KAPITEL 2

Die Methode der Finiten Elemente

1. Theoretische Grundlagen

Wir bezeichnen im Folgenden mit $H^m(\Omega) \subset L^2(\Omega)$, $\Omega \subset \mathbb{R}^n$ offen, den Sobolevraum aller Funktionen mit schwachen Ableitungen $\partial^\alpha u$ in $L^2(\Omega)$ für alle $|\alpha| \leq m$,

$$H^m(\Omega) = \{u | u \in L^2(\Omega), \mathrm{D}^\alpha u \in L^2(\Omega), |\alpha| \leq m\}$$

Der Raum $H^m(\Omega)$ ist ein Hilbertraum mit Skalarprodukt

$$\langle u, v\rangle_{H^m} = \langle u, v\rangle_m = \sum_{|\alpha|\leq m} \langle \mathrm{D}^\alpha u, \mathrm{D}^\alpha v\rangle_{L^2}$$

und der Sobolevnorm

$$\|u\|_{H^m(\Omega)} = \|u\|_m = (\langle u, u\rangle_m)^{1/2}$$

Weiter definieren wir die Halbnorm

$$|u|_m = \left(\sum_{|\alpha|=m} \langle \mathrm{D}^\alpha u, \mathrm{D}^\alpha u\rangle_{L^2(\Omega)} \right)^{1/2}$$

Die Vervollständigung von $\mathcal{C}_0^\infty(\Omega)$ bezüglich der Norm $\|\cdot\|_m$ bezeichnen wir mit $H_0^m(\Omega)$. Die Poincare–Friedrichs Ungleichung lautet dann

SATZ 2.1. *Sei Ω enthalten in einem Würfel in $\mathbb{R}^n$ mit Kantenlänge s. Dann gilt*

$$\|v\|_0 < s|v|_1$$

für alle Funktionen $v \in H_0^1(\Omega)$.

KOROLLAR 2.2. *Sei Ω beschränkt und enthalten in einem Würfel mit Kantenlänge s in $\mathbb{R}^n$. Dann gilt*

$$|v|_m \leq \|v\|_m \leq (1+s)^m |v|_m \qquad \forall v \in H_0^m(\Omega)$$

d.h. für beschränkte Gebiete Ω sind die beiden Normen $|\cdot|_m$ und $\|\cdot\|_m$ in $H_0^m(\Omega)$ äquivalent.

Zur Herleitung der Finite–Element Methode verwendet man die schwache Formulierung von Randwertaufgaben für elliptische Differentialgleichungen. Wir betrachten also das Dirichletproblem

$$\begin{cases} Lu = f & \text{in } \Omega \\ u = g & \text{auf } \partial\Omega \end{cases} \tag{2.1.1}$$

wobei Lu durch

$$Lu = -\sum_{i,j=1}^{n} \partial_i(a_{ij}\partial_j u) + a_0 u \tag{2.1.2}$$

mit $a_0 \geq 0$ gegeben ist.
Der Differentialoperator L heißt elliptisch, falls die Matrix $A = (a_{ij})$ positiv definit ist. Weiter nennt man L gleichmäßig elliptisch, falls L elliptisch ist und ein $\alpha > 0$ existiert, sodass

$$\rho^T A(x)\rho \geq \alpha \rho^T \rho$$

wobei $\rho \in \mathbb{R}^n$, $x \in \Omega$ und $A(x) = (a_{ij}(x))$.
Dirchletprobleme der Form (2.1.1) lassen sich stets auf Probleme mit homogenen Randbedingungen zurückführen: definiert man

$$w = u - u_0$$

wobei u_0 eine Funktion ist, die auf $\partial\Omega$ die Randbedingung $u_0 = g$ erfüllt. Dann löst w in Ω die elliptische Gleichung $Lw = \tilde{f} = f - Lu_0$ mit homogenen Randbedingungen $w = 0$ auf $\partial\Omega$.

Satz 2.3. *Jede starke Lösung von $Lu = f$ in Ω, $u = 0$ auf $\partial\Omega$, wobei L ein elliptischer Operator der Form (2.1.2) ist, löst gleichzeitig das Variationsproblem*

$$I(v) = \int_\Omega \left(\frac{1}{2} \sum_{i,j} a_{ij}\partial_i v \partial_j v + a_0 v^2 - fv \right) \mathrm{d}\mathbf{x} \to \min$$

unter allen Funktionen $v \in \mathcal{C}^2(\Omega) \cap \mathcal{C}^0(\bar{\Omega})$ mit homogenen Randbedingungen, d.h. $v|_{\partial\Omega} = 0$.

Das Funktional $I(v)$ läßt sich auch in der Form

$$I(v) = \frac{1}{2} a(v, v) - l(v)$$

darstellen, wobei

$$a(u, v) = \int_\Omega \left(\sum_{i,j} a_{ij}\partial_i u \partial_j v + a_0 uv \right) \mathrm{d}\mathbf{x}$$

die zum Operator L zugehörige Bilinearform ist und

$$l(v) = \int_\Omega fv dx$$

als ein Element des Dualraums interpretiert werden kann.
Die Aussage in Satz 2.3 bezeichnet man auch als Dirichletprinzip. Dieses Prinzip läßt sich auch in einer abstrakten Form formulieren:

Definition 2.4. *Eine Bilinearform $a : H \times H \to \mathbb{R}$ auf einem Hilbertraum H heißt* stetig, *falls ein $C > 0$ existiert, sodass für alle $u, v \in H$ gilt*

$$|a(u, v)| \leq C \cdot \|u\| \|v\|$$

Eine symmetrische und stetige Bilinearform a nennt man H–elliptisch, falls mit $\alpha > 0$ gilt

$$a(v,v) \geq \alpha \|v\|^2, \quad v \in H$$

Die Norm $\|v\|_a := (a(v,v))^{1/2}$ bezüglich einer H–elliptischen Bilinearform a nennt man auch Energienorm.

Dann gilt der Satz von Lax und Milgram

SATZ 2.5. *Sei V abgeschlossen und konvex in H. Die Bilinearform a sei H–elliptisch. Dann besitzt das Variationsproblem*

$$I(v) = \frac{1}{2}a(v,v) - l(v) \to \min \tag{2.1.3}$$

für alle l aus dem Dualraum H' von H eine eindeutige Lösung in V.

Neben dem abstrakten Prinzip im Lax–Milgram Satz läßt sich das Minimum u des Funktionals $I(v)$ in (2.1.3) noch folgendermaßen charakterisieren:

SATZ 2.6. *Sei V ein linearer Raum und $a : V \times V \to \mathbb{R}$ eine Bilinearform mit $a(v,v) > 0$ für alle $v \in V$, $v \neq 0$ und $l : V \to \mathbb{R}$ ein lineares Funktional. Dann ist u das Minimum von*

$$I(v) = \frac{1}{2}a(v,v) - l(v)$$

genau dann, wenn für alle $v \in V$ gilt: $a(u,v) = l(v)$.

Mit Hilfe des Charakterisierungssatzes kann man in $H_0^1(\Omega)$ schwache Lösungen von elliptischen Problemen mit homogenen Randbedingungen definieren.

DEFINITION 2.7. *Die Funktion $u \in H_0^1(\Omega)$ ist eine schwache Lösung des elliptischen Randwertproblems*

$$\left\{ \begin{array}{rcll} Lu & = & f & \text{in } \Omega \\ u & = & 0 & \text{auf } \partial\Omega \end{array} \right.$$

falls für die zugehörige Bilinearform a gilt

$$a(u,v) = l(v) = \langle f, v \rangle_{L^2(\Omega)} \quad \forall\, v \in H_0^1(\Omega)$$

SATZ 2.8. *Der Differentialoperator L zweiter Ordnung gegeben durch (2.1.2) sei gleichmäßig elliptisch und $a_0 > 0$. Dann existiert in $H_0^1(\Omega)$ eine schwache Lösung des zugehörigen Dirchletproblems mit homogenen Randbedingungen. Gleichzeitig erfüllt diese Lösung in $H_0^1(\Omega)$ das Variationsproblem*

$$\frac{1}{2}a(v,v) - \langle f, v \rangle_{L^2(\Omega)} \to \min$$

Man beweist diesen Satz durch Anwendung des Satzes von Lax und Milgram.
Bei Problemen mit inhomogenen Randbedingungen verwendet man eine zu Definition 2.7 äquivalente schwache Formulierung, in dem man das Problem auf homogene Randbedingungen zurückführt: sei $u_0 \in \mathcal{C}^2(\Omega) \cap \mathcal{C}^0(\bar{\Omega}) \cap H^1(\Omega)$ eine Funktion, die auf $\partial\Omega$ die Randbedingung $u_0 = g$ erfüllt. Dann suchen wir in der schwachen Formulierung eine Funktion $w \in H_0^1(\Omega)$, sodass für alle $v \in H_0^1(\Omega)$ gilt

$$a(w,v) = \langle f - Lu_0, v \rangle_{L^2(\Omega)}$$

Mittels partieller Integration erhält man die Beziehung $a(u_0, v) = \langle Lu_0, v\rangle_{L^2(\Omega)}$ und die schwache Form läßt sich folgendermaßen angeben: finde eine Funktion $u \in H^1(\Omega)$, sodass

$$a(u, v) = \langle f, v\rangle_{L^2(\Omega)}$$

für alle $v \in H_0^1(\Omega)$ gilt und $u - u_0$ in $H_0^1(\Omega)$ ist. Die Bedingung $u - u_0 \in H_0^1(\Omega)$ kann man auch als schwache Form der Randbedingung $u|_{\partial\Omega} = g$ interpretieren.
Wir erhalten also folgende Definition einer schwachen Lösung des Dirichletproblms mit inhomogenen Randbedingungen.

DEFINITION 2.9. *Die Function $u \in H^1(\Omega)$ ist eine schwache Lösung des Randwertproblems*

$$\begin{cases} Lu &= f \quad \text{in } \Omega \\ u &= g \quad \text{auf } \partial\Omega \end{cases}$$

falls für alle $v \in H_0^1(\Omega)$ gilt

$$a(u, v) = \langle f, v\rangle_{L^2(\Omega)}$$

und

$$u - u_0 \in H_0^1(\Omega)$$

wobei $u_0 \in \mathcal{C}^2(\Omega) \cap \mathcal{C}^0(\bar{\Omega}) \cap H^1(\Omega)$ eine Funktion mit $u_0|_{\partial\Omega} = g$ ist.

2. Die Ritz–Galerkin Methode

In diesem Abschnitt diskutieren wir die Ritz–Galerkin Methode zur numerischen Lösung des Variationsproblems

$$I(v) = \frac{1}{2}a(v, v) - l(v) \to \min! \tag{2.2.4}$$

in $H^m(\Omega)$ oder $H_0^m(\Omega)$, beziehungsweise das äquivalente Problem

$$a(u, v) = l(v)$$

für alle $v \in H^m(\Omega)$ oder $H_0^m(\Omega)$.
In der Ritz–Galerkin Methode sucht man das Minimum in (2.2.4) nicht im Funktionenraum $H^m(\Omega)$ oder $H_0^m(\Omega)$, sondern in einem endlichdimensionalen Unterraum S_h, dem Finite–Elementraum, wobei der Parameter h einen Gitterparameter zur Diskretisierung des Ortsraums bezeichnet. Anschliessend untersucht man die Konvergenz der numerischen Lösung in S_h gegen eine Lösung in $H^m(\Omega)$ beziehungsweise $H_0^m(\Omega)$ im Grenzfall $h \to 0$.
Wir wissen bereits, dass das Variationsproblem

$$I(v) = \frac{1}{2}a(v, v) - l(v) \to \min_{S_h}!$$

in S_h lösbar ist: u_h ist eine Lösung, falls für alle $v \in S_h$ gilt

$$a(u_h, v) = l(v) \tag{2.2.5}$$

Sei $\{\psi_1, \psi_2, ..., \psi_N\}$ eine Basis des endlichdimensionalen FE–Raums S_h, dann sind die Gleichungen

$$a(u_h, \psi_i) = l(\psi_i), \quad i = 1, \ldots, N$$

äquivalent zu (2.2.5). Zusammen mit dem Ansatz

$$u_h = \sum_{j=1}^{N} z_j \psi_j$$

erhalten wir das Gleichungssystem

$$\sum_{j=1}^{N} a(\psi_j, \psi_i) z_j = l(\psi_i), \quad i = 1, \ldots, N$$

das man auch in Matrixform

$$Az = b$$

schreiben kann. Die Matrix A ist dabei gegeben durch $A = (a_{ij}) = a(\psi_j, \psi_i)$, die rechte Seite gegeben durch $b_i = l(\psi_i)$.
Ist die Bilinearform a H–elliptisch, so ist die Matrix A positiv definit, denn es gilt

$$z^T A z = \sum_{i,j} z_i A_{ij} z_j = a(\sum_j z_j \psi_j, \sum_i z_i \psi_i) = a(u_h, u_h) \geq \alpha \|u_h\|_m^2$$

Die Matrix A wird gewöhnlich als Steifigkeitsmatrix bezeichnet.
Die Qualität der numerischen Approximation u_h im FE–Raum S_h läßt sich mittels des Céa–Lemmas zeigen.

LEMMA 2.10. *Sei a eine V–elliptische Bilinearform, $V = H^m$ oder $V = H_0^m$ und seien u und u_h die Lösungen des zugehörigen Variationsproblems in V und $S_h \subset V$. Dann gilt*

$$\|u - u_h\|_m \leq \frac{C}{\alpha} \inf_{v_h \in S_h} \|u - v_h\|_m$$

BEWEIS. Nach Definition gilt

$$\begin{aligned} a(u, v) &= l(v) \quad \text{for } v \in V \\ a(u_h, v) &= l(v) \quad \text{for } v \in S_h \end{aligned}$$

Da $S_h \subset V$, können wir die beiden Gleichung subtrahieren und erhalten

$$a(u - u_h, v) = 0 \quad \text{for } v \in S_h$$

Sei $v_h \in S_h$. Dann gilt mit $v = v_h - u_h \in S_h$

$$a(u - u_h, v_h - u_h) = 0$$

und

$$\begin{aligned} \alpha \|u - u_h\|_m^2 &\leq a(u - u_h, u - u_h) \\ &= a(u - u_h, u - v_h) + \underbrace{a(u - u_h, v_h - u_h)}_{=0} \\ &\leq C \|u - u_h\|_m \|u - v_h\|_m \end{aligned}$$

Daraus folgt aber $\alpha \|u - u_h\|_m \leq C \|u - v_h\|_m$ für alle $v_h \in S_h$. □

Das Lemma zeigt, dass die Genauigkeit der numerischen Approximation sehr stark von der Wahl des FE–Raumes S_h abhängt. Der FE–Raum wird gewöhnlich mit polynomialen Approximationen identifiziert, wobei man allerdings keinen Polynome mit beliebig hohem Grad, sondern stückweise definierte Polynome niedrigen Grades verwendet. Die Genauigkeit kann dann auch über die Wahl des Diskretisierungsparameters h des FE–Raums S_h gesteuert werden.
Also eine erste Anwendung der Methode der Finiten–Elemente betrachten wir folgendes Modellproblem: auf dem Einheitsquadrat sei die Poissongleichung mit homogenen Dirichlet–Randbedingungen gegeben,

$$\begin{cases} -\Delta u &= f & \text{in } \Omega = [0,1]^2 \\ u &= 0 & \text{auf } \partial\Omega \end{cases}$$

Wir konstruieren zunächst die in Bild 3.1 angegebene Triangulierung des Einheitsquadrats und betrachten dann den FE–Raum S_h aller Funktionen $v \in \mathcal{C}^0([0,1]^2)$, die auf den gegebenen Dreiecken lineare Funktionen sind und die Bedingung $v = 0$ auf dem Rand $\partial\Omega$ erfüllen.

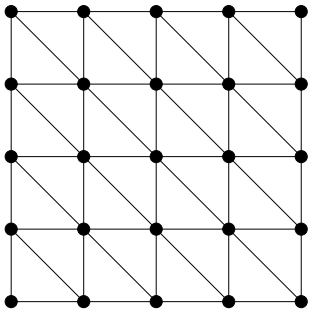

Bild 3.1 Triangulierung des Einheitsquadrats.

Als lineare Funktion besitzt $v \in S_h$ auf jedem Dreieck die Darstellung

$$v(x,y) = a + bx + cy$$

und die drei Koeffizienten a, b und c lassen sich etwa durch die Funktionswerte an den Ecken des Dreiecks festlegen. Die Ecken der Dreiecke sind dabei gerade die Giiterpunkte der Triangulierung. Damit ist die Dimension des Funktionenraums S_h auf Grund der vorgegebenen homogenen Dirichlet–Daten gerade gleich N, der Anzahl der inneren Gitterpunkte. Die Funktion v ist gleichzeitig eindeutig durch ihre Werte an den N Gitterpunkten (x_j, y_j), $j = 1, \ldots, N$, bestimmt.
Als Basis des FE–Raums S_h wählen wir die (bezüglich der gegebenen Triangulierung) stückweiese linearen Ansatzfunktionen $\{\psi_1, \ldots, \psi_N\}$, die an den inneren Gitterpunkten die Bedingungen $\psi_i(x_j, y_j) = \delta_{ij}$ genügen.
Verwenden wir die Anordnung wie in Bild 3.2 und setzen $C = (0,0)$, so erhalten wir als

Darstellung der Basisfunktion ψ_C mit $N = (0, h)$

$$\psi_C(x,y) = \begin{cases} 1 - x/h - y/h & (x,y) \in I \\ 0 & (x,y) \in II \\ 1 + x/h & (x,y) \in III \\ 1 - y/h & (x,y) \in IV \\ 0 & (x,y) \in V \\ 1 + x/h + y/h & (x,y) \in VI \\ 1 + y/h & (x,y) \in VII \\ 1 - x/h & (x,y) \in VIII \end{cases}$$

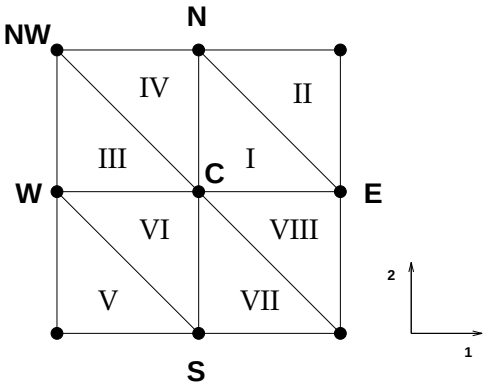

Bild 3.2 Anordnung der Dreiecke um den Mittelpunkt C.

Die Ableitung der Basisfunktion ψ_C auf den acht Dreiecken ist in Tabelle 3.1 angegeben.

Tabelle 3.1 Ableitungen der Basisfunktion ψ_C

	I	II	III	IV	V	VI	VII	VIII
$\partial_x \psi_C$	$-h^{-1}$	0	h^{-1}	0	0	h^{-1}	0	$-h^{-1}$
$\partial_y \psi_C$	$-h^{-1}$	0	0	$-h^{-1}$	0	h^{-1}	h^{-1}	0

Damit erhalten wir für die Auswertung der zur Poissongleichung gehörenden Bilinearform

$$\begin{aligned} a(\psi_C, \psi_C) &= \int_{I,\ldots,VIII} (\nabla \psi_C)^2 dxdy \\ &= 2 \int_{I+III+IV} \left((\partial_x \psi_C)^2 + (\partial_y \psi_C)^2 \right) dxdy \\ &= 2 \int_{I+III} (\partial_x \psi_C)^2 dxdy + 2 \int_{I+IV} (\partial_y \psi_C)^2 dxdy \\ &= 2h^{-2} \int_{I+III} dxdy + 2h^{-2} \int_{I+IV} dxdy = 4 \end{aligned}$$

sowie

$$\begin{aligned} a(\psi_C, \psi_N) &= \int\limits_{I+IV} (\nabla\psi_C) \cdot (\nabla\psi_N) dxdy \\ &= \int\limits_{I+IV} \partial_y\psi_C \partial_y\psi_N dxdy \\ &= \int\limits_{I+IV} (-h^{-1})h^{-1} dxdy = -1 \end{aligned}$$

Aus Symmetriegründen gilt dann auch

$$a(\psi_C, \psi_E) = a(\psi_C, \psi_S) = a(\psi_C, \psi_W) = a(\psi_C, \psi_N) = -1$$

Für die restlichen Eckpunkte gilt

$$a(\psi_C, \psi_{NW}) = \int\limits_{III+IV} (\partial_1\psi_C \partial_1\psi_{NW} + \partial_2\psi_C \partial_2\psi_{NW})\, \mathrm{d}x\mathrm{d}y = 0$$

und $a(\psi_C, \psi_{NW}) = a(\psi_C, \psi_{NW}) = a(\psi_C, \psi_{NW}) = 0$. Insgesamt erhalten wir also wie bei der Methode der finiten Differenzen den gewöhnlichen 5–Punktestern des Laplace–Operators

$$\begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}$$

3. Einige Beispiele zu Finite–Element Räumen

Im letzten Abschnitt hatten wir stückweise lineare $\mathcal{C}^0$–Funktionen auf der Basis einer Triangulierung des Einheitsquadrats $[0,1]^2$ verwendet. Es gibt eine Reihe von anderen Möglichkeiten Fe–Räume zu konstruieren. Im Allgemeinen benötigt man die folgenden Zutaten:

(1) Eine Partition des Rechengebiets Ω: hier ist die zentrale Frage, welche Elemente man bei der Zerlgung verwendet. Bei unserem Modellproblem aus dem letzten Abschnitt haben wir eine Triangulierung verwendet, d.h. die Basiselemente waren Dreiecke. Andere Möglichkeiten wären Rechtecke, in höheren Dimensionen allgemeine Polyeder.

(2) Die Auswahl der Basisfunktionen des FE–Raums: hier verwendet man gewöhnlich Polynome vom Grad kleiner gleich t, die jeweils stückweise auf den vorgegebenen Elementen definiert sind. Sei also (im Fall zweier Raumdimensionen)

$$\mathcal{P}_t = \left\{ u(x,y) = \sum_{i,j\geq 0,\, i+j\leq t} c_{ij} x^i y^j \right\}$$

Verwendet man für alle Elemente Polynome mit Grad kleiner gleich t, so bezeichnet man die Methode als Finite–Elemente mit vollständigen Polynomen.

(3) Die Festlegung von Stetigkeits– und Differenzierbarkeitseigenschaften: Finite Elemente nennt man $\mathcal{C}^k$–Elemente, falls die numerische Approximation global in $\mathcal{C}^k(\Omega)$ liegt.

Ein zentraler Begriff zur Partition des Rechengebiets ist der einer zulässigen Partition.

DEFINITION 2.11. *Eine Partition oder Zerlegung* $\mathcal{T} = \{T_1, T_2, ..., T_M\}$ *des Gebiets* Ω *in Elemente heißt* zulässig, *falls gilt:*

1) $\bar{\Omega} = \bigcup_{i=1}^{M} T_i$
2) *Besteht* $T_i \cap T_j$ *aus nur einem Punkt, so ist der Punkt ein Eckpunkt von* T_i *und* T_j
3) *Besteht* $T_i \cap T_j$ *mit* $i \neq j$ *aus mehr als einem Punkt, dann ist* $T_i \cap T_j$ *eine Kante von* T_i *und* T_j.

Eine nicht–zulässige Partition ist in Bild 3.3 dargestellt.

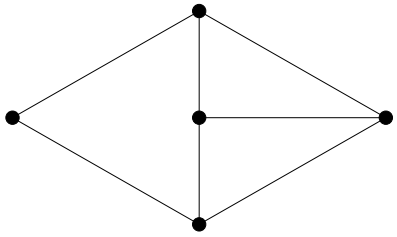

Bild 3.3 Nicht–zulässige Zerlegung bestehend aus drei Dreiecken.

In diesem Fall ist die Bedingung 2) aus Definition 2.11 verletzt: Die Zerlegung in Bild 3.3 besitzt einen hängenden Knoten, der bezüglich des linken Dreiecks im Inneren einer Kante liegt. Wir schreiben die Zerlegung auch als $\mathcal{T}_h$ statt $\mathcal{T}$, falls jedes Element in $\mathcal{T}_h$ einen Durchmesser kleiner als $2h$ besitzt.

DEFINITION 2.12.

1) *Eine Partition $\mathcal{T}_h$ heißt* quasi–gleichförmig, *falls ein $k > 0$ existiert, sodass jedes Element $T \in \mathcal{T}$ einen Kreis mit Radius ρ_T enthält, wobei $\rho_T \geq \frac{\tilde{h}_T}{k}$ und $\tilde{h}_T$ den halben Durchmesser von T bezeichnet.*
2) *Eine Zerlegung $\mathcal{T}_h$ heißt gleichförmig, falls ein $k > 0$ existiert, sodass jedes Element $T \in \mathcal{T}_h$ einen Kreis mit Radius $\rho_T \geq \frac{\tilde{h}}{K}$ enthält, wobei $\tilde{h} = \max_{T \in \mathcal{T}_h} \tilde{h}_T$ gilt.*

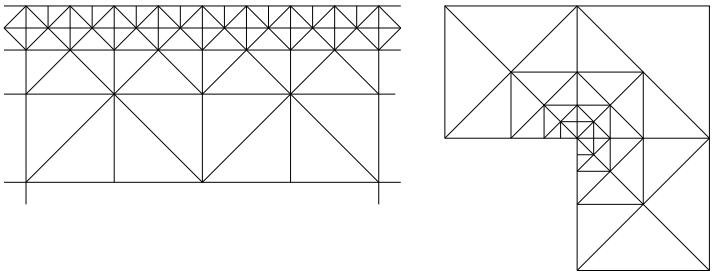

Bild 3.4 Quasi–gleichförmige, aber nicht gleichförmige Triangulierungen.

Im Modellproblem aus dem letzten Abschnitt hatten wir globale $\mathcal{C}^0$–Dreieckselemente verwendet, die auf linearen Ansatzfunktionen basierten. Andere globale $\mathcal{C}^0$–Elemente auf einer Dreicksstruktur erhält man, in dem man als Ansatzfunktionen stückweise–definierte Polynome höheren Grades verwendet: ein Polynom vom Grad t in $\mathbb{R}^2$ besitzt genau $s = 1+2+\cdots+(t+1) = (t+1)(t+2)/2$ Freiheitsgrade (Koeffizienten). Möchte man also auf einem vorgegebenen Dreieck ein Polynom vom Grad t festlegen, muss man genau s Funktionswerte des Polynoms auf dem Dreieck vorgeben. Das zugehörige Interpolationsproblem: finde ein Polynom, das auf einem gegebenen Dreiecke mit den ausgezeichneten Punkten $z_1, \ldots, z_s$, wobei jeweils $t+1$ auf einer Kante des Dreiecks liegen, die Interpolationsaufgabe

$$p(z_i) = f(z_i), \quad i = 1, 2, ..., s$$

mit $f \in \mathcal{C}(T)$ vorgegeben, löst, besitzt eine eindeutige Lösung.
In Bild 3.5 sind die zugehörigen Knotenpunkte, auf denen man Funktionswerte vorschreibt, für den Fall linearer, quadratischer und kubischer Elemente dargestellt.

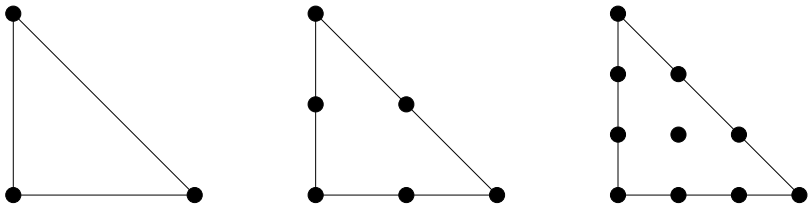

Bild 3.5 Knoten der nodalen Basis für lineare, quadratische und kubische Elemente.

Allgemein kann man zur Konstruktion globaler $\mathcal{C}^0$–Elemente auf einer gegebenen Triangulierung folgendes Schema angeben:

1) Man definiert eine zulässige Triangulierung des Rechengebiets Ω.

2) Auf jedem Dreieck definiert man $s = (t+1)(t+2)/2$ Knotenpunkte, die zur Interpolationsaufgabe verwendet werden.
3) Generiere auf jedem Dreieck ein Polynom vom Grad kleiner gleich t, das die zugehörige Interpolationsaufgabe löst.
4) Globale Stetigkeit entlang der Kanten der Dreiecke ist dann automatisch erfüllt: ein Polynom vom Grad kleiner gleich t ist entlang einer Kante vollständig durch die Vorgabe von $t+1$ Punkten auf dieser Kante bestimmt.

Der Finite–Element Raum S_h ist dann gegeben durch

$$M_{0,0}^t := S_h = \{v \in \mathcal{C}^0(\Omega),\ v|_T \in \mathcal{P}_t \text{ for } T \in \mathcal{T}_h,\ v|_{\partial\Omega} = 0\}$$

Für diesen Finite–Element Raum S_h definiert man die sogenannte Nodale Basis. Diese besteht gerade aus den Basisfunktionen, die an genau einem Punkt der Knoten, die zur Interpolationsaufgabe auf einem Dreieck der Triangulierung verwendet werden, ungleich sind, d.h. $\psi_i(x_j, y_j) = \delta_{ij}$.
Die Kosntruktion von globalen $\mathcal{C}^1$–Elementen ist deutlich komplizierter. Als Beispiel konstruieren wir $\mathcal{C}^1$–Elemente auf einer Dreieckszerlegung unter Verwendung von Polynomen vom Grad 5. In diesem Fall besitzt ein Polynom pro Dreieck genau 21 Freiheitsgrade (Koeffizienten). Die ersten 18 Freiheitsgrade werden dadurch festgelegt, dass man an jedem Eckpunkt des Dreiecks die Werte bis zu den zweiten Ableitungen vorschreibt. Beim sogenannten Argyris–Element bestimmt man die drei restlichen Freiheitsgrade durch Vorgabe der Normalenableitung in den drei Kantenmittelpunkten, siehe auch Bild 3.6.

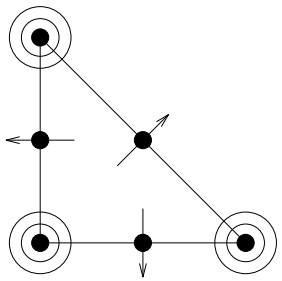

Bild 3.6 Das globale $\mathcal{C}^1$–Element nach Argyris.

Man überlegt sich leicht, dass das Argyris–Element ein globales $\mathcal{C}^1$–Element ist: die Einschränkung des Polynoms auf eine Kante des Dreiecks ist ein Polynom in einer Variablen mit Grad 5 und besitzt daher 6 Freiheitsgrade. Durch die Vorgabe der Ableitungen bis zur zweiten Ordnung an den beiden Eckpunkten, sind diese Freiheitsgrade eindeutig bestimmt. Daher ist die Polynomfunktion entlang Kanten stetig. Die Normalenableitung entlang einer Kante ist ein Polynom in einer Variablen mit Grad 4 und besitzt daher 5 Freiheitsgrade. Diese sind durch die Werte der Normalenableitung und die erste Ableitung an den Eckpunkten einer Kante sowie den Wert der Normalenableitung im Mittelpunkt einer Kante eindeutig festgelegt. Daher ist die Normalenableitung entlang einer Kante eine stetige Funktion.

Bei rechteckigen $\mathcal{C}^0$–Elementen verwendet man Ansatzfunktionen aus der Menge

$$Q_t = \left\{ u(x,y) = \sum_{0 \le i,j \le t} c_{ij} x^i y^j \right\}$$

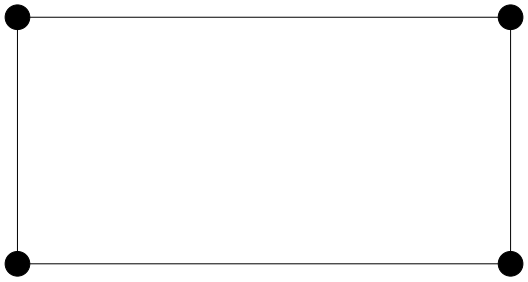

Bild 3.7 Knotenpunkte eines rechteckigen und linearen $\mathcal{C}^0$–Elements.

Im Fall $t = 1$ ist dies eine Ansatzfunktion $u(x, y)$ der Form

$$u(x,y) = a + bx + cy + dxy \tag{2.3.6}$$

d.h. u ist im Inneren des Rechtecks ein Polynom zweiten Grades, die Einschränkung auf eine Kante ist eine lineare Funktion. Globale Stetigkeit ist gewährleistet duch Vorgabe der Funktionswerte an den Eckpunkten des Rechtecks.
Werden die Rechteckelemente gedreht, sodass die Kanten nicht entlang der Koordinatenachsen liegen, muss man den Ansatz (2.3.6) entsprechend ändern (siehe auch Bild 3.8).

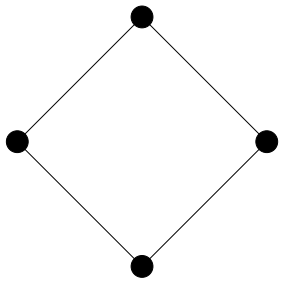

Fig. 3.8 Gedrehtes Rechteck zur Definition eines globalen $\mathcal{C}^0$–Elements.

In diesem Fall definiert man den FE–Raum S_h in der Form

$$\begin{aligned} S_h &= \{v \in \mathcal{C}^0(\Omega),\ v|_T \in Q_2,\ v = 0 \text{ on } \partial\Omega,\ T \in \mathcal{T}_h\} \\ &= \{v \in \mathcal{C}^0(\Omega),\ v|_T \in P_2,\ \text{ restiction of } v \text{ to the edges is in } P_1 \\ &\quad v = 0 \text{ on } \partial\Omega\ ,\ T \in \mathcal{T}_h\} \end{aligned}$$

Bei einer Gebietszerlegung mit Rechtecken verwendet man häufig Polynome vom Grad 3, deren Restriktionen auf Kanten vom Grad 2 sind. Die allgemeine Form mit 8 Freiheitsgraden schreibt man als

$$\begin{aligned} u(x,y) &= a + bx + cy + dxy \\ &\quad + e(x^2-1)(y-1) + f^2(x^2-1)(y+1) \\ &\quad + g(x-1)(y^2-1) + h(x+1)(y^2-1). \end{aligned}$$

und bezeichnet die resultierenden Elemente als Elemente der Serendipity–Klasse (siehe auch Bild 3.9). Die ersten Koeffizienten werden durch die Funktionswerte an den Eckpunkten festgelegt, die restlichen 4 durch die Werte an den Kantenmittelpunkten.

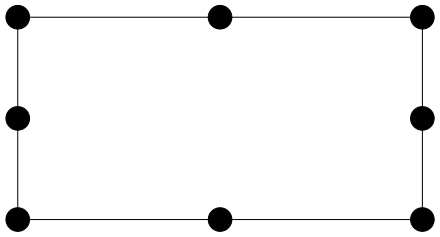

Bild 3.9 Rechteck–Elemente der Serendipity–Klasse.

4. Approximationstheorie der Finite–Element Methode

In diesem Abschnitt untersuchen die Genauigkeit von numerischen Approximationen auf der Basis von Finite–Element Methoden. Nach dem Lemma von Céa hängt die Genauigkeit allein davon ab, wie gut eine exakte Lösung durch ein Element des zugrundliegenden FE–Raums S_h approximiert werden kann.
Im folgenden beschränken wir uns auf globale $\mathcal{C}^0$–Elemente, die gleichzeitig Elemente des $H^1(\Omega)$ sind. Die Genauigkeit der numerischen Approximation werden wir im Folgenden bezüglich der Gitter–abhängigen Normen $\|\cdot\|_h$ und $\|\cdot\|_{m,h}$ charakterisieren.

Definition 2.13.
Consider a partition $\mathcal{T}_h = \{T_1, T_2, ..., T_m\}$ of Ω into triangles, rectangles, etc. and let $m \geq 1$. Then

$$\|v\|_{m,h} = \sqrt{\sum_{T_j \in \mathcal{T}_h} \|v\|^2_{m,T_j}}$$

and obviously we have

$$\|v\|_{m.h} = \|v\|_{m,\Omega} \quad \textit{for } v \in H^m(\Omega)$$

Let $m \geq 2$. Then, using Sobolev imbedding theorems, $H^m(\Omega) \subset \mathcal{C}^0(\Omega)$. Hence, from the lemma of the previous section, there are $t(t+1)/2$ points in each element T_i, such that there is an unique piecewise polynomial of degree $\leq t-1$, which we will denote by $I_h v$, interpolating v in Ω, where $v \in H^m(\Omega)$. Hence, we like to find an estimate for the term $\|v - I_h v\|_{m,h}$ by $\|v\|_{t,\Omega}$, where $m \leq t$.
We start by estimating the interpolation with polynomials on an arbitrary domain Ω.

LEMMA 2.14. *(Bramble–Hilbert)*
Let $\Omega \in \mathbb{R}^2$ with Lipschitz–continuous boundary. Let $t \geq 2$ and $t(t+1)/2$ points $z_1, ..., z_s$ be given, defining the interpolation $I : H^t(\Omega) \to P_{t-1}$ by polynomials of degree $\leq t-1$. Then, with $c = c(\Omega, z_1, ..., z_s)$ we have the estimate

$$\|u - Iu\|_t \leq c|u|_t \tag{2.4.7}$$

for $u \in H^t(\Omega)$.

BEWEIS. We use as norm in the Sobolev space $H^t(\Omega)$

$$|||v||| = |v|_t + \sum_{i=1}^{s} |v(z_i)|$$

If we can prove, that $|||\cdot|||$ and $\|\cdot\|_t$ are equivalent in $H^t(\Omega)$, we would have

$$\begin{aligned} \|u - Iu\|_t &\leq c|||u - Iu||| \\ &= c\left\{ |u - Iu|_t + \sum_{i=1}^{s} |(u - Iu)(z_i)| \right\} \\ &= c|u - Iu|_t = c|u|_t \end{aligned}$$

where we used, that $Iu = u$ at the interpolating points and $D^t Iu = 0$, because Iu is a polynomial of degree $t-1$.
Hence, we prove, that the two norms are equivalent: one direction is based on the fact that the embeddings $H^t \hookrightarrow H^2 \hookrightarrow \mathcal{C}^0$ are continuous and therefore

$$|v(z_i)| \leq c\|v\|_t, \quad i = 1, ..., s$$

and $|||v||| \leq c\|v\|_t$.
For the other direction, assume that the reversal direction does not hold, i.e. the estimate

$$\|v\|_t \leq c|||v|||$$

is wrong for any positive number c. Then there exists a sequence (v_k) in $H^t(\Omega)$ with

$$\|v_k\|_t = 1, \quad |||v||| \leq \frac{1}{k}, \quad k = 1, 2, \ldots$$

because for $c = k$ we have $\|v_k\|_t > k|||v|||$.
Since the embedding $H^t \hookrightarrow H^{t-1}$ is compact (Rellich), there exists a subsequence, again denote by (v_k), which converges in $H^{t-1}(\Omega)$. Moreover, (v_k) is a Cauchy sequence in $H^{t-1}(\Omega)$. With $|v_k|_t \to 0$ and

$$\|v_k - v_l\|_t^2 \leq \|v_k - v_l\|_{t-1}^2 + (|v_k|_t + |v_l|_t)^2$$

we conclude, that the sequence (v_k) is even a Cauchy sequence in $H^t(\Omega)$. Hence, we have a convergence to some $v^* \in H^t(\Omega)$ and due to continuity

$$\|v^*\|_t = 1, \quad \text{and } |||v^*||| = 0$$

But this is a contradiction, since, because $|v^*|_t = 0$ means, that v^* is a polynomial in $\mathcal{P}_{t-1}$ and because $v^*(z_i) = 0$ for $i = 1, 2, \ldots, s$, v^* must be the zero polynomial. □

An abstract formulation of the Bramble–Hilbert Lemma is following:

LEMMA 2.15.
Let $\Omega \subset \mathbb{R}^2$ with Lipschitz–continuous boundary, $t \geq 2$ and $L : H^t(\Omega) \to Y$ a bounded linear operator, where Y is a normed space. If $P_{t-1} \subset \ker L$, then we have with $c = c(\Omega, L) \geq 0$

$$\|Lv\| \leq c|v|_t$$

for all $v \in H^t(\Omega)$.

BEWEIS. Since $\|Lv\| \leq c_0\|v\|_t$ we get for the interpolating operator $I : H^t(\Omega) \to P_{t-1}$, because $Iv \in \ker L$

$$\begin{aligned} \|Lv\| &= \|L(v - Iv)\| \leq c_0\|v - Iv\|_t \\ &\leq c \cdot c_0|v|_t \end{aligned}$$

where c is the constant given in (2.4.7). □

Let us consider in more detail a partition of the domain Ω using triangular $\mathcal{C}^0$–elements, i.e. we take again piecewise polynomials of degree $t-1$, $t \geq 2$ on triangles. Then, $\mathcal{T}_h$ and $\mathcal{S}_h$ define the interpolation operator $I_h : H^t(\Omega) \to S_h$. Moreover, let the triangulation $\mathcal{T}_h$ be quasi–uniform and K the parameter of $\mathcal{T}_h$ according to the definition of a quasi–uniform triangulation. Then we have the following theorem.

SATZ 2.16.
Let $t \geq 2$ and $\mathcal{T}_h$ a quasiuniform triangulation of Ω into triangles. Then we have for the interpolation using piecewise polynomials of degree $t-1$ with $c = c(\Omega, K, t)$ the estimate

$$\|u - I_h u\|_{m,h} \leq c \cdot h^{t-m}|u|_{t,\Omega}, \quad u \in H^t(\Omega), \quad 0 \leq m \leq t$$

where c is the constant given in the (first) Brambert–Hilbert lemma.

BEWEIS. We restrict the proof to the case of a grid formed by congruent triangles: consider the reference triangle T_1 with

$$\begin{aligned} T_h &= h \cdot T_1 \\ &= \{\mathbf{x} = h\mathbf{y}, \quad \mathbf{y} \in T_1\} \end{aligned}$$

such that the triangles of $\mathcal{T}_h$ are congruent to T_h (in particular, $T_1 \neq T_{\text{ref}}$).
If we prove the estimate

$$\|u - Iu\|_{m,T_h} \leq c \cdot h^{t-m}|u|_{t,T_h} \tag{2.4.8}$$

then the theorem follows by taking squares and summation over all triangles.
The proof of (2.4.8) is done using a transformation theorem: define for $u \in H^t(T_h)$ the function $v \in H^t(T_1)$ by $v(y) = u(hy)$. Then $\partial^\alpha v = h^{|\alpha|}\partial^\alpha u$ with $\alpha \leq t$A and the transformation of the domain in $\mathbb{R}$ gives another factor h^{-2}, i.e. we have

$$\begin{aligned} |v|^2_{l,T_1} &= \sum_{|\alpha|=l} \int_{T_1} (\partial^\alpha v)^2 dy \\ &= \sum_{|\alpha|=l} \int_{T_h} h^{2l}(\partial^\alpha u)^2 h^{-2} dx \\ &= h^{2l-2}|u|^2_{l,T_h} \end{aligned} \tag{2.4.9}$$

Since $h \leq 1$, the smallest power in h is important when summing up over all triangles

$$\begin{aligned} \|u\|^2_{m,T_h} &= \sum_{l\leq m} |u|^2_{l,T_h} \\ &= \sum_{l\leq m} h^{-2l+2}|v|^2_{l,T_1} \\ &\leq h^{-2m+2}\|v\|^2_{m,T_1} \end{aligned}$$

Taking $u - Iu$ instead of u we get with the same arguments

$$\begin{aligned} \|u - Iu\|_{m,T_h} &\leq h^{-m+1}\|v - Iv\|_{m,T_1} \\ &\leq h^{-m+1}\|v - Iv\|_{t,T_1} \quad (m \leq t) \\ &\leq h^{-m+1}c|v|_{t,T_1} \quad \text{(Bramble-Hilbert)} \\ &= h^{t-m}c|u|_{t,T_h} \quad (2.4.9) \end{aligned}$$

□

Für lineare Elemene, d.h. $t = 2$, erhält man eine Approximationsordnung von $O(h^2)$ in der Norm $\|\cdot\|_{0,h}$ beziehungsweise $0(h)$ für $\|\cdot\|_{1,h}$. Bei rechteckigen, globalen $\mathcal{C}^0$–Elementen erhält man die folgende äquivalente Aussage.

Satz 2.17. *Sei $\mathcal{T}_h$ eine quasi–gleichförmige Zerlegung des Rechengebietes Ω in Parallelogramme. Dann gilt für den Interpolationsfehler mit bilinearen Elementen die Abschätzung*

$$\|u - Iu\|_{m,h} \leq c \cdot h^{2-m}|u|_{2,\Omega}, \quad u \in H^2(\Omega)$$

Für Elemente der Serendipity–Klasse zeigt man folgende Abschätzung

$$\|u - I_h u\|_{m,h} \leq ch^{t-m}|u|_{t,\Omega}, \quad u \in H^t(\Omega), \quad m = 0, 1,\ t = 2, 3$$