

2. Finite Differenzen für elliptische Gleichungen

Im Gegensatz zu hyperbolischen Gleichungen haben elliptische Gleichungen einen Glättungseffekt, d.h. im Allgemeinen besitzen solche Gleichungen eine starke (glatte) Lösung. Man erwartet daher, dass Differenzenverfahren weniger anfällig reagieren als dies bei hyperbolischen Gleichungen der Fall ist. Wir werden uns zunächst mit Differenzenverfahren in einer Raumdimension beschäftigen, d.h. also mit Verfahren mit Randwertprobleme von gewöhnlichen Differentialgleichungen und dabei bereits die wesentlichen Theorien von Differenzenverfahren für elliptische Gleichungen studieren.

2.1. Differenzenverfahren in einer Raumdimension. Im Falle einer Raumdimension betrachtet man Differentialoperatoren der Form

$$Lu = -au'' + bu' + cu, \quad c \geq 0$$

wobei L für $a > 0$ elliptisch ist. Wir verwenden im Folgenden $a = 1$ und betrachten Dirichlet-Probleme auf dem Einheitsintervall $(0, 1)$,

$$\begin{cases} Lu = -u'' + bu' + cu = f & \text{in } (0, 1) = \Omega \\ u = g & \text{auf } \Gamma = \{0, 1\} \end{cases}$$

Zur Herleitung eines Differenzenverfahrens führen wir in Analogie zu hyperbolischen Gleichungen die folgenden Schritte durch

- 1) wir definieren auf Ω und dem Rand Γ ein Gitter mittels

$$\begin{aligned} \Omega_h &= \{x | x = ih, i = 1, \dots, n-1\}, h = \frac{1}{n} \\ \Gamma_h &= \{0, 1\} \end{aligned}$$

und bezeichnen mit u_i die numerischen Approximationen von der exakten Lösung $u(x_i)$ mit $x_i \in \Omega_h \cup \Gamma_h$,

- 2) wir diskretisieren die Differentialoperatoren in der elliptischen Gleichung mit Hilfe von Differenzenquotienten, zum Beispiel durch

$$\begin{aligned} u'' &= \frac{u(x+h) - 2u(x) + u(x-h)}{h^2} + 0(h^2), \quad u \in \mathcal{C}^4 \\ u' &= \frac{u(x+h) - u(x-h)}{2h} + 0(h^2), \quad u \in \mathcal{C}^3 \end{aligned}$$

Damit erhalten wir eine diskrete Version des elliptischen Problems $Lu = f$ an den Gitterpunkten $x_i \in \Omega_h$ in der Form

$$(1.2.1) \quad -\frac{(u_{i+1} - 2u_i + u_{i-1}))}{h^2} + b_i \frac{u_{i+1} - u_{i-1}}{2h} + c_i u_i = f_i$$

wobei die Koeffizienten in (1.2.1) von den Gitterpunkten abhängen, $b_i = b(x_i)$, $c_i = c(x_i)$ und $f_i = f(x_i)$. Die Gleichung (1.2.1) ist eine lineare Gleichung für die drei Unbekannten u_{i-1} , u_i und u_{i+1} , die man auch als

$$l_i u_{i-1} + d_i u_i + r_i u_{i+1} = f_i$$

mit

$$l_i = -\frac{1}{h^2} - \frac{b_i}{2h}, \quad d_i = \frac{2}{h^2} + c_i, \quad r_i = -\frac{1}{h^2} + \frac{b_i}{2h}$$

schreiben kann. Zusätzlich haben wir diskrete Randbedingungen der Form

$$u_0 = g(0) = g_0, \quad u_n = g(1) = g_n$$

Insgesamt erhalten wir also ein lineares Gleichungssystem der Ordnung $n + 1$ gegeben durch

$$\begin{pmatrix} 1 & & & & & \\ l_1 & d_1 & r_1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & l_{n-1} & d_{n-1} & r_{n-1} & \\ & & & & & 1 \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix} = \begin{pmatrix} g_0 \\ f_1 \\ \vdots \\ f_{n-1} \\ g_n \end{pmatrix}$$

das wir im Folgenden als

$$(1.2.2) \quad L_h \mathbf{u}_h = \mathbf{f}_h$$

Um also eine numerische Approximation der exakten Lösung zu berechnen, müssen wir das lineare System (1.2.2) lösen und damit sind die folgenden zentralen Fragen zu beantworten:

- 1) Ist das lineare System stets lösbar und existiert eine eindeutige Lösung?
- 2) Was passiert mit der Lösung (oder dem System) im Fall $h \rightarrow 0$?
- 3) Wie gut ist die numerische Lösung?

Analog zu hyperbolischen Gleichungen definieren wir ein konsistentes Differenzenverfahren. Sei dazu D_h ein Operator der kontinuierliche Funktionen auf $[0, 1]$ auf unser gegebenes Gitter abbildet,

$$D_h(g)v = \begin{pmatrix} g(x_0) \\ v(x_1) \\ \vdots \\ v(x_{n-1}) \\ g(x_n) \end{pmatrix}$$

mit $g \in \mathcal{C}(\Gamma)$ und $v \in \mathcal{C}(\Omega)$.

DEFINITION 1.72. Das Differenzenverfahren definiert durch L_h ist konsistent mit Konsistenzordnung m für $v \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\bar{\Omega})$, falls für $h \rightarrow 0$ gilt

$$\|L_h D_h(v)v - D_h(v)Lv\|_{h,\infty} \leq ch^m$$

wobei

$$\|(w_0, \dots, w_n)^t\|_{h,\infty} = \max_{i=0,\dots,n} |w_i|$$

Den Ausdruck $\mathbf{I}_h(v)$ definiert durch

$$\mathbf{I}_h(v) = L_h D_h(v)v - D_h(v)Lv$$

nennt man auch lokalen Diskretisierungsfehler.

Ein konvergentes Verfahren wird folgendermassen definiert:

DEFINITION 1.73. Das Differenzenverfahren definiert durch L_h ist konvergent mit Konvergenzordnung m , falls für den Fehler \mathbf{e}_h definiert durch

$$\mathbf{e}_h = \mathbf{u}_h - D_h(g)u,$$

für $h \rightarrow 0$ gilt

$$\|\mathbf{e}_h\|_{h,\infty} = O(h^m)$$

Dabei ist $\mathbf{u}_h = L_h^{-1}\mathbf{f}_h$, u die exakte Lösung und g sind die vorgegebenen Randbedingungen.

Wie beim Äquivalenzsatz von Lax für lineare hyperbolische Gleichungen ist ein Verfahren (1.2.2) für lineare elliptische Gleichungen konvergent, falls es konsistent und stabil ist. Wir berechnen

$$\begin{aligned} L_h \mathbf{e}_h &= L_h \mathbf{u}_h - L_h D_h(g)u = L_h \mathbf{u}_h - L_h D_h(u)u \\ &= \mathbf{f}_h - (\mathbf{I}_h(u) + D_h(u)Lu) \\ &= \mathbf{f}_h - D_h(u)f - \mathbf{I}_h(u) \\ &= \mathbf{f}_h - D_h(g)f - \mathbf{I}_h(u) = -\mathbf{I}_h(u) \end{aligned}$$

Existiert also L_h^{-1} , so erhalten wir $\mathbf{e}_h = -L_h^{-1}\mathbf{I}_h(u)$ und

$$\|\mathbf{e}_h\|_{h,\infty} \leq \|L_h^{-1}\|_{h,\infty} \|\mathbf{I}_h(u)\|_{h,\infty}$$

Ist das Verfahren konsistent, so gilt $\|\mathbf{I}_h(u)\|_{h,\infty} \leq ch^m$ und damit ist das Verfahren konvergent, falls $\|L_h^{-1}\|_{h,\infty}$ unabhängig von h beschränkt ist. Wir definieren daher:

DEFINITION 1.74. Ein Verfahren ist stabil, falls die Abschätzung $\|L_h^{-1}\|_{h,\infty} \leq K$ gilt, wobei K unabhängig von h ist.

SATZ 1.75. Das Differenzenverfahren definiert durch L_h sei stabil und konsistent. Dann konvergiert das Verfahren im Sinne der Definition 1.73 und die Konvergenzordnung ist gleich der Konsistenzordnung.

Wir versuchen nun notwendige Bedingungen abzuleiten, die die Stabilität von Differenzenverfahren gewährleisten.

DEFINITION 1.76. Eine $(n \times n)$ -Matrix A nennt man

- 1) L_0 -Matrix, falls $a_{ij} \leq 0$ für alle $i \neq j$,
- 2) M -Matrix, falls A eine reguläre L_0 -matrix ist und jeder Eintrag von A^{-1} nicht-negativ ist.

SATZ 1.77. Eine L_0 -Matrix A ist genau dann eine M -Matrix, falls ein $\mathbf{y} \in \mathbb{R}^n$ mit $\mathbf{y} > 0$ existiert, sodass $A\mathbf{y} > 0$. Dann gilt zusätzlich

$$\|A^{-1}\|_{\infty} \leq \frac{\|\mathbf{y}\|_{\infty}}{\min_k (A\mathbf{y})_k}$$

wobei $\|\mathbf{y}\|_{\infty} = \max_{i=1,\dots,n} |y_i|$.

BEWEIS. Ist A eine M -Matrix zu setzen wir $\mathbf{y} = A^{-1}(1, \dots, 1)^T$. Dann gilt offensichtlich $\mathbf{y} > 0$ und $A\mathbf{y} > 0$. Für die andere Richtung sei $\mathbf{y} > 0$ ein Vektor mit $A\mathbf{y} > 0$. Dann gilt

$$(1.2.3) \quad (A\mathbf{y})_i = \sum_{j=1}^n a_{ij}y_j > 0$$

Da A eine L_0 -Matrix ist, muss daher für die Diagonalelemente die Beziehung $a_{ii} > 0$, $i = 1, \dots, n$ gelten. Dann ist $D := \text{diag}(a_{11}, \dots, a_{nn})$ eine reguläre Diagonalmatrix und mit $C := D^{-1}(D - A)$ erhalten wir für A^{-1} die Darstellung

$$A^{-1} = (I - C)^{-1}D^{-1}$$

Ausgehend von dieser Darstellung möchten wir zeigen, dass alle Einträge in A^{-1} nicht-negativ sind.

Zunächst sind alle Einträge in D^{-1} und $D - A$ nicht-negativ, also gilt auch $C \geq 0$. Da $A\mathbf{y}$ strikt positiv ist und $D > 0$ gilt

$$(I - C)\mathbf{y} = D^{-1}A\mathbf{y} > 0$$

und damit erhalten wir die Abschätzung

$$(1.2.4) \quad C\mathbf{y} < \mathbf{y}$$

Wir definieren nun die gewichtete Maximumnorm

$$\|\mathbf{x}\|_y = \max_i \frac{|x_i|}{y_i}$$

auf \mathbb{R}^n und die zugehörige Matrixnorm $\|C\|_y$,

$$\|C\|_y = \max_{\|\mathbf{x}\|_y=1} \|C\mathbf{x}\|_y$$

Wie man leicht sieht, gilt nun wegen (1.2.4) die Abschätzung $\|C\|_y < 1$: für einen beliebigen Vektor \mathbf{x} mit $\|\mathbf{x}\|_y = 1$ gilt $-\mathbf{y} \leq \mathbf{x} \leq \mathbf{y}$. Da aber C strikt positiv ist, folgt auch

$$|(C\mathbf{x})_i| \leq |(C\mathbf{y})_i|$$

und daher $\|C\mathbf{x}\|_y \leq \|C\mathbf{y}\|_y$. Mit der Definition der Matrixnorm gilt

$$\|C\|_y = \|C\mathbf{y}\|_y$$

und wegen (1.2.4) $\frac{|(C\mathbf{y})_i|}{|y_i|} < 1$.

Damit läßt sich die Matrix $(I - C)^{-1}$ in eine konvergente Reihe entwickeln und mit $C \geq 0$ gilt

$$(I - C)^{-1} = \sum_{k=0}^{\infty} C^k \geq 0$$

Daraus folgt aber bereits

$$A^{-1} = (I - C)^{-1}D^{-1} \geq 0,$$

Zum Beweis der Normabschätzung für A^{-1} schreiben wir zunächst für den gegebenen Vektor $\mathbf{y} > 0$ mit $A\mathbf{y} > 0$ ebenfalls komponentenweise $A\mathbf{y} \geq \min_i (A\mathbf{y})_i \cdot (1, \dots, 1)^T$. Da

A^{-1} nur nicht-negative Einträge besitzt, folgt $\mathbf{y} \geq \min_i (A\mathbf{y})_i \cdot A^{-1}(1, \dots, 1)^T$, was man auch in der Form $A^{-1}(1, \dots, 1)^T \leq \mathbf{y} / \min_i (A\mathbf{y})_i$ schreiben kann. Dann folgt aber

$$(1.2.5) \quad \|A^{-1}(1, \dots, 1)^T\|_\infty \leq \frac{\|\mathbf{y}\|_\infty}{\min_i (A\mathbf{y})_i}$$

Ist $\mathbf{x} \in \mathbb{R}^n$ beliebig, so gilt die Abschätzung $-\|\mathbf{x}\|_\infty \cdot (1, \dots, 1)^T \leq \mathbf{x} \leq \|\mathbf{x}\|_\infty \cdot (1, \dots, 1)^T$ und daher auch

$$-\|\mathbf{x}\|_\infty \cdot A^{-1}(1, \dots, 1)^T \leq A^{-1}\mathbf{x} \leq \|\mathbf{x}\|_\infty \cdot A^{-1}(1, \dots, 1)^T$$

Damit erhält man aber die Abschätzung

$$\|A^{-1}\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_\infty \|A^{-1}(1, \dots, 1)^T\|$$

Zusammen mit (1.2.5) ergibt sich

$$\|A^{-1}\|_\infty = \sup_{\|\mathbf{x}\|_\infty=1} \|A^{-1}\mathbf{x}\|_\infty \leq \|A^{-1}(1, \dots, 1)^T\|_\infty \leq \frac{\|\mathbf{y}\|_\infty}{\min_i (A\mathbf{y})_i}$$

□

Wir versuchen nun das Resultat des Satzes auf unser Differenzenverfahren (1.2.2) anzuwenden. Aus der Struktur der Matrix ist direkt ersichtlich, dass L_h unter der Bedingung

$$\frac{1}{h^2} \geq \frac{|b_i|}{2h}, \quad i = 1, \dots, n$$

eine L_0 -Matrix ist. Eine hinreichende Bedingung dafür ist

$$h \leq \frac{2}{\max_{x \in \Omega} |b(x)|} = \frac{2}{\|b\|_{\infty, \Omega}} =: h_0$$

d.h. die Schrittweite h muss hinreichend klein sein. Ist das Ausgangsproblem konvektionsdominant, d.h. die Ausbreitungsgeschwindigkeit b ist groß, so muss h entsprechend klein gewählt werden.

Den Vektor \mathbf{y} aus Satz 1.77 konstruieren wir folgendermassen: aus der klassischen Theorie starker Lösungen weiß man, dass das Randwertproblem

$$\begin{cases} L_s u = 1 & \text{in } \Omega \\ u = 1 & \text{auf } \Gamma \end{cases}$$

mit $L_s = -u'' + bu'$ eine eindeutige Lösung besitzt, die dem Maximumprinzip genügt, d.h. $u(x) \geq \min_{y \in \Gamma} u(y) = 1$ für all $x \in \Omega$. Wir setzen wir $Lu = L_s + cu = 1 + cu = f$, wobei wegen $u \geq 1, c \geq 0$ offensichtlich $f \geq 1$ gilt. Dann ist u auch eine Lösung des Randwertproblems

$$\begin{cases} Lu = f & \text{in } \Omega \\ u = 1 & \text{auf } \Gamma \end{cases}$$

Wir definieren nun

$$\mathbf{y}_h = D_h(u)u = \begin{pmatrix} u(x_0) \\ \vdots \\ u(x_n) \end{pmatrix} \geq \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} > 0$$

Da u eine starke Lösung ist, gilt $u \in \mathcal{C}(\bar{\Omega})$ und $\|u\|_{\infty, \Omega} < \infty$. Man erhält demnach für \mathbf{y}_h die von h unabhängige Abschätzung

$$\|\mathbf{y}_h\|_{h, \infty} = \max_i |u(x_i)| \leq \|u\|_{\infty, \bar{\Omega}}$$

Weiter betrachtet man

$$L_h \mathbf{y}_h = L_h D_h(u)u = D_h(u)Lu + \mathbf{l}_h(u) = D_h(u)f + \mathbf{l}_h(u)$$

Da $f \geq 1$ und $u|_{\Gamma} = 1$ gilt

$$D_h(u)f = (u(x_0), f(x_1), \dots, f(x_{n-1}), u(x_n))^T \geq (1, \dots, 1)^T$$

Nehmen wir an, dass das Verfahren konsistent ist, so gilt $\|\mathbf{l}_h(u)\|_{h, \infty} \rightarrow 0$ für $h \rightarrow 0$. Für hinreichend kleines h , zum Beispiel $h < h_1$, gilt dann $-(1, \dots, 1)/2 < \mathbf{l}_h(u) < (1, \dots, 1)/2$ und daher

$$L_h \mathbf{y}_h = D_h(u)f + \mathbf{l}_h(u) > \frac{1}{2} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} > 0$$

für alle $h < h_1$.

Zusammenfassend können wir feststellen, dass das von uns konstruierte konsistente Verfahren (1.2.2) zweiter Ordnung unter den obigen Bedingungen stabil ist und damit konvergent ist.

SATZ 1.78. Sei $u \in \mathcal{C}^4$ eine glatte Lösung vom Randwertproblem

$$\begin{cases} Lu = f & \text{in } \Omega, \quad (c \geq 0) \\ u = g & \text{auf } \Gamma \end{cases}$$

Dann ist das Differenzenverfahren (1.2.2) zweiter Ordnung konvergent unter den oben angegebenen Bedingungen an die Schrittweite h .

BEMERKUNG 1.79. Stabilität eines Differenzenverfahren kann man auch über ein diskretes Maximumprinzip definieren. Wir verzichten hier auf die Einzelheiten und verweisen auf den Abschnitt über parabolische Gleichungen.

Statt des von uns angegebenen linearen Systems $L_h \mathbf{u}_h = \mathbf{f}_h$ kann man die vorgegebenen Dirichlet-Daten $g(0)$ und $g(1)$ auch direkt in das lineare System einbauen und erhält dann ein System der Ordnung $n - 1$ der Form

$$\begin{pmatrix} d_1 & r_1 & & & \\ l_2 & d_2 & r_2 & & \\ & \ddots & \ddots & \ddots & \\ & & l_{n-2} & d_{n-2} & r_{n-2} \\ & & & l_{n-1} & d_{n-1} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-2} \\ u_{n-1} \end{pmatrix} = \begin{pmatrix} f_1 - l_1 g_0 \\ f_2 \\ \vdots \\ f_{n-2} \\ f_{n-1} - r_{n-1} g_n \end{pmatrix}$$

Die zugehörige Systemmatrix \tilde{L}_h ist dann, im Gegensatz zum Ausgangsmatrix, häufig symmetrisch, was für numerische Verfahren von Vorteil sein kann. Insbesondere erhält man

für den Spezialfall $b(x) = c(x) = 0$ die für elliptische Gleichungen klassische symmetrische Systemmatrix

$$\tilde{L}_h = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}$$

Die Konvergenz- und Stabilitätsaussagen bleiben davon unberührt.

2.2. Differenzenverfahren in zwei Raumdimensionen. Der Fall von zwei Raumdimensionen ist im Wesentlichen wegen zweier Eigenschaften komplizierter als der eindimensionale Fall:

- 1) Elliptische Operatoren in zwei Raumdimensionen sind komplizierter und können insbesondere gemischte Ableitungen enthalten.
- 2) Das Gebiet $\Omega \subset \mathbb{R}^2$ kann kompliziert sein.

Wir betrachten daher zunächst kurz ein einfaches Modellproblem, nämlich die Poisson-Gleichung im Einheitsquadrat mit Dirichlet-Daten auf dem Rand,

$$\begin{cases} -\Delta u = f & \text{in } \Omega = (0, 1)^2 \\ u = g & \text{auf } \Gamma = \partial\Omega \end{cases}$$

Wir überziehen das Einheitsquadrat Ω und den Rand Γ mit einem äquidistanten Gitter in beiden Richtungen, d.h.

$$x_i = ih, \quad y_j = jh, \quad i, j = 0, \dots, n, \quad h = \frac{1}{n}$$

und bezeichnen wie im vorigen Abschnitt die Menge der inneren Gitterpunkte mit Ω_h , die Menge der Gitterpunkte auf dem Rand mit Γ_h . Weiter bezeichnet u_{ij} die numerische Approximation der exakten Lösung $u(x_i, y_j)$ für $i, j = 1, \dots, n-1$ und wir verwenden folgende Differenzenquotienten auf Ω_h

$$\begin{aligned} u_{xx}(x_i, y_i) &\sim \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} \\ u_{yy}(x_i, y_i) &\sim \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{h^2} \end{aligned}$$

sodass

$$\Delta u(x_i, y_i) \sim \frac{-u_{i-1,j} - u_{i,j-1} + 4u_{i,j} - u_{i,j+1} - u_{i+1,j}}{h^2}$$

Das Verfahren basiert also auf einem 5-Punkte-Stern

$$\begin{array}{c} x \\ x \quad x \quad x \\ x \end{array}$$

wobei man für die Gewichte auf dem Stern auch die Notation

$$\frac{1}{h^2} \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

verwendet. Wir erhalten wiederum für jeden inneren Punkt aus Ω_h eine lineare Gleichung und für jeden Randpunkt aus Γ_h aufgrund der vorgegebenen Dirichlet-Daten eine Zusatzbedingung, die man auch als lineare Gleichung lesen kann. Die Anzahl der Gleichungen stimmt dabei wieder mit der Anzahl der Unbekannten überein.

Im Gegensatz zum eindimensionalen Fall ist aber nicht direkt einsichtig, wie die resultierende Struktur der Systemmatrix aussehen wird, da im Gegensatz zum Eindimensionalen auf dem \mathbb{R}^2 keine natürliche Ordnung existiert. Man kann also etwa eine zeilenweise Anordnung der Gitterpunkte verwenden, d.h. man ordnet die Gitterpunkte zunächst in x -Richtung,

$$\mathbf{u}_h = (u_{00}, \dots, u_{n0}, u_{01}, \dots, u_{nn})^T$$

oder aber man ordnet die Gitterpunkte spaltenweise an, d.h. zunächst in y -Richtung,

$$\mathbf{u}_h = (u_{00}, \dots, u_{0n}, u_{10}, \dots, u_{nn})^T$$

Daneben existieren eine Reihe von anderen Anordnungsprinzipien, etwa das Schachbrettmuster oder ähnliches.

Offensichtlich hängt die Systemmatrix stark von der verwendeten Anordnung der Gitterpunkte ab und dies kann zum Beispiel dazu verwendet werden, um numerische Lösungsverfahren für lineare System möglichst effizient zu machen. Für die zeilenweise Anordnung erhält man zum Beispiel eine Systemmatrix in der Form

$$L_h = \begin{pmatrix} I & & & & & \\ B & A & B & & & \\ & \ddots & \ddots & \ddots & & \\ & & B & A & B & \\ & & & & I & \end{pmatrix}$$

mit $(n+1) \times (n+1)$ -Matrizen der Form

$$B = \frac{1}{h^2} \begin{pmatrix} 0 & & & & & \\ & -1 & & & & \\ & & \ddots & & & \\ & & & -1 & & \\ & & & & -1 & \\ & & & & & 0 \end{pmatrix}, A = \frac{1}{h^2} \begin{pmatrix} h^2 & & & & & \\ -1 & 4 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & -1 & 4 & -1 \\ & & & & & h^2 \end{pmatrix}$$

und I der Einheitsmatrix der Ordnung $(n+1)$. Die rechte Seite des Systems $L_h \mathbf{u}_h = \mathbf{f}_h$ ist dann gegeben durch

$$\mathbf{f}_h = (g_{00}, \dots, g_{n0}, g_{01}, f_{11}, \dots, f_{n-1,1}, g_{n,1}, g_{02}, f_{12}, \dots, f_{n-1,n-1}, g_{n,n-1}, g_{0n}, \dots, g_{nn})^T$$

Wie im eindimensionalen Fall kann man die gegebenen Dirichlet-Daten wieder aus dem linearen Gleichungssystem eliminieren und erhält gewöhnlich eine symmetrische Systemmatrix \tilde{L}_h .

Für die Poissongleichung $-\Delta u = f$ lassen sich die Aussagen über Konsistenz, Stabilität und Konvergenz für eindimensionale Probleme direkt übertragen. Die oben angegebenen Differenzenquotienten liefern dann unter den entsprechenden Bedingungen an die Schrittweite h ein konvergentes Verfahren zweiter Ordnung. Dabei sind die resultierenden linearen Gleichungssysteme für die zeilen- und spaltenweise Anordnung identisch.

Wir betrachten nun das allgemeine Dirichlet-Problem auf dem Einheitsquadrat gegeben durch die Gleichung

$$(1.2.6) \quad Lu = -(a_{11}u_{xx} + (a_{21} + a_{12})u_{xy} + a_{22}u_{yy}) + b_1u_x + b_2u_y + cu$$

wobei A eine symmetrische, positiv definite Matrix ist,

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$$

Insbesondere gilt dann $a_{11}, a_{22} > 0$. Gilt $a_{12} \neq 0$, so benötigen wir geeignete Differenzenquotienten für die gemischte Ableitung $u_{xy}(x, y)$. Der einfachste Ansatz basiert auf zentralen Differenzen in jeder Koordinatenrichtung, sodass wir folgende Approximation zweiter Ordnung erhalten

$$\begin{aligned} u_{xy}(x_i, y_j) &\approx \left(\frac{u_{i+1,j+1} - u_{i+1,j-1}}{2h} - \frac{u_{i-1,j+1} - u_{i-1,j-1}}{2h} \right) / 2h \\ &= \frac{1}{4h^2} (u_{i+1,j+1} - u_{i+1,j-1} - u_{i-1,j+1} + u_{i-1,j-1}) \end{aligned}$$

Diese Approximation kombinieren wir mit einer Zerlegung der restlichen Ableitungen in zwei unabhängige Differenzenquotienten in x - und y -Richtung analog zu Abschnitt 2.1. Damit erhält man einen 9-Punkte-Stern mit Gewichtungsmatrix

$$\begin{bmatrix} +m_{ij} & r_{ij}^y & -m_{ij} \\ l_{ij}^x & d_{ij}^x + d_{ij}^y & r_{ij}^x \\ -m_{ij} & l_{ij}^y & +m_{ij} \end{bmatrix}$$

und

$$\begin{aligned} d_{ij}^x &= \frac{2a_{11}}{h^2} + c & d_{ij}^y &= \frac{2a_{22}}{h^2} + c \\ l_{ij}^x &= -\frac{a_{11}}{h^2} - \frac{b_1}{2h} & l_{ij}^y &= -\frac{a_{22}}{h^2} - \frac{b_2}{2h} \\ r_{ij}^x &= -\frac{a_{11}}{h^2} + \frac{b_1}{2h} & r_{ij}^y &= -\frac{a_{22}}{h^2} + \frac{b_2}{2h} \end{aligned}$$

sowie

$$m_{ij} = \frac{a_{12}}{2h^2}$$

Dabei können alle Koeffizienten der linearen Differentialgleichung von den Gitterpunkten (x_i, y_j) abhängen. Offensichtlich kann die resultierende Systemmatrix L_h keine L_0 -Matrix sein, denn wir haben wechselnde Vorzeichen in den Nicht-Diagonalelementen. Es stellt sich also die Frage, ob man andere Ansätze finden kann, die gewährleisten, dass die Systemmatrix eine L_0 -Matrix ist.

Man kann zeigen, dass dies unter Verwendung einer einzigen konsistenten Approximation der gemischten Ableitung U_{xy} unmöglich ist. Wir müssen daher mindestens zwei

konsistente Approximationen miteinander kombinieren: ist $a_{12}(x_i, y_j) \leq 0$, so setzen wir zunächst als Approximation der gemischten Ableitung Approximation der gemischten Ableitung

$$(1.2.7) \quad u_{xy}(x_i, y_j) \approx \frac{1}{2} (u_{xy}(x_{i-1/2}, y_{j+1/2}) + u_{xy}(x_{i+1/2}, y_{j-1/2}))$$

Für $a_{12}(x_i, y_j) > 0$ verwenden wir

$$(1.2.8) \quad u_{xy}(x_i, y_j) \approx \frac{1}{2} (u_{xy}(x_{i+1/2}, y_{j+1/2}) + u_{xy}(x_{i-1/2}, y_{j-1/2}))$$

Dann setzen wir für jeden der Terme in (1.2.7) und (1.2.8) zentrale Differenzen mit Gitterweite $h/2$ an: für (1.2.7) bedeutet dies

$$\begin{aligned} & \frac{1}{2} (u_{xy}(x_{i-1/2}, y_{j+1/2}) + u_{xy}(x_{i+1/2}, y_{j-1/2})) \\ & \approx \frac{1}{2h} (u_x(x_{i-1/2}, y_{j+1}) - u_x(x_{i-1/2}, y_j) + u_x(x_{i+1/2}, y_j) - u_x(x_{i+1/2}, y_{j-1})) \\ & \approx \frac{1}{2h^2} (u_{i,j+1} - u_{i-1,j+1} - u_{i,j} + u_{i-1,j} + u_{i+1,j} - u_{i,j} - u_{i+1,j-1} + u_{i,j-1}) \end{aligned}$$

und die zugehörige Gewichtungsmatrix lautet

$$\frac{1}{2h^2} \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

Dementsprechend erhalten wir als Gewichtungsmatrix für (1.2.8)

$$\frac{1}{2h^2} \begin{bmatrix} 0 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 0 \end{bmatrix}$$

Beide Approximationen sind zweiter Ordnung.

Insgesamt verwenden wir also für L eine 9-Punkte Stern der Form

$$\begin{bmatrix} \frac{a_{12}(x_i, y_j)}{h^2} & * & 0 \\ * & * & * \\ 0 & * & \frac{a_{12}(x_i, y_j)}{h^2} \end{bmatrix}$$

für $a_{12}(x_i, y_j) \leq 0$ und

$$\begin{bmatrix} 0 & * & -\frac{a_{12}(x_i, y_j)}{h^2} \\ * & * & * \\ -\frac{a_{12}(x_i, y_j)}{h^2} & * & 0 \end{bmatrix}$$

für $a_{12}(x_i, y_j) > 0$. Die Einträge in den Ecken sind offensichtlich größer gleich Null. Allerdings ändern sich auch die restlichen Werte und wir müssen überprüfen, dass diese weiterhin größer gleich Null sind. Für die modifizierten Koeffizienten \tilde{l}_{ij}^x erhalten wir

$$\tilde{l}_{ij}^x = -\frac{a_{11}}{h^2} - \frac{b_1}{2h} - \frac{a_{12}}{h^2}$$

Man kann allerdings leicht zeigen, dass die resultierende Systemmatrix wieder eine L_0 -Matrix ist und die Stabilitätseigenschaft erhalten bleibt. Damit ist das Verfahren erster Ordnung konvergent. Eine genauere Analyse liefert für den lokalen Diskretisierungsfehler die Abschätzung

$$\|l_n\|_{h,\infty} \leq \alpha h^n + \beta h^{2+m}$$

mit den beiden Konstanten α und β , wobei h den maximalen Abstand zwischen zwei Randpunkten bezeichnet und n bzw. m die Konsistenzordnungen entfernt und nahe zum Rand sind. Damit ist das Verfahren weiterhin zweiter Ordnung, solange mit einer Konstanz erster Ordnung am Rand ansetzt.

Zum Abschluss geben wir noch kurz an, wie Neumann-Randbedingungen $u_n = g$ auf dem Rand Γ umgesetzt werden. Dazu verwendet man etwa eine Approximation erster Ordnung der Form

$$\partial_n u \approx \frac{u_{n,j} - u_{n-1,j}}{h}, \quad j = 0, \dots, n$$

oder eine Approximation zweiter Ordnung, etwa

$$\partial_n u \approx \frac{3u_{n,j} - 4u_{n-1,j} + u_{n-2,j}}{2h}$$

die man aus einer Taylor-Entwicklung am Rand ableitet. Man sieht leicht, dass die Approximation erster Ordnung im Gegensatz zur Approximation zweiter Ordnung keine L_0 -Matrix liefert.

Bei der Umsetzung von Neumann-Randbedingungen treten zusätzliche Probleme auf:

- 1) Ein Randwertproblem mit Neumann-Randbedingungen besitzt keine eindeutige Lösung, da man zu jeder gegebenen Lösung eine Konstante addieren kann und wiederum eine Lösung erhält. In der Numerik bedeutet dies, dass die Matrix L_h singular ist. Die einfachste Möglichkeit dies zu umgehen ist, einen Wert am Rand vorzuschreiben. Dadurch wird eine Zeile in der Systemmatrix gestrichen und die resultierende Matrix ist regulär.
- 2) Die Methode zur Untersuchung der Stabilität bei Dirichletproblemen lässt sich nicht auf den Fall von Neumann-Randbedingungen übertragen. Insbesondere greift in diesem Fall nicht die verwendete Klassifikation einer M -Matrix. Mit Hilfe anderer Methoden kann man allerdings wieder die Stabilität und damit die Konvergenz des Verfahrens beweisen.