$ can be inferred from $L_h^{-1} > \mathbf{0}$ (cf. Theorem 4.3.11). ■

The bound $g_h(\mathbf{x},\xi) \le h^{-2}/8$ is too pessimistic and can be improved considerably.

Remark 4.4.8′.

$$0 < g_h(\mathbf{x},\xi) \le \frac{1}{4} - \frac{\log(|\mathbf{x}-\xi|^2 + 2h^2)}{\log 16} \le \frac{|\log h|}{\log 4}. \tag{4.4.7}$$

This estimate via $O(|\log h|)$ reflects the logarithmic singularity of the singularity function $s(\mathbf{x},\xi) = -\log(|\mathbf{x}-\xi|^2)/4\pi$.

Exercise 4.4.9. For the proof of inequality (7) define

$$s_h(\mathbf{x},\xi) := 1/4 - \log(|\mathbf{x}-\xi|^2 + 2h^2)/\log 16, \quad u_h(\mathbf{x}) := s_h(\mathbf{x},\mathbf{0})$$

and carry out the following steps:
(a) $s_h(\mathbf{x},\xi) \ge 0$ for all $\mathbf{x} \in \overline{\Omega}_h$, $\xi \in \Omega_h$,
(b) $-\Delta_h u_h(\mathbf{0}) = h^{-2}$,
(c) $-\Delta_h u_h(\mathbf{x}) \ge 0$ for all $\mathbf{x} \in \mathbb{R}^2$ (longer calculation !),
(d) $-\Delta_h(s_h(\cdot,\xi) - g_h(\cdot,\xi)) \ge 0$ for fixed $\xi \in \Omega_h$,
(e) $g_h(\mathbf{x},\xi) \le s_h(\mathbf{x},\xi)$.

Let φ_h be defined as in (2.6b). The solution of the discrete potential equation

$$-\Delta_h u_h = 0 \quad \text{in } \Omega_h, \quad u_h = \varphi \quad \text{on } \Gamma_h$$

is given by $u_h := L_h^{-1}\varphi_h$ (if one continues the grid function, at first only defined on Ω_h, through φ on Γ_h; cf. Remark 4.2.5). The representation with the aid of g_h reads

$$u_h(\mathbf{x}) = h^2 \sum_{\xi\in\Omega_h} g_h(\xi,\mathbf{x})\varphi_h(\xi), \quad \mathbf{x} \in \Omega_h.$$

Since $\varphi_h(\mathbf{x})$ vanishes at all points $\mathbf{x}$ far from the boundary, it suffices to extend the sum over the points near the boundary. If one sums over the neighbouring boundary points of

$$\Gamma_h' := \{\xi \in \Gamma_h : \xi \text{ is not a corner point } (0,0), (1,0), (0,1), \text{ or } (1,1)\}$$

instead of over the points near the boundary, then the definition of φ_h results in the representation

$$u_h(\mathbf{x}) = -h \sum_{\xi\in\Gamma_h'} \partial_n^- \overline{g}_h(\xi,\mathbf{x})\varphi(\xi), \quad \mathbf{x} \in \Omega_h, \tag{4.4.8}$$

where ∂_n^- is the backward difference with respect to the direction of the normal $\mathbf{n}$:

$$\partial_n^- \overline{g}_h(\xi,\mathbf{x}) = h^{-1}[\overline{g}_h(\xi,\mathbf{x}) - \overline{g}_h(\xi - h\mathbf{n},\mathbf{x})]$$

(note that $\overline{g}_h(\xi,\mathbf{x}) = 0$ for $\xi \in \Gamma_h' \subset \Gamma_h$). The variable $\xi - h\mathbf{n}$ ranges over all points next to the boundary.

Remark 4.4.10. Equation (8) corresponds to the representation in Theorem 3.2.10. The summation $h\sum_{\xi\in\Gamma_h'}$ approximates the integral $\int_\Gamma$.

Finally we want to take a closer look at the estimate for the solution $u_h = L_h^{-1}q_h$ through $\|u_h\|_\infty \le \|L_h^{-1}\|_\infty \|q_h\|_\infty \le \|q_h\|_\infty/8$ (cf. (1c)). According to (2.6b), $q_h = f_h + \varphi_h$ contains the right-hand side $f_h(\mathbf{x}) = f(\mathbf{x})$ of the discrete Poisson equation and the boundary values $\varphi(\mathbf{x})$, $\mathbf{x} \in \Gamma_h$ hidden in φ_h. The following theorem gives a bound in which these components are separated.

Theorem 4.4.11. *According to* (2.6b), *let $q_h = f_h + \varphi_h$ be constructed from f and φ. The discrete solution $u_h = L_h^{-1}q_h$ of the Poisson boundary value problem can be bounded by*

$$\|u_h\|_\infty \le \|L_h^{-1}\|_\infty \max_{\mathbf{x}\in\Omega_h} |f(\mathbf{x})| + \max_{\mathbf{x}\in\Gamma_h'} |\varphi(\mathbf{x})| \le \frac{1}{8}\max_{\mathbf{x}\in\Omega_h} |f(\mathbf{x})| + \max_{\mathbf{x}\in\Gamma_h'} |\varphi(\mathbf{x})|. \tag{4.4.9}$$

PROOF. Set $u_h' := L_h^{-1}f_h$ and $u_h'' := L_h^{-1}\varphi_h$. The estimate for the first summand in $u_h = u_h' + u_h''$ results in the first summand in (9). To bound u_h'' use the inequality (4a) with $u_h^1 = u_h''$, $\varphi^1 = \varphi$ and $u_h^2 = 0$, $\varphi^2 = 0$. ■

The corresponding inequality

$$\|u\|_\infty \le C\|f\|_\infty + \|\varphi\|_\infty \tag{4.4.10}$$

for the solution of the boundary value problem (2.2a,b) has not been mentioned until now, but will be proved in a more general context in Section 5.1.3. The maximum norm $\|\cdot\|_\infty$ in (9) can be replaced by the Euclidean norms

$$\|u_h\|_{2,\Omega_h} := \sqrt{h^2 \sum_{\mathbf{x}\in\Omega_h} |u_h(\mathbf{x})|^2}, \quad \|\varphi\|_{2,\Gamma_h'} := \sqrt{h \sum_{\mathbf{x}\in\Gamma_h'} |\varphi(\mathbf{x})|^2}.$$

Here, $h^2\sum_{\Omega_h}$ and $\int_\Omega$, $h\sum_{\Gamma_h'}$ and $\int_\Gamma$ correspond to each other.

Theorem 4.4.12. *Under the assumptions of Theorem 11 there holds*

$$\|u_h\|_{2,\Omega_h} \le \|L_h^{-1}\|_2 \|f\|_{2,\Omega_h} + \frac{1}{\sqrt{2}}\|\varphi\|_{2,\Gamma_h'} \le \frac{1}{16}\|f\|_{2,\Omega_h} + \frac{1}{\sqrt{2}}\|\varphi\|_{2,\Gamma_h'}. \tag{4.4.11}$$

PROOF. (1) It suffices to consider the case of the potential equation (i.e., $f = 0$). Let the restriction of φ on Γ_h' result in the grid function Φ_h: $\Phi_h(\mathbf{x}) = \varphi(\mathbf{x})$

for $\mathbf{x} \in \Gamma'_h$. Let the mapping $\Phi_h \mapsto u_h = L_h^{-1}\varphi_h$ be given by the rectangular matrix $\mathbf{A}$: $u_h = \mathbf{A}\varphi_h$. According to Equation (8) the entries of $\mathbf{A}$ read

$$a_{\mathbf{x}\xi} = -h\partial_n^- g_h(\xi, \mathbf{x}) = g_h(\xi - h\mathbf{n}, \mathbf{x}) = g_h(\mathbf{x}, \xi - h\mathbf{n}), \quad \mathbf{x} \in \Omega_h,\ \xi \in \Gamma'_h.$$

Since $\mathbf{A} \geq \mathbf{0}$ (cf. Remark 8), one obtains the row sum norm $||\mathbf{A}||_\infty := \max_{\mathbf{x}} \sum_\xi a_{\mathbf{x}\xi}$ as $||\mathbf{A}\varphi_h||_\infty$ for the choice of $\varphi_h(\mathbf{x}) = 1$ in all $\mathbf{x} \in \Gamma'_h$. The solution $v_h = \mathbf{A}\varphi_h$ then reads $v_h = \mathbb{1}$ (why?), so it follows that

$$||\mathbf{A}||_\infty = ||\mathbf{A}\varphi_h||_\infty = ||\mathbb{1}||_\infty = 1.$$

(2) The column sums of $\mathbf{A}$ are $s(\xi) := \sum_{\mathbf{x}\in\Omega_h} a_{\mathbf{x}\xi} = \sum_{\mathbf{x}\in\Omega_h} g_h(\mathbf{x}, \xi - h\mathbf{n})$ for $\xi \in \Gamma'_n$. The grid function $v_h := h^{-2}L_h^{-1}\mathbb{1}$ at the points $\xi - h\mathbf{n}$ near the boundary has the values

$$s(\xi) = v_h(\xi - h\mathbf{n}), \quad \xi \in \Gamma'_h,\ \mathbf{n} \text{ normal direction},$$

as is implied by Remark 7. Let $\xi = (\xi_1, \xi_2) \in \Gamma'_h$ be a point of the left or right boundary (i.e., $\xi_1 = 0$ or 1). As mentioned in the proof of (1c), $L_h w_h \geq \mathbb{1}$ holds for $w_h(x, y) := x(1-x)/2$. Since $L_h^{-1} \geq \mathbf{0}$, then so is $w_h \geq L_h^{-1}\mathbb{1}$ and hence $v_h \leq h^{-2}w_h$. In particular we have the following estimate

$$s(\xi) = v_h(\xi - h\mathbf{n}) \leq h^{-2}w_h(\xi - h\mathbf{n}) = h^{-2}h(1-h)/2 \leq h^{-1}/2$$

at the point near the boundary $\xi - h\mathbf{n} = (h, \xi_2)$ or $(1-h, \xi_2)$. For a point ξ from the upper or from the lower boundary one obtains the same estimate if one uses $w_h(x, y) := y(1-y)/2$. Since the column sums $s(\xi)$ are the row sums of $\mathbf{A}^\mathsf{T}$, we have proved

$$||\mathbf{A}^\mathsf{T}||_\infty = \max\{s(\xi) : \xi \in \Gamma'_h\} \leq h^{-1}/2.$$

(3) We have $||\mathbf{A}^\mathsf{T}\mathbf{A}||_2 = \rho(\mathbf{A}^\mathsf{T}\mathbf{A}) \leq ||\mathbf{A}^\mathsf{T}\mathbf{A}||_\infty \leq ||\mathbf{A}^\mathsf{T}||_\infty ||\mathbf{A}||_\infty \leq h^{-1}/2$ (cf. Exercise 4.3.20a, (3.10d), (3.10a)) so that the solution $u_h = L_h^{-1}\varphi_h = \mathbf{A}\Phi_h$ satisfies the following estimate

$$\begin{aligned} ||u_h||^2_{2,\Omega_h} &= h^2 \sum_{\mathbf{x}\in\Omega_h} u^2(\mathbf{x}) = h^2 \sum_{\mathbf{x}\in\Omega_h} |(\mathbf{A}\Phi_h)(\mathbf{x})|^2 \\ &= h^2 \sum_{\mathbf{x}\in\Gamma'_h} \Phi_h(\mathbf{x})(\mathbf{A}^\mathsf{T}\mathbf{A}\Phi_h)(\mathbf{x}) \\ &\leq h^2 ||\mathbf{A}^\mathsf{T}\mathbf{A}||_2 \sum_{\mathbf{x}\in\Gamma'_h} \Phi_h^2(\mathbf{x}) \\ &\leq h/2 \sum_{\mathbf{x}\in\Gamma'_h} \Phi_h^2(\mathbf{x}) = \frac{1}{2}||\Phi_h||^2_{2,\Gamma'_h}. \end{aligned}$$

Since $||\varphi||_{2,\Gamma'_h} = ||\Phi_h||_{2,\Gamma'_h}$, the assertion follows. ■

4.5 Convergence

Let U_h be the vector space of the grid functions on $\overline{\Omega}_h$. The discrete solution $u_h \in U_h$ and the continuous solution $u \in C^0(\overline{\Omega})$ cannot be compared directly because of the different domains of definition. In difference methods it is customary to compare both functions on the grid $\overline{\Omega}_h$. To this end one must map the solution u by means of a "restriction"

$$u \mapsto R_h u \in U_h \quad (u: \text{continuous solution}) \tag{4.5.1}$$

to U_h. In the following we choose R_h as restriction on $\overline{\Omega}_h$:

$$(R_h u)(\mathbf{x}) = u(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \overline{\Omega}_h. \tag{4.5.2}$$

The limit $h \to 0$ is made precise as follows. Let $H \subset \mathbb{R}_+$ be a subset with accumulation point zero: $0 \in \overline{H}$. For example, the step sizes considered so far form the set $H = \{1/n : n \in \mathbb{N}\}$. For each $h \in H$ let U_h be equipped with the norm $||\cdot||_h$.

Definition 4.5.1. (Convergence) The discrete solutions $u_h \in U_h$ converge (with respect to the system of norms $||\cdot||_h, h \in H$) to u if

$$||u_h - R_h u||_h \to 0. \tag{4.5.3a}$$

We have convergence of order k if

$$||u_h - R_h u||_h = O(h^k). \tag{4.5.3b}$$

The proof of convergence is usually carried out with the aid of the concepts of "stability" and "consistency". The discretisation $\{L_h : h \in H\}$ is said to be stable with respect to $||\cdot||_\infty$ if

$$\sup_{h\in H} ||L_h^{-1}||_\infty < \infty. \tag{4.5.4}$$

For the discretisation defined in Section 4.2, the stability has been proved in (4.1c) with respect to the row sum norm, and in (4.1d) with respect to the spectral norm.

The grid function f_h in $-\Delta_h u_h = f_h$ is the restriction

$$f_h = \tilde{R}_h f \tag{4.5.5a}$$

of f, where in (2.6b) $\tilde{R}_h$ was chosen as the restriction to Ω_h:

$$(\tilde{R}_h f)(\mathbf{x}) = f(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Omega_h. \tag{4.5.5b}$$

The formulation of consistency becomes simpler if instead of the matrix L_h one considers the difference operator $-\Delta_h$ on which it is based (cf. Remark 4.2.5). Let D_h be a general difference operator, and let $-D_h u_h = f_h$ discretise $-\Delta u = f$. The discretisation described by the triple $(D_h, R_h, \tilde{R}_h)$ is said to be consistent of order k with respect to $\|\cdot\|_\infty$ (consistent with the Laplace operator), if

$$\|D_h R_h u - \tilde{R}_h \Delta u\|_\infty \le K h^k \|u\|_{C^{k+2}(\bar{\Omega})} \tag{4.5.6}$$

for all $u \in C^{k+2}(\bar{\Omega})$. Here K is independent of h and u.

Remark 4.5.2. Let R_h and $\tilde{R}_h$ be given by (2) and (5b). The five-point formula Δ_h is consistent of order 2: estimate (6) holds with $k = 2$ and $K = 1/6$.

PROOF. The expansion (1.4c) can be applied in the x and y directions and yields

$$\Delta_h R_h u(x,y) = \Delta u(x,y) + h^2(R_x + R_y) \quad \text{with } |R_x|, |R_y| \le (1/12)\|u\|_{C^4(\bar{\Omega})}. \tag{4.5.7}$$

■

Theorem 4.5.3. *Let the discretisation $(D_h, R_h, \tilde{R}_h)$ be consistent of order k. Let the matrix L_h associated to the difference operator D_h (cf. Remark 4.2.5) be stable with respect to $\|\cdot\|_\infty$. Then the method is convergent of order k if $u \in C^{k+2}(\bar{\Omega})$.*

PROOF. $w_h := u_h - R_h u$ satisfies the difference equations

$$\begin{aligned} -D_h w_h &= -D_h u_h + D_h R_h u = f_h + D_h R_h u \\ &= \tilde{R}_h f + D_h R_h u = D_h R_h u - \tilde{R}_h \Delta u \quad \text{in } \Omega_h. \end{aligned}$$

Because of the homogeneous boundary condition $w_h = 0$ on Γ_h we have that $w_h = L_h^{-1}(D_h R_h u - \tilde{R}_h \Delta u)$ (cf. Remark 4.2.1). (4) and (6) imply (3b). ■

If one substitutes the concrete values $\|L_h^{-1}\|_\infty \le 1/8$ and $K = 1/6$ one obtains

Corollary 4.5.4. *Let the continuous solution of the boundary value problem* (2.2a,b) *belong to $C^4(\bar{\Omega})$. Let u_h be the discrete solution defined in* (2.4a,b). *Then the convergence of u_h to u is of second order:*

$$\|u_h - R_h u\|_\infty \le \frac{h^2}{48}\|u\|_{C^4(\bar{\Omega})}. \tag{4.5.8}$$

The assumption $u \in C^4(\bar{\Omega})$ can be weakened to

Corollary 4.5.5. *Under the condition $u \in C^{3,1}(\bar{\Omega})$ we also have*

$$\|u_h - R_h u\|_\infty \le (h^2/48)\|u\|_{C^{3,1}(\bar{\Omega})}. \tag{4.5.8'}$$

PROOF. The remainder term R_4 in (1.5b) may also be written as

$$R_4 = h^{-4}\int_x^{x\pm h} [u'''(\xi) - u'''(x)](x \pm h - \xi)^2/2!\,d\xi.$$

The Lipschitz estimate $|u'''(\xi) - u'''(x)| \le |\xi - x|\,\|u\|_{C^{3,1}(\bar{\Omega})}$ implies $R_4 \le \|u\|_{C^{3,1}(\bar{\Omega})}/4!$ so that in (7), (6), and (8) the norm $\|u\|_{C^4(\bar{\Omega})}$ can be replaced by $\|u\|_{C^{3,1}(\bar{\Omega})}$. ■

If, however, one further weakens $u \in C^{3,1}(\bar{\Omega})$ to $u \in C^s(\bar{\Omega})$, $2 < s < 4$, one obtains a weaker order of convergence.

Corollary 4.5.6. *Under the condition $u \in C^s(\bar{\Omega})$, $2 < s \le 4$, u_h converges of order $s-2$:*

$$\|u_h - R_h u\|_\infty \le K_s h^{s-2}\|u\|_{C^s(\bar{\Omega})}, \tag{4.5.8''}$$

where $K_s := 1/[2s(s-1)]$ for $2 \le s \le 3$ and $K_s := 1/[2s(s-1)(s-2)]$ for $3 \le s \le 4$.

The proof results from

Exercise 4.5.7. Show $\|\Delta_h R_h u - \tilde{R}_h \Delta u\|_\infty \le 8K_s h^{s-2}\|u\|_{C^s(\bar{\Omega})}$, $2 < s \le 4$.

Even though the proofs of convergence are simple, the results remain unsatisfactory. As can be seen from Example 2.1.3, the continuous solution of the boundary value problem (2.2a,b) generally does not even satisfy $u \in C^2(\bar{\Omega})$, although one needs at least $u \in C^s(\bar{\Omega})$ with $s > 2$ in Corollary 6 for convergence. Stronger results can be obtained by an analysis which will be discussed in Section 9.2. That errors of the order of magnitude of $O(h^2)$ occur even under weaker conditions, is shown in

Example 4.5.8. If one solves the difference equation (2.4a, b) for $-\Delta u = 1$ in $\Omega = (0,1) \times (0,1)$ and $u = 0$ on Γ, one obtains at the centre $x = y = 1/2$ the values $u_h(\frac{1}{2}, \frac{1}{2})$, which are shown in the first column of Table 1. The exact solution $u(\frac{1}{2}, \frac{1}{2}) = 0.0736713\cdots$ results from a representation that one can find in Example 8.1.11. The quotients ϵ_h/ϵ_{2h} of the errors $\epsilon_h = u(\frac{1}{2}, \frac{1}{2}) - u_h(\frac{1}{2}, \frac{1}{2})$ approximate $1/4$. This proves $u_h(\frac{1}{2}, \frac{1}{2}) = u(\frac{1}{2}, \frac{1}{2}) + O(h^2)$, although $u \notin C^2(\bar{\Omega})$. At $(\frac{1}{2}, \frac{1}{2})$ u_h has furthermore the asymptotic expansion $u_h(\frac{1}{2}, \frac{1}{2}) = u(\frac{1}{2}, \frac{1}{2}) + h^2 e(\frac{1}{2}, \frac{1}{2}) + O(h^4)$. The error term $e(\frac{1}{2}, \frac{1}{2})$ independent of h, is eliminated by using the Richardson extrapolation

$$u_{h,2h}\left(\frac{1}{2}, \frac{1}{2}\right) := \frac{1}{3}\left[4u_h\left(\frac{1}{2}, \frac{1}{2}\right) - u_{2h}\left(\frac{1}{2}, \frac{1}{2}\right)\right].$$

The extrapolated values are already very accurate for $h = 1/16$ (cf. the last column of Table 1).

Table 4.5.1 Difference solutions for Example 8

h	$u_h(\frac{1}{2},\frac{1}{2})$	$u_{h,2h}(\frac{1}{2},\frac{1}{2})$	$\epsilon_h = u_h(\frac{1}{2},\frac{1}{2}) - u_h(\frac{1}{2},\frac{1}{2})$	Quotient	$\epsilon_{h,2h}$	Quotient
1/8	0.0727826		8.89×10^{-4}			
1/16	0.07344576	0.0736668	2.26×10^{-4}	0.250	4.5×10^{-6}	
1/32	0.0736147373	0.07367106	5.66×10^{-5}	0.251	2.7×10^{-7}	0.06
1/64	0.0736571855	0.07367133	1.41×10^{-5}	0.249		

4.6 Discretisations of Higher Order

The five-point formula (2.11) is of second order. Even if the solution u belongs to $C^s(\overline{\Omega})$ with $s > 4$, no better bound for $\Delta_h R_h u - \tilde{R}_h \Delta u$ than $O(h^2)$ would result. An obvious method for constructing difference methods of higher order is the following. As an ansatz for the discretisation of the one-dimensional equation $u'' = f$ choose

$$(D_h u_h)(x) = h^{-2} \sum_{\nu=-k}^{k} c_\nu u_h(x + \nu h).$$

The Taylor expansion provides

$$(D_h R_h u)(x) = \sum_{\mu=0}^{2k} a_\mu h^{\mu-2} u^{\mu}(x) + O(h^{2k-1}), \quad a_\mu = \Big(\sum_{\nu=-k}^{k} c_\nu \nu^\mu\Big)/\mu!.$$

The $2k+1$ equations $a_0 = a_1 = a_3 = a_4 = \cdots = a_{2k} = 0$ and $a_2 = 1$ form a linear system for the $2k+1$ unknown coefficients c_ν. For $k = 1$ one obtains the usual difference formula (1.4c); for $k = 2$ a difference of fourth order results:

$$h^2(D_h u_h)(x) = \frac{-1}{12}[u_h(x-2h)+u_h(x+2h)]+\frac{4}{3}[u_h(x-h)+u_h(x+h)]-\frac{5}{2}u_h(x).$$

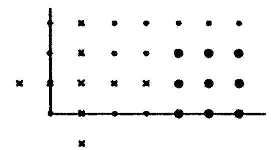
If one applies this approximation for u'' to the x and y coordinates, one obtains for $-\Delta$ the difference star

$$\frac{h^{-2}}{12}\begin{bmatrix} & & 1 & & \\ & & -16 & & \\ 1 & -16 & 60 & -16 & 1 \\ & & -16 & & \\ & & 1 & & \end{bmatrix} \tag{4.6.1}$$

(cf. (2.12)). The difference scheme (1) is of fourth order, but presents difficulties at points near the boundary. To set up the difference formula at $(h,h) \in \Omega_h$, for example, one needs the values $u_h(-h,h)$ and $u_h(h,-h)$ outside $\overline{\Omega}_h$ (cf. Figure 1). One possibility would be to use scheme (1) only at points far from the boundary and to use the five-point formula (2.11) at points near the boundary.

The above complications do not occur if one limits oneself to **compact nine-point formulae**; by this one means difference methods (2.12) which are characterised by "$c_{\alpha\beta} \neq 0$ only for $-1 \le \alpha, \beta \le 1$" (cf. Fig. 1). An ansatz with the nine free parameters $c_{\alpha\beta}, -1 \le \alpha, \beta \le 1$, leads, however, to a negative result: there is no compact nine-point formula with $(D_h u)(\mathbf{x}) = -\Delta u(\mathbf{x}) + O(h^3)$. In this sense the five-point formula is indeed optimal.

Nevertheless, nine-point procedures of fourth order can be obtained if one also selects the right side f_h of the system of equations (2.6a,b) in a suitable manner.



×: Difference Scheme (1)
•: Compact nine-point formula

Figure 4.6.1. Scheme (1) and the compact nine-point formula

If one applies the compact nine-point scheme

$$D_h := -\frac{h^{-2}}{6}\begin{bmatrix} -1 & -4 & -1 \\ -4 & 20 & -4 \\ -1 & -4 & -1 \end{bmatrix} \tag{4.6.2}$$

to $u \in C^8(\overline{\Omega})$, the Taylor expansion results in

$$-D_h u = -\Delta u - \frac{h^2}{12}\Delta^2 u - \frac{h^4}{360}\left[\frac{\partial^4}{\partial x^4} + 4\frac{\partial^4}{\partial x^2 \partial y^2} + \frac{\partial^4}{\partial y^4}\right]\Delta u + O(h^6). \tag{4.6.3}$$

Here it is crucial that the error term can be expressed by $-\Delta u$ and hence by f.

For the special choice of the restriction $\tilde{R}_h$ via

$$f_h = \tilde{R}_h f := \frac{1}{6}\begin{bmatrix} & 1/2 & \\ 1/2 & 4 & 1/2 \\ & 1/2 & \end{bmatrix} f, \tag{4.6.4}$$

i.e.,

$$\begin{aligned} f_h(x,y) =& (\tilde{R}_h f)(x,y) \\ :=& \frac{1}{12}[f(x-h,y) + f(x+h,y) + \\ & f(x,y-h) + f(x,y+h) + 8f(x,y)], \end{aligned} \tag{4.6.4'}$$

one obtains the expansion

$$f_h(x,y) = f(x,y) + \frac{h^2}{12}\Delta f(x,y) + O(h^4), \tag{4.6.5}$$

which, because $f = -\Delta u$, agrees with (3) up to $O(h^4)$.

The matrix L_h of the system of equations which results after the elimination of the boundary values $u_h(x) = \varphi(x)$, $x \in \Gamma_h$, has the entries

$$L_{\mathbf{x}\xi} = h^{-2}\begin{cases} 20/6 & \text{if } \mathbf{x} = \xi, \\ -1/6 & \text{if } \mathbf{x} - \xi = (\pm h, \pm h) \text{ or } (\pm h, \mp h), \\ -4/6 & \text{if } \mathbf{x} - \xi = (\pm h, 0) \text{ or } (0, \pm h), \\ 0 & \text{otherwise.} \end{cases} \tag{4.6.6}$$

The right side of the system of equations $L_h u_h = q_h$ is

$$q_h := f_h + \varphi_h, \quad f_h := \tilde{R}_h f \text{ according to } (4'), \quad \varphi_h(x) := -\sum_{\xi \in \Gamma_h} L_{\mathbf{x}\xi}\varphi(\xi). \tag{4.6.7}$$

The discretisation $(D_h, R_h, \tilde{R}_h)$ with D_h from (2), R_h from (5.2), $\tilde{R}_h$ from (4') is called the mehrstellen method (cf. Collatz [1]).

Exercise 4.6.1. Let D_h and L_h be defined respectively by (2) and (6). Prove that
(a) L_h is an M-matrix;
(b) $\|L_h^{-1}\|_\infty \le 1/8$ (stability), $\|L_h\|_\infty \le 20h^{-2}/3$;
(c) $\|D_h R_h u - \tilde{R}_h \Delta u\|_\infty \le \frac{11h^4}{180}\|u\|_{C^6(\overline{\Omega})}$, if $u \in C^6(\overline{\Omega})$ (consistency).

Theorem 4.6.2. *Let Ω_h be defined as in Section 4.2. Let u_h be the solution provided by the mehrstellen method $L_h u_h = q_h$ with L_h from (6), and q_h from (7). Let the solution u of the boundary value problem (2.2a,b) belong to $C^6(\overline{\Omega})$ or $C^{5,1}(\overline{\Omega})$. Then the following error estimates hold in the respective cases :*

$$\|u_h - R_h u\|_\infty \le \frac{11h^4}{1440}\|u\|_{C^6(\overline{\Omega})}, \quad \text{resp.} \quad \|u_h - R_h u\|_\infty \le \frac{11h^4}{1440}\|u\|_{C^{5,1}(\overline{\Omega})}. \tag{4.6.8a}$$

In the case of the potential equation, i.e. $f = 0$, we even have

$$\|u_h - R_h u\|_\infty \le Kh^6\|u\|_{C^8(\overline{\Omega})}, \quad \|u_h - R_h u\|_\infty \le Kh^6\|u\|_{C^{7,1}(\overline{\Omega})}, \tag{4.6.8b}$$

if $u \in C^8(\overline{\Omega})$, resp. $u \in C^{7,1}(\overline{\Omega})$.

PROOF. As in Theorem 4.5.3, (8a) results from Exercise (1b,c). In the case that $f = 0$ the $O(h^4)$-term also vanishes in (3). ∎

4.7 The Discretisation of the Neumann Boundary Value Problem

The Dirichlet boundary condition $u(\mathbf{x}) = \varphi(\mathbf{x})$ was used directly in the difference method; a discretisation was not necessary. A different situation arises for the Neumann boundary value problem

$$-\Delta u = f \quad \text{in} \quad \Omega = (0,1)\times(0,1), \quad \partial u/\partial n = \varphi \quad \text{on } \Gamma. \tag{4.7.1}$$

The normal derivative, which reads explicitly

$$\begin{aligned} \frac{\partial u}{\partial n} &= -u_y \quad \text{for } \mathbf{x} = (x,0) \in \Gamma, & \frac{\partial u}{\partial n} &= u_x \quad \text{for } \mathbf{x} = (x,1) \in \Gamma, \\ \frac{\partial u}{\partial n} &= -u_x \quad \text{for } \mathbf{x} = (0,y) \in \Gamma, & \frac{\partial u}{\partial n} &= u_y \quad \text{for } \mathbf{x} = (1,y) \in \Gamma, \end{aligned} \tag{4.7.2}$$

like the Laplace operator, must be replaced by a difference. We will investigate three different discretisations.

4.7.1 One-sided Difference for $\partial u/\partial n$

The Poisson equation leads to the $(n-1)^2 = (1/h-1)^2$ equations

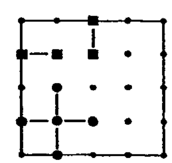
$$(-\Delta_h u_h)(\mathbf{x}) = f(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Omega_h, \tag{4.7.3}$$

which require the values of $u_h(\mathbf{x})$ for all $\mathbf{x} \in \overline{\Omega}'_h$, where

$$\overline{\Omega}'_h := \Omega_h \cup \Gamma'_h, \quad \Gamma'_h := \Gamma_h \backslash \{(0,0),(0,1),(1,0),(1,1)\}.$$

To obtain a further $4(n-1)$ equations for $\{u_h(\mathbf{x}) : \mathbf{x} \in \Gamma'_h\}$, we replace, at all $\mathbf{x} \in \Gamma'_h$, the normal derivative $\partial u/\partial n = \varphi$ by the backward difference

$$\partial_n^- u_h(\mathbf{x}) := h^{-1}[u_h(\mathbf{x}) - u_h(\mathbf{x} - h\mathbf{n})] = \varphi(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Gamma'_h. \tag{4.7.4}$$



■ Discretisation of the Neumann condition
● Discretisation of the Poisson equation

Figure 4.7.1.

If one inserts the corresponding normal directions $\mathbf{n}$ for the four sides of the square one obtains

$$\left.\begin{array}{l} h^{-1}[u_h(x,0)-u_h(x,h)]=\varphi(x,0) \\ h^{-1}[u_h(x,1)-u_h(x,1-h)]=\varphi(x,1) \end{array}\right\} \text{ for } x=h,2h,\ldots,1-h,$$
$$\left.\begin{array}{l} h^{-1}[u_h(0,y)-u_h(h,y)]=\varphi(0,y) \\ h^{-1}[u_h(1,y)-u_h(1-h,y)]=\varphi(1,y) \end{array}\right\} \text{ for } y=h,2h,\ldots,1-h. \tag{4.7.4'}$$

Equations (3) and (4) yield $(n+1)^2-4$ equations for as many unknowns.

Exercise 4.7.1. After a rescaling of equations (4) to $h^{-1}\partial_n^- u_h(\mathbf{x}) = h^{-1}\varphi(\mathbf{x})$, $\mathbf{x}\in\Gamma_h'$, these equations, together with (3), form a system $L_h u_h = q_h$. Show that L_h is symmetric and satisfies (3.1a).

As in the Dirichlet problem the variables $u_h(\mathbf{x})$, $\mathbf{x}\in\Gamma_h'$, can be eliminated with the aid of (4) in (3). At the point (h,y) near the boundary, for example, Equation (3) becomes

$$h^{-2}[3u_h(h,y)-u_h(h,y-h)-u_h(h,y+h)-u_h(2h,y)] = f(h,y)+h^{-1}\varphi(0,y).$$

The star $h^{-2}\begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}$ thus becomes

$$h^{-2}\begin{bmatrix} & -1 & \\ 0 & 3 & -1 \\ & -1 & \end{bmatrix}, \qquad h^{-2}\begin{bmatrix} & -1 & \\ -1 & 3 & 0 \\ & -1 & \end{bmatrix},$$
$$h^{-2}\begin{bmatrix} & 0 & \\ -1 & 3 & -1 \\ & -1 & \end{bmatrix}, \qquad h^{-2}\begin{bmatrix} & -1 & \\ -1 & 3 & -1 \\ & 0 & \end{bmatrix},$$

near, respectively, the left, right, upper and lower boundaries. At the corner points one even has to replace two boundary values, so that, for example, at $(h,h)\in\Omega_h$ the star reads $h^{-2}\begin{bmatrix} & -1 & \\ 0 & 2 & -1 \\ & 0 & \end{bmatrix}$. Except for the special case $h=1/2$, one obtains for the $(n-1)^2$ values $u_h(\mathbf{x})$, $\mathbf{x}\in\Omega_h$, the system of equations

$$L_h u_h = q_h \tag{4.7.5a}$$

with $$L_{\mathbf{xx}} = h^{-2}\begin{cases} 4 & \text{if } \mathbf{x}\in\Omega_h \text{ is far from the boundary,} \\ 2 & \text{if } \mathbf{x}=(h,h),(h,1-h),(1-h,h),(1-h,1-h), \\ 3 & \text{otherwise} \end{cases}$$
$$L_{\mathbf{x}\xi} = \begin{cases} -h^{-2} & \text{if } \mathbf{x},\xi\in\overline{\Omega}_h \text{ are neighbours,} \\ 0 & \text{otherwise for } \mathbf{x}\neq\xi \end{cases} \tag{4.7.5b}$$

and $$q_h = f_h+\varphi_h,\quad f_h(\mathbf{x}) = (\tilde{R}_h f)(\mathbf{x}) := f(\mathbf{x}),\quad \varphi_h(\mathbf{x}) = -h\textstyle\sum_{\xi\in\Gamma_h'} L_{\mathbf{x}\xi}\varphi(\xi). \tag{4.7.5c}$$

Remark 4.7.2. (a) L_h is symmetric and satisfies the sign condition (3.1a). (b) With lexicographical arrangement of the grid points of Ω_h, L_h has the form

$$L_h = h^{-2}\begin{bmatrix} T-I & -I & & & \\ -I & T & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & T & -I \\ & & & -I & T-I \end{bmatrix},$$
$$T = \begin{bmatrix} 3 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 3 \end{bmatrix}.$$

The matrix L_h is singular because the system $L_h u_h = q_h$, like the continuous boundary value problem (1), is, in general, not solvable. The analogue of Theorem 3.4.1 reads as follows.

Theorem 4.7.3. *The system of equations* (5a) *is solvable if and only if*

$$-h^2\sum_{\mathbf{x}\in\Omega_h} f(\mathbf{x}) = h\sum_{\mathbf{x}\in\Gamma_h'}\varphi(\mathbf{x}). \tag{4.7.6}$$

Any two solutions of (5a) *can only differ by a constant:* $u_h^1 - u_h^2 = c\mathbb{1}$, $c\in\mathbb{R}$.

PROOF. Evidently, $L_h\mathbb{1}=0$ holds, i.e., $\mathbb{1}\in\text{kernel}(L_h)$. Furthermore, Theorem 4 will then imply dim(kernel (L_h)) $=1$. This proves

$$\text{kernel}(L_h) = \{c\mathbb{1}: c\in\mathbb{R}\} \tag{4.7.7}$$

and thus the second part of the assertion. (5a) is solvable if and only if the scalar product $\langle v, q_h\rangle$ vanishes for all $v\in$ kernel (L_h^{T}) = range $(L_h)^{\perp}$. Because of $L_h^{\mathsf{T}} = L_h$ and (7)

$$\langle\mathbb{1}, q_h\rangle = 0 \quad\text{or}\quad \sum_{\mathbf{x}\in\Omega_h} q_h(\mathbf{x}) = 0 \tag{4.7.6'}$$

is sufficient and necessary. According to Definition (5c), (6) and (6′) agree. ■

Let condition (6) be satisfied. System (5a) can be solved as follows. Select an arbitrary $\mathbf{x}_0\in\Omega_h$ and normalise the solution u_h (determined except for one constant) by

$$u_h(\mathbf{x}_0) = 0. \tag{4.7.8}$$

Let $\hat{u}_h$ be the vector u_h without the component $u_h(\mathbf{x}_0)$. Let $\hat{L}_h$ be the submatrix of L_h in which the row and column with index $\mathbf{x}_0$ have been left out. Let $\hat{q}_h$ be constructed likewise. Then

$$\hat{L}_h \hat{u} = \hat{q}_h \tag{4.7.9}$$

is a system with $(n-1)^2 - 1$ equations and unknowns.

Theorem 4.7.4. *The system of equations* (9) *is solvable; in particular, $\hat{L}_h$ is a symmetric M-matrix. Under condition* (6), *$\hat{u}_h = \hat{L}_h^{-1}\hat{q}_h$, supplemented by* (8), *yields the solution u_h of system* (5a).

PROOF. (a) As a principal submatrix of L_h, $\hat{L}_h$ is symmetric. In $\Omega_h \backslash \{x_0\}$ any two grid points can be connected by a chain of neighbouring points so that $\hat{L}_h$ is irreducible. For all $x \in \Omega_h \backslash \{x_0\}$ there holds (3.4b); at neighbouring points of x_0 we even have (3.4a), so that $\hat{L}_h$ is irreducibly diagonally dominant. According to Criterion 4.3.10, $\hat{L}_h$ is an M-matrix, thus nonsingular.
(b) If u_h is the solution of (5a), one can assume (8) without loss of generality, so that u_h restricted to $\Omega_h \backslash \{x_0\}$ also solves Equation (9) and has to agree with the unique solution $\hat{u}_h$. ■

As a corollary of Theorem 4 one obtains that $\text{rank}(L_h) \geq \text{rank}(\hat{L}_h) = (n-1)^2 - 1$, i.e., $\dim(\text{kernel } (L_h)) = 1$ and hence (7) holds.

Another possibility for the solution of Equation (5a) is to pass to an extended system of equations

$$\overline{L}_h \overline{u}_h = \overline{q}_h \tag{4.7.10a}$$

$$\text{with} \quad \overline{L}_h = \begin{bmatrix} L_h & \mathbb{1} \\ \mathbb{1}^{\mathsf{T}} & 0 \end{bmatrix}, \quad \overline{u}_h = \begin{bmatrix} u_h \\ \lambda \end{bmatrix}, \quad \overline{q}_h = \begin{bmatrix} q_h \\ \sigma \end{bmatrix}, \tag{4.7.10b}$$

where σ can be prescribed arbitrarily.

Theorem 4.7.5. *Equation* (10a) *is always solvable. If for the last component of the solution $\overline{u}_h$ we have $\lambda = 0$, then condition* (6) *is satisfied and u_h represents the solution of system* (5a) *which is normalised by $\mathbb{1}^{\mathsf{T}} u_h = \sum_{x \in \Omega_h} u_h(x) = \sigma$. However, if $\lambda \neq 0$ holds one can interpret u_h as solution of $L_h u_h = \tilde{q}_h$ where $\tilde{q}_h = q_h - \lambda \mathbb{1}$ belongs to $\tilde{f}(x) := f(x) - \lambda$ and $\tilde{f}$ and φ satisfy condition* (6).

PROOF. $\mathbb{1}$ is linearly independent of the columns of L_h so that rank $[L_h, \mathbb{1}] = \text{rank}\,(L_h) + 1 = (n-1)^2$. Likewise, $(\mathbb{1}^{\mathsf{T}}, 0)$ is linearly independent of the rows of $[L_h, \mathbb{1}]$ so that rank $(\overline{L}_h) = (n-1)^2 + 1$, i.e., $\overline{L}_h$ is nonsingular. The other statements can be read from (10b). ■

Attention. One should either use Equation (10a,b) or Equation (9), after first replacing q_h by $\tilde{q}_h := q_h - (\mathbb{1}^{\mathsf{T}} q_h / \mathbb{1}^{\mathsf{T}} \mathbb{1})\mathbb{1}$. As a justification of this recommendation note that the condition for solvability of the continuous problem is $\int_\Omega f dx + \int_\Gamma \varphi d\Gamma = 0$ and that this does not at all imply the discrete solvability condition (6). For smooth functions f and φ Equation (6) can be shown to hold up to a remainder of order $O(h)$. Thus it is generally unavoidable to

replace f_h and q_h by $f_h - \lambda\mathbb{1}$ and $q_h - \lambda\mathbb{1}$ ($\lambda = \mathbb{1}^{\mathsf{T}} q_h / \mathbb{1}^{\mathsf{T}} \mathbb{1}$). In the case of Equation (10a,b) this correction is carried out implicitly. If, however, (9) is used without any correction, the resulting solution can be interpreted as a solution of $L_h u_h = \tilde{q}_h$ with $\tilde{q}_h(x) = q_h(x)$ for $x \neq x_0$ and $\tilde{q}_h(x_0) := -\sum_{x \neq x_0} q_h(x)$. That is, here too, an implicit correction of q_h is carried out, with the difference that the correction is not distributed over all components as before, but is concentrated on $q_h(x_0)$. If Equation (6) is satisfied up to order $O(h)$, then $q_h(x_0)$ and $\tilde{q}_h(x_0)$ differ by $O(h^{-1})$. Therefore the solution $\hat{u}_h$ of Equation (9) contains a singularity at the point x_0.

Theorem 4.7.6. (Convergence) *Let $u \in C^{3,1}(\overline{\Omega})$ be the solution of the Neumann problem* (1). *Let $\overline{u}_h = \begin{pmatrix} u_h \\ \lambda \end{pmatrix}$ be the solution of Equation* (10a). *Then there exists a $c \in \mathbb{R}$ and constants C, C' independent of u and h such that*

$$|\lambda| \leq C' h \|u\|_{C^{0,1}(\overline{\Omega})}, \quad \|u_h - R_h u - c\mathbb{1}\|_\infty \leq C[h\|u\|_{C^{1,1}(\overline{\Omega})} + h^2 \|u\|_{C^{3,1}(\overline{\Omega})}]. \tag{4.7.11}$$

PROOF. (a) We have

$$\lambda = \mathbb{1}^{\mathsf{T}} q_h / \mathbb{1}^{\mathsf{T}} \mathbb{1} = (h^2 \sum_{\Omega_h} f(x) + h \sum_{\Gamma'_h} \varphi(x)) h^{-2} (n-1)^{-2}.$$

Because $\int_\Omega f\, dx + \int_\Gamma \varphi\, d\Gamma = 0$, the first factor consists only of the quadrature error $O(h\|u\|_{C^{0,1}(\overline{\Omega})})$.

(b) According to Theorem 5, u_h is the solution of $L_h u_h = \tilde{q}_h := q_h - \lambda\mathbb{1}$. This corresponds to the difference equations

$$-\Delta_h u_h = \tilde{f}_h := f_h - \lambda\mathbb{1} \quad \text{in } \Omega_h, \quad \partial_n^- u_h = \varphi \quad \text{on } \Gamma'_h.$$

The difference $w_h := u_h - R_h u$ satisfies

$$\begin{aligned} -\Delta_h w_h &= -\Delta_h u_h + \Delta_h R_h u = \Delta_h R_h u + f_h - \lambda\mathbb{1} \\ &= \Delta_h R_h u - \tilde{R}_h \Delta u - \lambda\mathbb{1} =: c_h \quad \text{in } \Omega_h, \end{aligned} \tag{4.7.12a}$$

$$\partial_n^- w_h = \partial_n^- u_h - \partial_n^- R_h u = \varphi - \partial_n^- R_h u = \frac{\partial u}{\partial n} - \partial_n^- R_h u =: \psi \quad \text{on } \Gamma'_h. \tag{4.7.12b}$$

The errors of consistency are

$$\|c_h\|_\infty \leq \frac{1}{6} h^2 \|u\|_{C^{3,1}(\overline{\Omega})} \qquad \text{(cf. Remark 4.5.2)},$$

$$|\psi(x)| \leq \frac{1}{2} h \|u\|_{C^{1,1}(\overline{\Omega})} + |\lambda| \quad \text{(cf. Lemma 4.1.1)}.$$

Since the solution w_h exists, condition (6) is satisfied with c_h and ψ:

$$h^2 \sum_{\Omega_h} c_h(\mathbf{x}) + h \sum_{\Gamma_h'} \psi(\mathbf{x}) = 0.$$

Then $\tilde{w}_h := w_h - c\mathbb{1}$ with $c = \mathbb{1}^\mathsf{T} w_h / \mathbb{1}^\mathsf{T}\mathbb{1}$ is the solution of (12a,b) normalised by $\mathbb{1}^\mathsf{T} w_h = 0$. The application of the following Theorem 4.7.7 to Equation (12a,b) provides the inequality (11). ∎

Theorem 4.7.7. (Stability) *Let condition* (6) *be satisfied. Let the solution* u_h *of* (3), (4) *be normalised by* $\mathbb{1}^\mathsf{T} u_h = 0$. *Then there exist constants* C_1, C_2 *independent of* u *and* h *such that*

$$\|u_h\|_\infty \le C_1 \max_{\mathbf{x}\in\Omega_h} |f(\mathbf{x})| + C_2 \max_{\mathbf{x}\in\Gamma_h'} |\varphi(\mathbf{x})|. \tag{4.7.13}$$

The proof of this theorem, which corresponds to Theorem 4.4.11, will be supplied in Sect. 4.7.4.

4.7.2 Symmetric Difference for $\partial u/\partial n$

As can be seen from Theorem 6, the one-sided difference ∂_n^- causes the error term $O(h)$. It seems obvious to replace ∂_n^- by a symmetric difference. To this end the five-point discretisation is set up at all points $\mathbf{x} \in \overline{\Omega}_h = \Omega_h \cup \Gamma_h$ (cf. (2.1c)):

$$-\Delta_h u_h = f_h := \tilde{R}_h f \quad \text{in } \overline{\Omega}_h, \tag{4.7.14a}$$

where $\tilde{R}_h$ is the restriction to $\overline{\Omega}_h$. The symmetric difference ∂_n^0 is defined by

$$(\partial_n^0 u_h)(\mathbf{x}) := (2h)^{-1}[u_h(\mathbf{x}+h\mathbf{n}) - u_h(\mathbf{x}-h\mathbf{n})] = \varphi(\mathbf{x}) \text{ in } \mathbf{x} \in \Gamma_h. \tag{4.7.14b}$$

Here, we assign two normal directions to the corner points, so that for each of the corner points two equations of the form (14b) can be set up. In the corner $\mathbf{x} = (0,0)$ one has, for example, the normals

$$\mathbf{n} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \qquad \mathbf{n} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

The corresponding equations (14b) also contain different (!) values $\varphi(0+,0) = \lim_{x\to 0}\varphi(x,0)$ and $\varphi(0,0+) = \lim_{y\to 0}\varphi(0,y)$. For $\mathbf{x} \in \Gamma_h$ the difference formula (14a) needs the values in points $\mathbf{x} + h\mathbf{n}$ outside of $\overline{\Omega}_h$. These can be eliminated with the aid of (14b) so that a system of equations $L_h u_h = q_h$ remains for the $(n+1)^2$ components $u_h(\mathbf{x})$, $\mathbf{x} \in \overline{\Omega}_h$.

Exercise 4.7.8. (a) If the grid points of $\overline{\Omega}_h$ are arranged in lexicographical order, L_h has the form

$$L_h = h^{-2}\begin{bmatrix} T & -2I & & & \\ -I & T & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & T & -I \\ & & & -2I & T \end{bmatrix}, \quad T = \begin{bmatrix} 4 & -2 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -2 & 4 \end{bmatrix}.$$

(b) L_h is not symmetric, but $D_h L_h$ with $D_h = \operatorname{diag}\{d(x)d(y)\}$, $d(0) = d(1) = 1/2$, $d(\cdot) = 1$ otherwise, is symmetric.

The analogue of Theorem 3 reads:

Theorem 4.7.9. *Equation* (14a,b) *is solvable if and only if* $\mathbb{1}^\mathsf{T} D_h q_h = 0$ *for* D_h *in Exercise* 8b. *Any two solutions may differ by only a constant. The formulation of* $\mathbb{1}^\mathsf{T} D_h q_h = 0$ *with the aid of* f *and* φ *reads:*

$$-h^2 \sum_{(x,y)\in\overline{\Omega}_h} d(x)d(y)f(x,y) = 2h \sum_{(x,y)\in\Gamma_h} d(x)d(y)\varphi(x,y), \tag{4.7.15}$$

where the summand for the corner points occurs twice in the second sum, and both the different limits for φ *are taken into account.*

Remark 4.7.10. The sums in (15) represent summed trapezoidal formulae. Thus, from $\int_\Omega f dx + \int_\Gamma \varphi d\Gamma = 0$ follows Equation (15), except for a remainder $O(h^2\|u\|_{C^{1,1}(\overline{\Omega})})$.

Theorems 4 and 5 can be transferred without difficulty. The convergence theorem, Theorem 6, becomes

Theorem 4.7.11. *Let* $u \in C^{3,1}(\overline{\Omega})$ *be a solution of* (1). *Let* $\overline{u}_h = \begin{pmatrix} u_h \\ \lambda \end{pmatrix}$ *be the solution of* $\overline{L}_h \overline{u}_h = \begin{pmatrix} D_h q_h \\ 0 \end{pmatrix}$ *with* $\overline{L}_h = \begin{bmatrix} D_h L_h & \mathbb{1} \\ \mathbb{1}^\mathsf{T} & 0 \end{bmatrix}$. *We have convergence of second order:*

$$|\lambda| \le C' h^2 \|u\|_{C^{1,1}(\overline{\Omega})}, \quad \|u_h - R_h u - c\mathbb{1}\|_\infty \le C h^2 \|u\|_{C^{3,1}(\overline{\Omega})}. \tag{4.7.16}$$

The proof is essentially the same as for Theorem 6. An additional technical difficulty is the fact that the consistency error $\Delta_h R_h u - \tilde{R}_h \Delta u$ must also be determined in $\mathbf{x} \in \Gamma_h$ although u is only defined in $\overline{\Omega}$. Instead of treating the difference equation and the boundary discretisation separately as in (12a,b) one should directly analyse the equations $L_h u_h = q_h$ from which the values $u_h(\mathbf{x} + h\mathbf{n})$ outside of $\overline{\Omega}$ have already been eliminated.

4.7.3 Symmetric Difference for $\partial u/\partial n$ on an Offset Grid

If we offset the above grid by $h/2$ in the x and y directions, we obtain the grid

$$\Omega_h := \left\{(x,y) \in \Omega : \frac{x}{h} - \frac{1}{2}, \frac{y}{h} - \frac{1}{2} \in \mathbb{Z}\right\}$$

in Figure 2. The points near the boundary of Ω_h are at a distance $h/2$ from Γ. We set

$$\Gamma_h := \left\{(x,y) \in \Gamma : \frac{x}{h} - \frac{1}{2} \in \mathbb{Z} \text{ or } \frac{y}{h} - \frac{1}{2} \in \mathbb{Z}\right\}$$

(cf. Figure 2). To each point near the boundary $\mathbf{x} - h\mathbf{n}/2$ ($\mathbf{x} \in \Gamma_h$) corresponds an outlying neighbour $\mathbf{x} + h\mathbf{n}/2$.

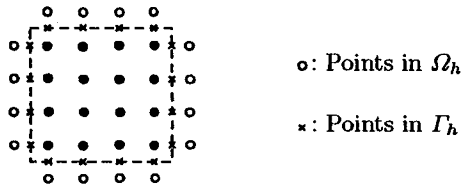


Figure 4.7.2. Offset grid

The discretisation of the Poisson equation (1) is

$$-\Delta_h u_h(\mathbf{x}) = f(\mathbf{x}) \quad \text{in } \mathbf{x} \in \Omega_h, \tag{4.7.17a}$$

$$h^{-1}[u_h(\mathbf{x} + h\mathbf{n}/2) - u_h(\mathbf{x} - h\mathbf{n}/2)] = \varphi(\mathbf{x}) \quad \text{in } \mathbf{x} \in \Gamma_h. \tag{4.7.17b}$$

The difference (17b) is symmetric with respect to the boundary point $\mathbf{x}$ and nevertheless agrees with the backward difference ∂_n^- at the grid point $\mathbf{x}+h\mathbf{n}/2$.

Remark 4.7.12. After eliminating the values $u_h(\mathbf{x} + h\mathbf{n}/2)$, $\mathbf{x} \in \Gamma_h$, we obtain a system of equations $L_h u_h = q_h$, where Remark 2 is also valid for this matrix L_h. In contrast to Section 4.7.1, L_h is of size $n^2 \times n^2$.

The Theorems 3, 4, 5, and 7 hold analogously. Theorem 6 holds with the inequality (16) instead of (11).

4.7.4 Proof of the Stability Theorem 7

In the case of Dirichlet boundary values the stability statement of Theorem 4.4.11 follows immediately from the maximum principle and the bound for L_h^{-1}. The corresponding statement of Theorem 7 for Neumann boundary values, however, cannot be proved that easily. In the literature one can only find weaker estimates which on the right-hand side of Equation (13) contain an additional factor $|\log h|$. However, this factor is not to be avoided if one uses Equation (9), $\hat{L}_h \hat{u}_h = \hat{q}_h$, without condition (6) being satisfied.

Let the discrete Green function $g_h(\mathbf{x},\xi)$, $\mathbf{x} \in \Omega_h \cup \Gamma_h$, $\xi \in \Omega_h$ be defined by

$$-\Delta_h g_h(\mathbf{x},\xi) = \delta(\mathbf{x},\xi) \tag{4.7.18a}$$
$$:= \begin{Bmatrix} h^{-2} & \text{for } \mathbf{x} = \xi \\ 0 & \text{otherwise} \end{Bmatrix} - \begin{Bmatrix} h^{-1}/(4-4h) & \text{if } \mathbf{x} \text{ is near the boundary} \\ 0 & \text{if } \mathbf{x} \text{ is far from the boundary} \end{Bmatrix},$$

$$\partial_n^- g_h(\mathbf{x},\xi) = 0 \quad \text{for } \mathbf{x} \in \Gamma_h' \tag{4.7.18b}$$

where Δ_h and ∂_n^- act on $\mathbf{x}$. g_h exists since $\sum_{\mathbf{x}\in\Omega_h} \delta(\mathbf{x},\xi) = 0$ proves the solvability condition (6′).

Lemma 4.7.13. *For arbitrary q_h with $\mathbb{1}^\mathsf{T} q_h = 0$ (cf. (6′)),*

$$u_h(\mathbf{x}) := h^2 \sum_{\xi\in\Omega_h} q_h(\xi) g_h(\mathbf{x},\xi) \tag{4.7.19}$$

represents a solution of $L_h u_h = q_h$.

PROOF. At points far from the boundary $\mathbf{x} \in \Omega_h$, $(L_h u_h)(\mathbf{x}) = -\Delta_h u_h(\mathbf{x}) = q_h(\mathbf{x})$. At points near the boundary $\mathbf{x} \in \Omega_h$, from $\partial_n^- u_h = 0$ and (6′), one has the identity

$$(L_h u_h)(\mathbf{x}) = (-\Delta_h u_h)(\mathbf{x}) = q_h(\mathbf{x}) - \frac{h^{-1}}{4-4h} \sum_{\xi\in\Omega_h} q_h(\xi) = q_h(\mathbf{x}). \qquad \blacksquare$$

Equation (18a,b) determine g_h up to a constant.

Theorem 4.7.14. *The Green function $g_h(\mathbf{x},\xi)$ can be so selected that*

$$|g_h(\mathbf{x},\xi)| \le C[1 + |\log(|\mathbf{x}-\xi| + h)|] \quad \textit{for } \mathbf{x},\xi \in \Omega_h. \tag{4.7.20}$$

This inequality corresponds to the bound (4.7) in the Dirichlet case. Before Theorem 14 is proved, we want to show that Theorem 7 follows from it. Analogously to $\int_\Omega [1 + |\log|\mathbf{x}-\xi||]\,d\mathbf{x} \le K_1$ and $\int_\Gamma [1 + |\log|\mathbf{x}-\xi||]\,d\Gamma_\mathbf{x} \le K_2$ we obtain

$$h^2 \sum_{\mathbf{x}\in\Omega_h} [1 + |\log(|\mathbf{x}-\xi| + h)|] \le K_1',\ h \sum_{\mathbf{x}\in\Gamma_h'} [1 + |\log(|\mathbf{x}-\xi| + h)|] \le K_2'.$$

From (19), (20), and $q_h = f_h + \varphi_h$ (cf. (5c)) thus follows the estimate

$$|u_h(\mathbf{x})| \le K_1 \|f_h\|_\infty + hk_2 \|\varphi_h\|_\infty \le K_1 \max_{\mathbf{x}\in\Omega_h} |f(\mathbf{x})| + 2K_2 \max_{\mathbf{x}\in\Gamma_h'} |\varphi(\mathbf{x})|.$$

If $\tilde{u}_h := u_h - \mathbb{1}^\mathsf{T} u_h / \mathbb{1}^\mathsf{T}\mathbb{1}$ is the solution of $L_h u_h = q_h$ normalised by $\mathbb{1}^\mathsf{T} \tilde{u}_h = 0$, one can see that

$$\|\tilde{u}_h\|_\infty \le 2\inf\{\|u_h - c\mathbb{1}\|_\infty : c \in \mathbb{R}\} \le 2\|\tilde{u}_h\|_\infty,$$

so that the inequality (11) and Theorem 7 are proved. $\blacksquare$

It remains to prove Theorem 14.

Lemma 4.7.15. *Let* $\mathbf{e}$ *be one of the unit vectors* $(1,0)$ *or* $(0,1)$. *Let* $\xi \in \Omega_h$ *be such that* $\xi + h\mathbf{e} \in \Omega_h$ *also. For all these* $\mathbf{e}$ *and* ξ *let there exist a function* $G_h(\mathbf{x}) = G_h(\mathbf{x};\xi,\mathbf{e})$ *with the properties*

$$-\Delta_h G_h(\mathbf{x}) = h^{-2}\begin{Bmatrix} 1 & \text{if } \mathbf{x}=\xi \\ -1 & \text{if } \mathbf{x}=\xi+h\mathbf{e} \\ 0 & \text{otherwise}\end{Bmatrix} \text{ in } \Omega_h, \tag{4.7.21a}$$

$$\partial_n^- G_h = 0 \quad \text{on } \Gamma_h', \tag{4.7.21b}$$

$$|G_h(\mathbf{x})| \le hC'/(|\mathbf{x}-\xi|+h), \tag{4.7.21c}$$

where C' *does not depend on* $\mathbf{e}$ *and* ξ. *Then Theorem 14 holds, and thus also Theorem 7.*

PROOF. For $\xi, \xi' \in \Omega_h$ let $g_h(\mathbf{x},\xi,\xi')$ be defined as a solution of

$$\partial_n^- g_h(\cdot,\xi,\xi') = 0 \text{ on } \Gamma_h', \quad -\Delta_h g_h(\mathbf{x},\xi,\xi') = h^{-2}\begin{Bmatrix} 1 & \text{if } \mathbf{x}=\xi \\ -1 & \text{if } \mathbf{x}=\xi' \\ 0 & \text{otherwise}\end{Bmatrix} \text{ in } \Omega_h.$$

For $\xi' = \xi + h\mathbf{e}$, $g_h(\cdot,\xi,\xi')$ agrees with G_h in Lemma 15. For arbitrary ξ, ξ', one finds a connection $\xi = \xi^0, \xi^1, \cdots, \xi^l = \xi'$ with $\xi^{k+1} - \xi^k = \pm h\mathbf{e}^k$, $\mathbf{e}^k = (1,0)$ or (0,1). Since $g_h(\cdot,\xi,\xi') = \sum_{k=1}^{l} g_h(\cdot,\xi^{k-1},\xi^k)$, (21c) implies the estimate

$$|g_h(\mathbf{x},\xi,\xi')| \le h\sum_{k=1}^{l} C'/(|\mathbf{x}-\xi^{k-1}|+h).$$

If one considers first the case $x_1 = \xi_1 = \xi_1'$, $x_2 \le \xi_2 < \xi_2'$ and one applies $\sum_{k=k_1}^{k_2} 1/k \le \text{const} \cdot (1 + \log(k_2 h) - \log(k_1 h))$, one obtains

$$|g_h(\mathbf{x},\xi,\xi')| \le C''[1 + |\log(|\mathbf{x}-\xi|+h)| + |\log(|\mathbf{x}-\xi'|+h)|].$$

In the general case, this bound also comes out, but one must choose the connection ξ^k such that $|\mathbf{x}-\xi^k| \ge \min\{|\mathbf{x}-\xi|, |\mathbf{x}-\xi'|\}$. Since we know that $h\sum_{\xi'\in\Gamma_h'} |\log(|\mathbf{x}-\xi'|+h)| \le \text{const}$ for all $\mathbf{x} \in \Omega_h$,

$$g_h(\mathbf{x},\xi) := \frac{h}{4-4h}\sum_{\xi'\in\Gamma_h'} g_h(\mathbf{x},\xi,\xi')$$

satisfies the inequality (20). Because $\sum_{\Gamma_h'} 1 = (4-4h)/h$ the equations (18a,b) are also satisfied, i.e., $g_h(\mathbf{x},\xi)$ is the Green function. ■

We shall construct the function G_h needed in Lemma 15 explicitly. The basic building block is the discrete singularity function $s_h(\mathbf{x},\xi)$ on the grid in $\mathbb{R}^2$

$$Q_h := \{(x,y) \in \mathbb{R}^2 : x/h, y/h \in \mathbb{Z}\}.$$

Lemma 4.7.16. *The singularity function defined by* $s_h(\mathbf{x},\xi) := \sigma_h(\mathbf{x}-\xi)$ *and*

$$\sigma_h(\mathbf{x}) := \frac{1}{16\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{e^{i(x_1\eta_1+x_2\eta_2)/h} - 1}{\sin^2(\eta_1/2)+\sin^2(\eta_2/2)}\,d\eta_1\,d\eta_2$$
$$= -\frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{\sin^2((x_1\eta_1+x_2\eta_2)/(2h))}{\sin^2(\eta_1/2)+\sin^2(\eta_2/2)}\,d\eta_1\,d\eta_2$$

for all $\mathbf{x},\xi \in Q_h$ *has the property* $-\Delta_h s_h(\mathbf{x},\xi) = h^{-2}$ *for* $\mathbf{x} = \xi$ *and* $-\Delta_h s_h = 0$ *otherwise.*

PROOF. Let $e(\mathbf{x},\eta) := \exp(i(x_1\eta_1 + x_2\eta_2)/h)$. Note that

$$-\Delta_h(e(\mathbf{x},\eta) - 1) = 4h^{-2}e(\mathbf{x},\eta)[\sin^2(\eta_1/2) + \sin^2(\eta_2/2)]$$

and

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e(\mathbf{x},\eta)\,d\eta = \begin{Bmatrix} 4\pi^2 & \text{for } \mathbf{x}=0 \\ 0 & \text{for } \mathbf{x}\ne 0\end{Bmatrix}.$$ ■

For multi-indices $\alpha = (\alpha_1,\alpha_2) \in \mathbb{Z}^2$ with $\alpha_1 \ge 0$, $\alpha_2 \ge 0$ one defines the partial difference operators of order $|\alpha| = \alpha_1 + \alpha_2$ by

$$\partial^\alpha = (\partial_x^+)^{\alpha_1}(\partial_y^+)^{\alpha_2}. \tag{4.7.22}$$

Starting with the representation in Lemma 16, Thomée [1] proves

Lemma 4.7.17. $|(\partial^\alpha \sigma_h)(\mathbf{x})| \le C(|\mathbf{x}|+h)^{-|\alpha|}$ *for all* $\mathbf{x} \in Q_h$.

For the construction of the function G_h in Lemma 15 we select, without loss of generality, $\mathbf{e} = (1,0)$ and keep $\xi \in \Omega_h$ with $\xi + h\mathbf{e} \in \Omega_h$. The function

$$G_h'(\mathbf{x}) := h\partial_x^-\sigma_h(\mathbf{x}-\xi) = \sigma_h(\mathbf{x}-\xi) - \sigma_h(\mathbf{x}-\xi-h\mathbf{e}) \tag{4.7.23}$$

satisfies (21a) but not the boundary condition (21b). Let the symmetrisation operators S_x and S_y be defined by

$$S_x u_h(x,y) := \frac{1}{2}[u_h(x,y) + u_h(h-x,y)], \quad S_y u_h(x,y) := \frac{1}{2}[u_h(x,y) + u_h(x,h-y)]$$

for $(x,y) \in Q_h$. The function

$$G_h'' := 4S_xS_yG_h' \tag{4.7.24}$$

is symmetric with respect to the axes $y = h/2$ and $x = h/2$. Thus we have (21b) on the left and lower boundary:

$$\partial_n^- G_h''(0,y) = \partial_n^- G_h''(x,0) = 0 \tag{4.7.25}$$

Furthermore, G_h'' satisfies condition (21a) just as G_h' does. For each $\beta \in \mathbb{Z}^2$ we define the operator P_β by

$$(P_\beta u_h)(x,y) = \frac{1}{4}[u_h(x+\beta_1 L, y+\beta_2 L) + u_h(x-\beta_1 L, y+\beta_2 L)$$
$$+u_h(x+\beta_1 L, y-\beta_2 L) + u_h(x-\beta_1 L, y-\beta_2 L)],$$

where $L = 2-2h$, and set

$$G_h^\beta(\mathbf{x}) := (P_\beta G_h'')(\mathbf{x}) - (P_\beta G_h'')(0). \tag{4.7.26}$$

Lemma 4.7.18. *For $\beta = (\beta_1, \beta_2)$ with $\|\beta\|_\infty \geq 2$ there holds*

$$|\partial^\alpha G_h^\beta(\mathbf{x})| \leq hK/|\beta|^3 \quad \textit{for all } |\alpha| \leq 2,\ \mathbf{x} \in \Omega_h. \tag{4.7.27}$$

Here, K is independent of the choice of the $\xi, \xi + he \in \Omega_h$.

PROOF. According to the definition we have

$$G_h^\beta(0,0) = 0. \tag{4.7.28a}$$

The operator P_β preserves the symmetry, i.e., $u_h = S_x u_h$ implies $P_\beta u_h = S_x P_\beta u_h$. Thus we have that $G_h^\beta = S_x G_h^\beta = S_y G_h^\beta$ and consequently

$$\partial_x^+ G_h^\beta(0,0) = \partial_y^+ G_h^\beta(0,0) = 0. \tag{4.7.28b}$$

$G_h^\beta(\mathbf{x})$ is the linear combination of $h\partial_x^- \sigma_h(\mathbf{x}-\tilde{\xi})$ for different $\tilde{\xi}$ with $|\mathbf{x}-\tilde{\xi}|+h \geq K'|\beta|$, if $\mathbf{x} \in \Omega_h$. According to Lemma 17,

$$|(\partial^\alpha G_h^\beta)(\mathbf{x})| \leq hCK''|\beta|^{-3} \quad \text{for all } |\alpha| = 2, \mathbf{x} \in \Omega_h, \tag{4.7.28c}$$

so that (27) follows for $|\alpha| = 2$. Let $|\alpha| = 1$. $\partial^\alpha G_h(\mathbf{x})$ can be written in the form

$$\partial^\alpha G_h^\beta(0) + \sum_{k=1}^{l} [\partial^\alpha G_h^\beta(\mathbf{x}^k) - \partial^\alpha G_h^\beta(\mathbf{x}^{k-1})]$$

with $\mathbf{x}^0 = \mathbf{x}$, $\mathbf{x}^l = \mathbf{x}$, $\|\mathbf{x}^k - \mathbf{x}^{k-1}\|_\infty = h$. Each summand has the form $\pm h\partial^\gamma G_h^\beta$ with $|\gamma| = 2$ so that (28b,c) lead to the estimate (28d)

$$|(\partial^\alpha G_h^\beta)(\mathbf{x})| \leq 2hCK''|\beta|^{-3} \quad \text{for } |\alpha| = 1,\ \mathbf{x} \in \Omega_h. \tag{4.7.28d}$$

Likewise one infers from (28a) and (28d) the inequality (27) for $|\alpha| = 0$, which proves Lemma 18. ■

Since $\sum_{\beta \in \mathbf{Z}^2 \setminus \{0\}} |\beta|^{-3} < \infty$,

$$G_h(\mathbf{x}) := \sum_{\beta \in \mathbf{Z}^2} G_h^\beta(\mathbf{x}) \tag{4.7.29}$$

exists. Since $-\Delta_h G_h^\beta(\mathbf{x}) = 0$ in Ω_h for all $\beta \neq (0,0)$, G_h also satisfies Equation (21a). As already mentioned in the proof of (28b), $G_h = S_x G_h = S_y G_h$ holds so that G_h also satisfies Equation (25): $\partial_n^- G_h = 0$ at the left and lower boundary. The proof of $\partial_n^- G_h = 0$ at the lower boundaries requires

Lemma 4.7.19. *G_h is L-periodic: $G_h(x,y) = G_h(x+L,y) = G_h(x,y+L)$.*

PROOF. Let $G_h''(x+\beta_1 L, y+\beta_2 L) - G_h''(\beta_1 L, \beta_2 L)$ be abbreviated to $\gamma_\beta(x,y)$. Definition (29) says that

$$G_h = \lim_{k\to\infty} \sum_{|\beta_1| \leq k} \sum_{|\beta_2| \leq k} \gamma_\beta. \tag{4.7.30a}$$

Now γ_β can be written as a sum over the differences $h\partial_h^+ G_h''(x+\beta_1 L, \beta_2 L + \mu h)$ from $\mu h = 0$ to $\mu h = x$, and over $h\partial_h^+ G_h''(\beta_1 L + \nu h, \beta_2 L)$ from $\nu h = 0$ to $\nu h = y$. For $|\beta_1| \geq 2$ the distance between the arguments of ξ and $\xi + he$ is always $\leq (|\beta_1| - 1)L$. Each summand is thus, by Lemma 17, bounded by $O(h^2/(|\beta_1|-1)^2) = O(h^2/\beta_1^2)$. As a sum of such terms then γ_β can be estimated by

$$|\gamma_\beta| \leq O(h/\beta_1^2) \quad \text{for } |\beta_1| \geq 2. \tag{4.7.30b}$$

We want now to show that the following equation (30c) also holds:

$$G_h = \lim_{k\to\infty} \sum_{|\beta_1 - 1| \leq k} \sum_{|\beta_2| \leq k} \gamma_\beta = \lim_{k\to\infty} \sum_{1-k \leq \beta_1 \leq 1+k} \sum_{|\beta_2| \leq k} \gamma_\beta. \tag{4.7.30c}$$

The sums (30a) and (30c) differ by $D_h := \sum_{|\beta_2| \leq k} [\gamma_{(1+k,\beta_2)} - \gamma_{(-k,\beta_2)}]$. The values $1+k$ and $-k$ that can be given to β_1 satisfy $|\beta_1| \leq 2k$. The absolute value of the sum D_h is then bounded, from (30b), by $\sum_{|\beta_2| \leq k} O(h/k^2) = (2k+1)O(h/k^2) = O(h/k)$, so that the limits of the double sums in (30a) and (30c) are the same. Changing the variable β_1 to $\beta_1 - 1$ then transforms (30c) into

$$G_h(x,y) = \lim_{k\to\infty} \sum_{-k \leq \beta_1, \beta_2 \leq k} [G_h''(x+\beta_1 L + L, y + \beta_2 L) - G_h''(\beta_1 L + L, \beta_2 L)] \tag{4.7.30d}$$

The first part of the identity

$$\sum_{\beta_1,\beta_2=-k}^{k} [G_h''(\beta_1 L, \beta_2 L) - G_h''(\beta_1 L + L, \beta_2 L)]$$
$$= \sum_{\beta_2=-k}^{k} [G_h''(-kL, \beta_2 L) - G_h''(kL + L, \beta_2 L)]$$
$$= \sum_{\beta_2=-k}^{k} [G_h''(h + kL, \beta_2 L) - G_h''(kL + L, \beta_2 L)] \tag{4.7.30e}$$

is elementary; the second results from the symmetry $G_h'' = S_x G_h''$. As above, the summands of the last sum can be bounded by $O(h/k^2)$ so that (30e) vanishes for $k \to \infty$. Together with (30d) one obtains

$$G_h(x,y) = \lim_{k\to\infty} \sum_{-k\le\beta_1,\beta_2\le k} [G_h''(x+\beta_1 L + L, y+\beta_2 L) - G_h''(\beta_1 L, \beta_2 L)]$$

$$= \lim_{k\to\infty} \sum_{-k\le\beta_1,\beta_2\le k} \gamma_\beta(x+L,y) = G_h(x+L,y).$$

The proof of $G_h(x,y) = G_h(x, y+L)$ is analogous. ■

Since $L = 2-2h$, the symmetry $G_h = S_x G_h$ and the periodicity yield

$$G_h(1,y) = G_h(h-1,y) = G_h(h-1+L,y) = G_h(1-h,y),$$

i.e., $\partial_n^- G_h(1,y) = 0$ on the upper boundary $(1,y) \in \Gamma_h'$. Likewise we show that $\partial_n^- G_h(x,1) = 0$ on the right boundary. Thus (21b) is also proved.
It remains to show (21c): $|G_h(\mathrm{x})| \le hC/(|\mathrm{x}-\xi|+h)$ in Ω_h. The G_h^β for $\|\beta\|_\infty \le 1$ are linear combinations of $h\partial_x^- \sigma_h(\mathrm{x}-\tilde\xi)$ for $\tilde\xi = \xi$ and other $\tilde\xi \notin \Omega_h$, which are generated by S_x, S_y, and P_β. For all ξ there holds $|\mathrm{x}-\tilde\xi| \ge |\mathrm{x}-\xi|$, so that from Lemma 17 follows

$$|G_h^\beta(\mathrm{x})| \le hC'/(|\mathrm{x}-\xi|+h) \quad \text{for } \|\beta\|_\infty \le 1, \quad \mathrm{x}\in\Omega_h.$$

On the other hand, Lemma 18 shows that

$$\Big|\sum_{\|\beta\|_\infty\ge 2} G_h^\beta(\mathrm{x})\Big| \le h \sum_{\|\beta\|_\infty \ge 2} K/|\beta|^3 = hK' \le hK''/(|\mathrm{x}-\xi|+h).$$

The assumptions of Lemma 15 are hence satisfied: G_h with (21a-c) exists. Thus theorems 14 and 7 are proved. ■

4.8 Discretisation in an Arbitrary Domain

4.8.1 Shortley-Weller Approximation

Let the boundary value problem

$$-\Delta u = f \quad \text{in } \Omega, \quad u = \varphi \quad \text{on } \Gamma \tag{4.8.1}$$

be given for an arbitrary domain Ω. If one places a square grid of step size h over Ω one obtains

$$\Omega_h := \{(x,y)\in\Omega\colon x/h\in\mathbb{Z} \text{ and } y/h\in\mathbb{Z}\} \tag{4.8.2}$$

as the set of grid points. By contrast, the set Γ_h of boundary points must be defined differently from the case of the square. The left neighbour point of $(x,y)\in\Omega_h$ reads $(x-h,y)$. If the connecting segment $\{(x-\vartheta h, y)\colon \vartheta\in(0,1]\}$ does not lie completely in Ω, there exists a boundary point

$$(x-sh,y)\in\Gamma \text{ with } (x,y)\in\Omega_h, s\in(0,1], (x-\vartheta h,y)\in\Omega \text{ for all } \vartheta\in[0,s), \tag{4.8.3a}$$

which now, instead of $(x-h,y)$, is called the left neighbour point of (x,y) (cf. Figure 1). Likewise the right, lower, and upper neighbour points can be boundary points of the following form:

$$(x+sh,y)\in\Gamma \text{ with } (x,y)\in\Omega_h, s\in(0,1], (x+\vartheta h,y)\in\Omega \text{ for all } \vartheta\in[0,s), \tag{4.8.3b}$$

$$(x,y-sh)\in\Gamma \text{ with } (x,y)\in\Omega_h, s\in(0,1], (x,y-\vartheta h)\in\Omega \text{ for all } \vartheta\in[0,s), \tag{4.8.3c}$$

$$(x,y+sh)\in\Gamma \text{ with } (x,y)\in\Omega_h, s\in(0,1], (x,y+\vartheta h)\in\Omega \text{ for all } \vartheta\in[0,s). \tag{4.8.3d}$$

We set

$$\Gamma_h := \{\text{boundary points which satisfy (3a), (3b), (3c), or (3d)}\}. \tag{4.8.4}$$

A grid point $(x,y)\in\Omega_h$ possessing a neighbour from Γ_h, is said to be near the boundary. All other points of Ω_h are said to be far from the boundary. As can be seen in Figure 1, a point may be near the boundary although $(x\pm h,y)$ and $(x,y\pm h)$ belong to Ω_h (namely if not all connecting segments lie completely in Ω).

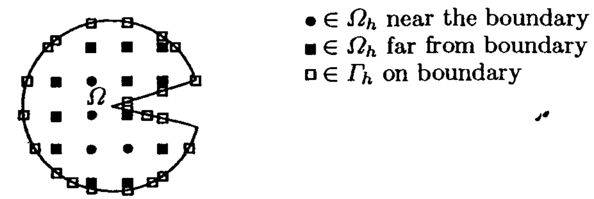


Figure 4.8.1. Ω_h and Γ_h

Exercise 4.8.1. For a convex domain Ω show that

$$\Gamma_h = \{(x,y)\in\Gamma\colon x/h\in\mathbb{Z} \text{ or } y/h\in\mathbb{Z}\}.$$

If one wishes to approximate the second derivative $u''(x)$ with the aid of the values of u at $x' < x < x''$, one can use Newton's divided differences:

$$u''(x) = 2\Big[\frac{u(x'')-u(x)}{x''-x} - \frac{u(x)-u(x')}{x-x'}\Big]/(x''-x') + \text{Rem}. \tag{4.8.5}$$

Exercise 4.8.2. Show that (a) the Taylor expansion implies

$$\begin{aligned}|\text{Rem}| &\le \frac{1}{3}\frac{(x''-x)^2+(x-x')^2}{x''-x'}\|u\|_{C^3([x',x''])} \\ &\le \frac{\max\{x''-x, x-x'\}}{3}\|u\|_{C^3([x',x''])}\end{aligned} \tag{4.8.6}$$

for the remainder in Equation (5) if $u\in C^3([x',x''])$.

(b) In Equation (6) one can replace the norm of $C^3([x',x''])$ by that of $C^{2,1}([x',x''])$.
(c) If $x'' = x+h$ and $x' = x-h$ the difference in (5) agrees with the usual second difference $\partial^-\partial^+u(x)$.

To set up the difference equation for $-\Delta u = f$ at $(x,y) \in \Omega_h$ we use the four neighbouring points

$$(x - s_l h, y),\ (x + s_r h, y),\ (x, y - s_u h),\ (x, y + s_o h) \in \Omega_h \cup \Gamma_h,\ 0 < s_* \le 1,$$

as defined above. The subscripts to the s are suggest respectively left, right, under and over. For points far from the boundary $s_* = 1$ holds; for points near the boundary (x,y) at least one neighbour lies on Γ_h and the corresponding distance $s_* h$ may be smaller than h. Equation (5) with $x' = x - s_l h$ and $x'' = x + s_r h$ provides an approximation for u_{xx}. Analogously one can replace u_{yy} by a divided difference. One obtains the difference scheme of Shortley and Weller [1]:

$$\begin{aligned} -D_h u(x,y) := h^{-2}\Bigg\{ &\left(\frac{2}{s_l s_r} + \frac{2}{s_o s_u}\right) u(x,y) - \frac{2}{s_r(s_r+s_l)} u(x+s_r h, y) \\ &- \frac{2}{s_l(s_r+s_l)} u(x - s_l h, y) - \frac{2}{s_o(s_o+s_u)} u(x, y+s_o h) \\ &- \frac{2}{s_u(s_o+s_u)} u(x, y - s_u h)\Bigg\}. \end{aligned} \tag{4.8.7}$$

Remark 4.8.3. If $s_l = s_r = s_u = s_o = 1$, D_h agrees with Δ_h.

The discrete boundary value problem assumes the form

$$-D_h u_h = f_h := \tilde{R}_h f \quad \text{on } \Omega_h \tag{4.8.8a}$$
$$\text{with } (\tilde{R}_h f)(\mathbf{x}) = f(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Omega_h,$$
$$u_h = \varphi \quad \text{on } \Gamma_h. \tag{4.8.8b}$$

The five coefficients on the right side of Equation (7) define the matrix elements $L_{\mathbf{x}\xi}$ for $\xi = \mathbf{x}$ and for the four neighbours ξ of $\mathbf{x}$. Otherwise we set $L_{\mathbf{x}\xi} = 0$. The right side of the system of equations

$$L_h u_h = q_h, \tag{4.8.9a}$$

in which u_h is interpreted as a grid function on Ω_h (not $\Omega_h \cup \Gamma_h$) is given by

$$q_h = f_h + \varphi_h, \quad \varphi_h(\mathbf{x}) := -\sum_{\xi \in \Gamma_h} L_{\mathbf{x}\xi}\varphi(\xi). \tag{4.8.9b}$$

Again $\varphi_h(\mathbf{x}) = 0$ holds for points $\mathbf{x} \in \Omega_h$ far from the boundary.

Theorem 4.8.4. *Let Ω be bounded and contained in a strip $(x_0, x_0+d) \times \mathbb{R}$ or $\mathbb{R} \times (y_0, y_0+d)$ of width d. For the matrix L_h belonging to the Shortley-Weller discretisation the following holds:*
(a) L_h is generally not symmetric;
(b) L_h is M-matrix with

$$\|L_h^{-1}\|_\infty \le d^2/8. \tag{4.8.10}$$

PROOF. (a) Let $\mathbf{x} = (x,y) \in \Omega_h$ be near the boundary, but its neighbour $\mathbf{x}' = (x+h, y) \in \Omega_h$ be far from the boundary. Then it holds that $L_{\mathbf{x}\mathbf{x}'} = -2h^{-2}/(1+s_l) \ne -h^{-2} = L_{\mathbf{x}'\mathbf{x}}$, if $s_l < 1$. Other than in Exercise 4.7.1, in general no scaling can be found so that $D_h L_h$ (D_h diagonal) becomes symmetric.
(b) L_h need not necessarily be irreducible and hence irreducibly diagonally dominant. But the weaker condition in Exercise 4.3.8 is satisfied and proves the M-matrix property.
(c) For the proof of (10) we use Theorem 4.3.16. If the domain Ω lies in the strip $(x_0, x_0+d) \times \mathbb{R}$ we select $w_h(x,y) := R_h w$, $w := (x-x_0)(x_0+d-x)/2$. The remainder in Equation (5) contains only third derivatives that vanish for w. Thus $D_h w_h$ agrees with $\Delta w = -1$:

$$-D_h w_h = \mathbb{1} \quad \text{in } \Omega_h, \quad w_h \ge 0 \quad \text{on } \Gamma_h.$$

The corresponding system of equations reads $L_h w_h = q_h := f_h + \varphi_h$ with $f_h = \mathbb{1}$ and $q_h \ge 0$. Thus we have $L_h w_h \ge \mathbb{1}$ and Theorem 4.3.16 proves $\|L_h^{-1}\|_\infty \le \|w_h\|_\infty \le (d/2)^2/2 = d^2/8$. ■

Exercise 4.8.5. Prove the analogue of the derivative (4.4.9):

$$\|u_h\|_\infty \le \|L_h^{-1}\|_\infty \|f_h\|_\infty + \max_{\xi\in\Gamma_h}|\varphi(\xi)| \le \frac{d^2}{8}\max_{\mathbf{x}\in\Omega_h}|f(\mathbf{x})| + \max_{\xi\in\Gamma_h}|\varphi(\xi)|.$$

With (10) stability has been proved. The order of consistency however is only 1, for at points near the boundary

$$c_h := D_h R_h u - \tilde{R}_h \Delta u$$

is of the order of magnitude $O(h^1)$. Here, R_h is the restriction to $\Omega_h \cup \Gamma_h$. $\tilde{R}_h$ has been defined in (8a). We want to show that nevertheless there is convergence of order 2. The difference $w_h := u_h - R_h u$ between the discrete solution $u_h = L_h^{-1} q_h$ and the solution $u \in C^{3,1}(\bar{\Omega})$ of (1) satisfies

$$\begin{aligned} -D_h w_h &= -D_h u_h + D_h R_h u = \tilde{R}_h f + D_h R_h u \\ &= D_h R_h u - \tilde{R}_h \Delta u = c_h \qquad \text{in } \Omega_h, \end{aligned} \tag{4.8.11a}$$

$$w_h = 0 \quad \text{on } \Gamma_h, \tag{4.8.11b}$$

so that $w_h = L_h^{-1}c_h$ follows. c_h can be written as $c_h^x + c_h^y$, where c_h^x (resp. c_h^y) is the error of discretisation of the x difference (resp. y difference). In turn, c_h^x is split into $c_h^x = c_h^{x,1} + c_h^{x,2}$:

$$c_h^{x,2}(x,y) := \begin{Bmatrix} c_h^x(x,y) & \text{if } s_l = s_r = 1 \\ 0 & \text{otherwise} \end{Bmatrix}, \qquad c_h^{x,1} := c_h^x - c_h^{x,2}.$$

Analogously one defines $c_h^{y,1}$ and $c_h^{y,2}$ and sets

$$c_h^1 := c_h^{x,1} + c_h^{y,1}, \quad c_h^2 := c_h^{x,2} + c_h^{y,2}, \quad w_h^i := L_h^{-1}c_h^i.$$

The errors c_h^2 are described by (5.7):

$$\|w_h^2\|_\infty \le \|L_h^{-1}\|_\infty \|c_h^2\|_\infty, \quad \|c_h^2\|_\infty \le \frac{1}{6}h^2\|u\|_{C^{3,1}(\overline{\Omega})}. \tag{4.8.12a}$$

With $K := \frac{1}{3}h^3\|u\|_{C^{2,1}(\overline{\Omega})}$ define

$$v_h = K\mathbb{1} \quad \text{in } \Omega_h, \quad v_h = 0 \quad \text{on } \Gamma_h, \quad \tilde{c}_h := L_h v_h.$$

For $\mathbf{x} \in \Omega_h$ far from the boundary $\tilde{c}_h(\mathbf{x}) = 0$; for $\mathbf{x} \in \Omega_h$ near the boundary, however, we have

$$\tilde{c}_h(\mathbf{x}) = K \sum_{\xi\in\Omega_h} L_{\mathbf{x}\xi} = -K \sum_{\xi\in\Gamma_h} L_{\mathbf{x}\xi} \quad (\mathbf{x} \in \Omega_h \text{ near boundary}).$$

Consider, for example, the case $\mathbf{x} = (x,y) \in \Omega_h$ near the boundary with $\xi = (\mathbf{x} - s_l h, y) \in \Gamma_h$. According to Exercise (2a,b) the x difference has the error

$$|c_h^{x,1}(\mathbf{x})| \le \frac{h}{3}\frac{s_r^2 + s_l^2}{s_r + s_l}\|u\|_{C^{2,1}(\overline{\Omega})} = h^{-2}\frac{s_r^2 + s_l^2}{s_r + s_l}K \le K\frac{2h^{-2}}{s_l(s_r + s_l)} = -KL_{\mathbf{x}\xi}.$$

The analogous estimate for $c_h^{y,1}(\mathbf{x})$ gives $|c_h^1(\mathbf{x})| \le \tilde{c}_h(\mathbf{x})$. Because $c_h^1(\mathbf{x}) = \tilde{c}_h(\mathbf{x}) = 0$ at the $\mathbf{x} \in \Omega_h$ far from the boundary one has $-\tilde{c}_h \le c_h^1 \le \tilde{c}_h$. Since $L_h^{-1} \ge 0$ it follows that $-v_h \le w_h^1 \le v_h$, i.e.,

$$\|w_h^1\|_\infty \le K = \frac{1}{3}h^3\|u\|_{C^{2,1}(\overline{\Omega})}. \tag{4.8.12b}$$

Equations (12a,b) together with $w_h^1 + w_h^2 = w_h = u_h - R_h u$ prove

Theorem 4.8.6. *Let Ω satisfy the assumption of Theorem 4. The convergence for the Shortley-Weller method is second order:*

$$\begin{aligned}\|u_h - R_h u\|_\infty &\le \frac{1}{3}h^3\|u\|_{C^{2,1}(\overline{\Omega})} + \|L_h^{-1}\|_\infty \frac{1}{6}h^2\|u\|_{C^{3,1}(\overline{\Omega})} \\ &\le \left(\frac{1}{3}h^3 + \frac{d^2}{48}h^2\right)\|u\|_{C^{3,1}(\overline{\Omega})},\end{aligned} \tag{4.8.13}$$

if $u \in C^{3,1}(\overline{\Omega})$.

Exercise 4.8.7. Show that if one uses the Shortley-Weller discretisation for all grid points near the boundary, but the mehrstellen method from Section 4.6 for all points far from the boundary, we obtain a method with convergence of third order: $\|u_h - R_h u\|_\infty = O(h^3)$.

4.8.2 Interpolation at Points near the Boundary

Instead of discretising the Poisson equation at grid points $\mathbf{x} \in \Omega_h$ near the boundary, one could also try to determine $u_h(\mathbf{x})$ by interpolation from the neighbouring points. If, for example, $\mathbf{x} = (x,y) \in \Omega_h$ is near the boundary, $(x - s_l h, y) \in \Gamma_h$ and $(x + s_r h, y) \in \Omega_h \cup \Gamma_h$, then linear interpolation yields

$$u_h(x,y) = [s_l u_h(x + s_r h, y) + s_r u_h(\mathrm{x} - s_l h, y)]/(s_r + s_l).$$

Thus, at the point $\mathbf{x}$ we set up the equation

$$(s_l + s_r)u_h(x,y) - s_l u_h(x + s_r h, y) - s_r u_h(\mathrm{x} - s_l h, y) = 0. \tag{4.8.14a}$$

Since $\xi = (x - s_l h, y)$ should be a boundary point, one can replace $u_h(\xi)$ by $\varphi(\xi)$. If, however, $(x, y + s_o h)$ or $(x, y - s_u h)$ is a boundary point, we choose interpolation in the y direction:

$$(s_u + s_o)u_h(x,y) - s_u u_h(x, y + s_o h) - s_o u_h(x, y - s_u h) = 0. \tag{4.8.14b}$$

At any $\mathbf{x} \in \Omega_h$ far from the boundary the five-point formula (2.3) is used:

$$-(\Delta_h u_h)(\mathbf{x}) = f_h(\mathbf{x}) = \tilde{R}_h f(\mathbf{x}) = f(\mathbf{x}) \quad (\mathbf{x} \text{ far from boundary}). \tag{4.8.14c}$$

Theorem 4.8.8. *Let Ω satisfy the assumptions of Theorem 4. Let the discretisation be given by (14a-c) with the choice between (14a) and (14b) being made in such a way that always (at least) one boundary point is used for the interpolation. Let the system of equations resulting after the elimination of the boundary values $u_h(\xi) = \varphi(\xi)$, $\xi \in \Gamma_h$ be $L_h u_h = q_h$. L_h is an (in general unsymmetric) M-matrix which satisfies the estimate (10). The discrete solutions u_h converge with the order 2, if $u \in C^{3,1}(\overline{\Omega})$:*

$$\|u_h - R_h u\|_\infty \le h^2\|u\|_{C^{1,1}(\overline{\Omega})} + \|L_h^{-1}\|_\infty \frac{1}{6}h^2\|u\|_{C^{3,1}(\overline{\Omega})}. \tag{4.8.15}$$

PROOF. The M-matrix property and (10) are proved as in Theorem 4.2. Let $\mathbf{x} \in \Omega_h$ be a point near the boundary at which (14a) is used. The error of interpolation is

$$c_h^1(x,y) := (s_l + s_r)R_h u(x,y) - s_l R_h u(x + s_r h, y) - s_r R_h u(x - s_l h, y),$$

$$|c_h^1(x,y)| \le \frac{1}{2}s_r s_l(s_l + s_r)h^2\|u\|_{C^{1,1}(\overline{\Omega})}.$$

With this c_h^1 and $K := h^2\|u\|_{C^{1,1}(\bar{\Omega})}$, one can essentially just repeat the proof of Theorem 6. ∎

By rescaling and adding the equations (14a,b), one obtains

$$h^{-2}\left\{\left(\frac{s_l+s_r}{s_l s_r}+\frac{s_o+s_u}{s_o s_u}\right)u_h(x,y)-\frac{1}{s_l}u_h(x-s_l h,y)\right.$$
$$-\frac{1}{s_r}u_h(x+s_r h,y)-\frac{1}{s_o}u_h(x,y+s_o h)$$
$$\left.-\frac{1}{s_u}u_h(x,y-s_u h)\right\}=0. \tag{4.8.16}$$

Using this device one can obtain a symmetric matrix L_h, even at arbitrary Ω.

Exercise 4.8.9. Show that the discretisation (14c), (16) leads to a symmetric M-matrix. The estimate of convergence reads

$$\|u_h - R_h u\|_\infty \le 2h^2\|u\|_{C^{1,1}(\bar{\Omega})} + \frac{1}{6}d^2h^2\|u\|_{C^{3,1}(\bar{\Omega})}.$$

In (14a,b) linear interpolation was chosen because the values at the neighbouring points were sufficient for it. Constant interpolation by

$$u(x,y)=u(x-s_l h,y)=\varphi(x-s_l h,y), \quad \text{if } (x,y)\in\Omega_h, (x-s_l h,y)\in\Gamma_h,$$

is less desirable since it only provides first-order convergence : $\|u_h - R_h u\|_\infty = O(h)$. By contrast, interpolation of higher order is very applicable indeed (cf. Pereyra–Proskurowski–Widlund [1]). However, it is described by an equation which also contains points at an distance of $\ge 2h$. Higher boundary approximations are in particular then necessary if one wants to apply extrapolation methods (cf. Marchuk-Shaidurov [1, p. 162 ff.]).

5 General Boundary Value Problems

5.1 Dirichlet Boundary Value Problems for Linear Differential Equations

5.1.1 Posing the Problem

In Section 1.2 we have already formulated the general linear differential equation of second order:

$$Lu = f \quad \text{in } \Omega \tag{5.1.1a}$$

with

$$L=\sum_{i,j=1}^{n}a_{ij}(\mathbf{x})\frac{\partial^2}{\partial x_i\partial x_j}+\sum_{i=1}^{n}a_i(\mathbf{x})\frac{\partial}{\partial x_i}+a(\mathbf{x}). \tag{5.1.1b}$$

We mentioned that, without loss of generality, one can assume

$$a_{ij}(\mathbf{x}) = a_{ji}(\mathbf{x}) \tag{5.1.1c}$$

so that the matrix

$$\mathbf{A}(\mathbf{x}) := (a_{ij}(\mathbf{x}))_{i,j=1,\cdots,n} \tag{5.1.1d}$$

is symmetric. A differential operator apparently more general than (1b) is

$$L=\sum_{i,j=1}^{n}\left[a_{ij}^{\mathrm{I}}\frac{\partial^2}{\partial x_i\partial x_j}+\frac{\partial}{\partial x_j}a_{ij}^{\mathrm{II}}\frac{\partial}{\partial x_i}+\frac{\partial^2}{\partial x_i\partial x_j}a_{ij}^{\mathrm{III}}\right]+\sum_{i=1}^{n}\left[a_i^{\mathrm{I}}\frac{\partial}{\partial x_i}+\frac{\partial}{\partial x_i}a_i^{\mathrm{II}}\right]+a. \tag{5.1.2}$$

But since, for example, $\frac{\partial}{\partial x_j}(a_{ij}^{\mathrm{II}}\frac{\partial}{\partial x_i}u) = a_{ij}^{\mathrm{II}}u_{x_ix_j} + (\partial a_{ij}^{\mathrm{II}}/\partial x_j)u_{x_i}$, the operator (2) can be described in the form (1b), provided the coefficients are sufficiently often differentiable. According to Definition 1.2.3, Equation (1) is elliptic in Ω if all eigenvalues of $\mathbf{A}(\mathbf{x})$ have the same sign. One can assume without loss of generality that all eigenvalues are positive so that $\mathbf{A}(\mathbf{x})$ is positive definite (cf. Exercise 4.3.22a).

Thus, L is elliptic in Ω if

$$\sum_{i,j=1}^{n}a_{ij}(\mathbf{x})\xi_i\xi_j > 0 \quad \text{for all } \mathbf{x}\in\Omega, \quad 0\ne\xi\in\mathbb{R}^n. \tag{5.1.3a}$$