

# Zusammenfassung Numerik von PDEs

© Tim Baumann, <http://timbaumann.info/uni-spicker>

Dies ist ein verkürztes Skript zur gleichnamigen Vorlesung von Frau Prof. Dr. Tatjana Stykel an der Universität Augsburg im WS 15/16.

**Def.** Sei  $\Omega \subseteq \mathbb{R}^n$  offen. Eine DGL der Form

$$F(x, u, Du, \dots, D^k u) = 0$$

heißt **partielle DGL/PDE** der Ordnung  $k \geq 1$ , wobei

$$F : \Omega \times \mathbb{R} \times \mathbb{R}^n \times \dots \times \mathbb{R}^{n^k} \rightarrow \mathbb{R}$$

eine gegebene Funktion und  $u : \Omega \rightarrow \mathbb{R}$  gesucht ist.

**Def (Klassifikation von PDEs).**

- Die PDE heißt **linear**, wenn sie die Form

$$\sum_{|\alpha| \leq k} a_\alpha(x) D^\alpha u = f(x)$$

mit Funktionen  $a_\alpha, f : \Omega \rightarrow \mathbb{R}$  besitzt.

- Die PDE heißt **semilinear**, wenn sie die Form

$$\sum_{|\alpha| = k} a_\alpha(x) D^\alpha u + a_0(x, u, Du, \dots, D^{k-1}u) = 0$$

besitzt, wobei  $a_\alpha : \Omega \rightarrow \mathbb{R}$  und  $a_0 : \Omega \times \mathbb{R} \times \mathbb{R}^n \times \dots \times \mathbb{R}^{n^k} \rightarrow \mathbb{R}$  gegeben sind.

- Die PDE heißt **quasilinear**, wenn sie die Form

$$\sum_{|\alpha| = k} a_\alpha(x, u, Du, \dots, D^{k-1}u) D^\alpha u + a_0(x, u, Du, \dots, D^{k-1}u) = 0$$

hat, wobei  $a_\alpha, a_0 : \Omega \times \mathbb{R} \times \mathbb{R}^n \times \dots \times \mathbb{R}^{n^k}$  gegeben sind.

- Die PDE heißt **nichtlinear**, falls die Ableitungen der höchsten Ordnung nicht linear vorkommen.

**Def.** Sei  $\Omega \subseteq \mathbb{R}^n$  offen und  $F : \Omega \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$  gegeben. Eine **PDE zweiter Ordnung** ist eine PDE der Form

$$F(x, u, \partial_{x_1} u, \dots, \partial_{x_n} u, \partial_{x_1} \partial_{x_1} u, \dots, \partial_{x_1} \partial_{x_n} u, \dots, \partial_{x_n} \partial_{x_n} u) = 0.$$

**Notation.**  $p_i := \partial_{x_i} u$ ,  $p_{ij} := \partial_{x_i x_j}^2 u$

$$M(x) := \begin{pmatrix} \frac{\partial F}{\partial p_{11}} & \dots & \frac{\partial F}{\partial p_{1n}} \\ \vdots & & \vdots \\ \frac{\partial F}{\partial p_{n1}} & \dots & \frac{\partial F}{\partial p_{nn}} \end{pmatrix} = M(x)^T.$$

**Def (Typeneinteilung für PDEs der 2. Ordnung).**

Obige PDE zweiter Ordnung heißt

- elliptisch** in  $x$ , falls die Matrix  $M(x)$  positiv o. negativ definit ist.
- parabolisch** in  $x$ , falls genau ein EW von  $M(x)$  gleich null ist und alle anderen dasselbe Vorzeichen haben.
- hyperbolisch** in  $x$ , falls genau ein EW von  $M(x)$  ein anderes Vorzeichen als die anderen EWe hat.

## Lösungstheorie elliptischer PDEs

**Def.** Sei  $\Omega \subset \mathbb{R}^n$  offen, zusammenhängend und beschränkt.

- $\mathcal{C}(\overline{\Omega}, \mathbb{R}^m) := \{u : \overline{\Omega} \rightarrow \mathbb{R}^m \mid u \text{ stetig}\}$ ,  $\mathcal{C}(\overline{\Omega}) := \mathcal{C}(\overline{\Omega}, \mathbb{R})$ , mit Norm

$$\|u\|_{\mathcal{C}(\overline{\Omega}, \mathbb{R}^m)} = \sup_{x \in \overline{\Omega}} \|u(x)\|. \quad (\text{Supremumsnorm})$$

- $\mathcal{C}^k(\overline{\Omega}, \mathbb{R}^m)$ ,  $k \in \mathbb{N}$  ist der Raum aller auf  $\Omega$   $k$ -mal stetig diff'baren Funktionen  $u : \Omega \rightarrow \mathbb{R}^m$ , die zusammen mit ihren Ableitungen bis zur Ordnung  $k$  stetig auf  $\overline{\Omega}$  fortgesetzt werden können.

$$\|u\|_{\mathcal{C}^k(\overline{\Omega}, \mathbb{R}^m)} = \sum_{|\alpha| \leq k} \|D^\alpha u\|_{\mathcal{C}(\overline{\Omega}, \mathbb{R}^m)}$$

- Sei  $\alpha \in [0, 1)$ .  $\mathcal{C}^{0,\alpha}(\overline{\Omega}, \mathbb{R}^m) = \{u \in \mathcal{C}(\overline{\Omega}, \mathbb{R}^m) \mid H_\alpha(u, \overline{\Omega}) < \infty\}$  mit

$$H_\alpha(u, \overline{\Omega}) := \sup_{x, y \in \overline{\Omega}, x \neq y} \frac{\|u(x) - u(y)\|}{\|x - y\|^\alpha} \quad (\text{Hölder-Koeffizient})$$

heißt **Raum der glm. Hölder-stetigen Fktn** zum Exponent  $\alpha$ . Der Hölder-Koeffizient ist dabei eine Seminorm auf  $\mathcal{C}^{0,\alpha}(\overline{\Omega}, \mathbb{R}^m)$ .

- $\mathcal{C}^{k,\alpha}(\overline{\Omega}, \mathbb{R}^m) := \{u \in \mathcal{C}^k(\overline{\Omega}, \mathbb{R}^m) \mid \forall |\gamma| = k : D^\gamma u \in \mathcal{C}^{0,\alpha}(\overline{\Omega}, \mathbb{R}^m)\}$  heißt **Hölder-Raum**. Eine Norm ist gegeben durch

$$\|u\|_{\mathcal{C}^{k,\alpha}(\overline{\Omega}, \mathbb{R}^m)} := \|u\|_{\mathcal{C}^k(\overline{\Omega}, \mathbb{R}^m)} + \sum_{|\gamma|=k} H_\alpha(D^\gamma u, \overline{\Omega}).$$

**Bem.** • Jede Hölder-stetige Funktion ist gleichmäßig stetig.

- $\mathcal{C}^{0,1}(\overline{\Omega}, \mathbb{R}^m)$  heißt **Raum der Lipschitz-stetigen Funktionen**.
- $\mathcal{C}$ ,  $\mathcal{C}^k$  und  $\mathcal{C}^{k,\alpha}$  sind Banach-Räume mit den jeweiligen Normen.

**Def.** Sei  $\Omega \subset \mathbb{R}^n$  offen, zusammenhängend und beschränkt. Das Gebiet  $\Omega$  gehört zur **Klasse  $\mathcal{C}^{k,\alpha}$** , wenn in jedem Punkt  $x \in \partial\Omega$  eine Umgebung in  $\partial\Omega$  existiert, die sich in einem geeigneten Koordinatensystem als ein Graph einer Funktion aus  $\mathcal{C}^{k,\alpha}$  darstellen lässt und  $\Omega$  lokal immer auf einer Seite von  $\partial\Omega$  liegt.

**Satz (Gauß'scher Integralsatz).** Sei  $\Omega \subset \mathbb{R}^n$  ein Lipschitz-Gebiet und  $u \in \mathcal{C}(\overline{\Omega}, \mathbb{R}^n) \cap \mathcal{C}^1(\Omega, \mathbb{R}^n)$ . Dann gilt

$$\int_\Omega \operatorname{div} u \, dx = \int_\Omega \sum_{i=1}^n \frac{\partial u_i}{\partial x_i} \, dx = \int_{\partial\Omega} \sum_{i=1}^n u_i \nu_i \, d\rho(x) = \int_{\partial\Omega} u \cdot \nu \, d\rho(x),$$

wobei  $\nu$  der äußere Normalenvektor an den Rand von  $\Omega$  ist.

**Problem.** Wir betrachten das Randwertproblem

$$(\text{RWP}) \quad \begin{cases} \mathcal{L}u = f & \text{in } \Omega \\ \mathcal{R}u = g & \text{auf } \partial\Omega \end{cases} \quad \begin{matrix} (\text{PDE}) \\ (\text{Randbedingung}) \end{matrix}$$

wobei  $\mathcal{L}$  der lineare Differentialoperator

$$\mathcal{L}u = - \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x) \frac{\partial u}{\partial x_i} + c(x)u$$

mit Fktn  $a_{ij}, b_i, c, f : \Omega \rightarrow \mathbb{R}$ ,  $g : \partial\Omega \rightarrow \mathbb{R}$  ist, sodass  $A(x) := (a_{ij}(x))$  symmetrisch ist. Als Randbedingung (RB) verlangen wir:

$$\begin{array}{lll} \text{Dirichlet-RB:} & u & = g \quad \text{auf } \partial\Omega, \\ \text{Neumann-RB:} & (A(x)\nabla u) \cdot \nu & = g \quad \text{auf } \partial\Omega \text{ oder} \\ \text{Robin-RB:} & (A(x)\nabla u) \cdot \nu + \delta u & = g \quad \text{auf } \partial\Omega. \end{array}$$

**Bem.** • Man kann auch auf verschiedenen Teilstücken des Randes verschiedene Bedingungen stellen.

- Falls die Funktionen  $a_{ij}$  differenzierbar sind, so kann  $\mathcal{L}$  in **Divergenzform** geschrieben werden:

$$\begin{aligned} \mathcal{L}u &= - \sum_{i,j=1}^n \frac{\partial}{\partial x_j} \left( a_{ij}(x) \frac{\partial u}{\partial x_i} \right) + \sum_{i=1}^n \underbrace{\left( \left( \sum_{j=1}^n \frac{\partial}{\partial x_j} a_{ij}(x) \right) + b_i(x) \right)}_{\tilde{b}(x) :=} \frac{\partial u}{\partial x_i} + c(x)u \\ &= - \operatorname{div}(A(x)\nabla u) + \tilde{b}(x) \cdot \nabla u + c(x)u \end{aligned}$$

**Voraussetzung.** Wir nehmen im Folgenden an:

- $\mathcal{L}$  ist **gleichmäßig elliptisch**, d. h.

$$\exists \lambda_0 > 0 : \forall \xi \in \mathbb{R}^n : \forall x \in \Omega : \xi^T A(x) \xi \geq \lambda_0 \|\xi\|^2.$$

Dabei heißt  $\lambda_0$  **Elliptizitätskonstante**.

- $a_{ij}, b_i, c, f \in \mathcal{C}(\overline{\Omega})$ ,  $g \in \mathcal{C}(\partial\Omega)$

**Bem.**  $\mathcal{L}$  ist elliptisch auf  $\Omega \iff A(x) > 0$  (spd) für alle  $x \in \Omega$

**Def.** Eine Fkt  $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$  heißt **klassische Lsg** vom (RWP) mit  $\mathcal{R}u := u$ , wenn die beiden Gleichungen in (RWP) in jedem Punkt von  $\Omega$  bzw. des Randes  $\partial\Omega$  erfüllt sind.

**Satz (Maximumprinzip).** Sei  $\Omega \subset \mathbb{R}^n$  offen, zshgd u. beschränkt. Sei  $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$  eine Lösung vom (RWP),  $f \leq 0$  in  $\Omega$  und  $c \equiv 0$ . Dann nimmt  $u$  sein Maximum auf dem Rand  $\partial\Omega$  an, d. h.

$$\sup_{x \in \overline{\Omega}} u(x) = \sup_{x \in \partial\Omega} u(x) = \sup_{x \in \partial\Omega} g(x)$$

**Kor.** Sei  $c \geq 0$  und  $f \leq 0$ . Dann gilt  $\sup_{x \in \overline{\Omega}} u(x) \leq \max\{\sup_{x \in \partial\Omega} u(x), 0\}$ .

**Kor (Vergleichsprinzip).** Für  $u_1, u_2 \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$  und  $c \geq 0$  gelte  $\mathcal{L}u_1 \leq \mathcal{L}u_2$  in  $\Omega$  und  $u_1 \leq u_2$  auf  $\partial\Omega$ . Dann gilt  $u_1 \leq u_2$  auf  $\overline{\Omega}$ .

**Kor (Eindeutigkeit).** Sei  $c \geq 0$ . Dann hat (RWP) höchstens eine Lösung  $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$ .

**Satz.** Sei  $\Omega$  ein beschr. Lipschitz-Gebiet,  $a_{ij}, b_i, c \in \mathcal{C}(\overline{\Omega})$ ,  $c \geq 0$ ,  $\mathcal{L}$  glm. elliptisch,  $f \in \mathcal{C}^{0,\alpha}(\overline{\Omega})$  für ein  $\alpha \in (0, 1)$  und  $g \in \mathcal{C}(\partial\Omega)$ . Dann besitzt (RWP) genau eine Lsg  $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$ .

**Achtung.** Es muss aber nicht  $u \in \mathcal{C}^2(\overline{\Omega})$  gelten!

## Differenzenverfahren

**Verfahren (DV).** Am Beispiel des Poisson-Problems

$$(RWP_1) \quad \begin{cases} -\Delta u = f & \text{in } \Omega = (0, 1) \\ u(0) = g_0, u(1) = g_1 & \text{auf } \partial\Omega \end{cases}$$

Wir führen folgende Schritte durch:

1. Diskretisierung: Wähle  $n \in \mathbb{N}$ , setze  $h := \frac{1}{n}$  und

$$\begin{aligned} \Omega_h &:= \{x_i := ih \mid i = 1, \dots, n-1\} & (\text{innere Gitterpunkte}) \\ \partial\Omega_h &:= \{x_0 = 0, x_n = 1\} & (\text{Randpunkte}) \end{aligned}$$

2. Approx. der Ableitungen durch Differenzenquotienten (DQ)

$$u'(x_i) \approx \frac{1}{h} (u(x_i + h) - u(x_i)) \quad (\text{Vorwärts-DQ})$$

$$u'(x_i) \approx \frac{1}{h} (u(x_i) - u(x_i - h)) \quad (\text{Rückwärts-DQ})$$

$$u'(x_i) \approx \frac{1}{2h} (u(x_i + h) - u(x_i - h)) \quad (\text{zentraler DQ})$$

Für die zweite Ableitung ergibt sich

$$\begin{aligned} u''(x_i) &= (u'(x_i))' \approx \frac{1}{h} (u'(x_i + h) - u'(x_i)) \approx \\ &\approx \frac{1}{h} \cdot \left( \frac{1}{h} (u(x_i + h) - u(x_i)) - \frac{1}{h} (u(x_i) - u(x_i - h)) \right) \\ &= \frac{1}{h^2} (u(x_i + h) - 2 \cdot u(x_i) + u(x_i - h)) =: \Delta_h u \end{aligned}$$

Dabei heißt  $\Delta_h$  der diskrete eindim. Laplace-Operator.  
Das diskretisierte Randwertproblem ist nun

$$(RWP_1)_h \quad \begin{cases} -\Delta_h u_h = f & \text{in } \Omega_h, \\ u_h(0) = g_0, u_h(1) = g_1 & \text{auf } \partial\Omega_h. \end{cases}$$

3. Aufstellen des linearen Gleichungssystems

$$\begin{aligned} \frac{1}{h^2} (2u_h(x_1) - u_h(x_2)) &= f(x_1) + \frac{g_0}{h^2} & (i=1) \\ \frac{1}{h^2} (-u_h(x_{i-1}) + 2u_h(x_i) - u_h(x_{i+1})) &= f(x_i) & (i=2, \dots, n-2) \\ \frac{1}{h^2} (-u_h(x_{n-2}) + 2u_h(x_{n-1})) &= f(x_{n-1}) + \frac{g_1}{h^2} & (i=n-1) \end{aligned}$$

Als lineares Gleichungssystem:  $-\tilde{\Delta}_h \tilde{u}_h = \tilde{f}_h$  mit

$$-\tilde{\Delta}_h = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \\ 0 & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{pmatrix} \in \mathbb{R}^{(n-1) \times (n-1)},$$

$$\tilde{u}_h = \begin{pmatrix} u_h(x_1) \\ \vdots \\ u_h(x_{n-1}) \end{pmatrix}, \quad \tilde{f}_h = \begin{pmatrix} f(x_1) + \frac{g_0}{h^2} \\ f(x_2) \\ \vdots \\ f(x_{n-2}) \\ f(x_{n-1}) + \frac{g_1}{h^2} \end{pmatrix}$$

## Konvergenz, Konsistenz und Stabilität des DV

**Ziel.** Herausfinden, was die Lösung  $u_h$  von  $(RWP)_h$  (die man durch Lösen von (LGS) erhält) mit der Lösung  $u$  zum ursprünglichen Problem (RWP) zu tun hat. Ist etwa  $u_h$  die Einschränkung von  $u$ , oder zumindest annäherungsweise? Wenn ja, wie klein muss man  $h$  wählen, damit die Approximation gut wird?

$$(RWP) \quad \begin{cases} -\mathcal{L}u = f & \text{in } \Omega, \\ u = g & \text{auf } \partial\Omega \end{cases}$$

$$(RWP)_h \quad \begin{cases} -\mathcal{L}_h u = f_h & \text{in } \Omega_h, \\ u_h = g_h & \text{auf } \partial\Omega_h \end{cases}$$

$$(LGS) \quad \tilde{\mathcal{L}}_h \tilde{u}_h = \tilde{f}_h$$

**Notation.**  $U_h := \{\Omega_h \rightarrow \mathbb{R}\}$ ,  $R_h : \mathcal{C}(\bar{\Omega}) \rightarrow U_h$ ,  $u \mapsto u|_{\Omega_h}$

**Def.** Das Differenzenverfahren  $(RWP)_h$  heißt

- **konvergent** von der Ordnung  $p$ , falls  $C > 0$ ,  $h_0 > 0$  existieren, sodass für die Lsg  $u$  von (RWP) und die Lsg  $u_h$  von  $(RWP)_h$  gilt:

$$\|u_h - R_h u\|_h \leq Ch^p \quad \text{für alle } 0 < h \leq h_0,$$

wobei  $\|\cdot\|_h$  eine Norm zu  $U_h$  ist, wie z. B.  $\|u_h\|_h := \max_{x \in \Omega_h} |u_h(x)|$ .

- **konsistent** von der Ordnung  $p$ , falls

$$\|\mathcal{L}_h R_h u - R_h \mathcal{L}u\|_h \leq ch^p \|u\|_{\mathcal{C}^{p+2}(\bar{\Omega})} \quad \forall u \in \mathcal{C}^{p+2}(\bar{\Omega}).$$

- **stabil**, falls  $\tilde{\mathcal{L}}_h$  invertierbar ist und ein  $h_0 > 0$  existiert mit

$$\sup_{0 < h \leq h_0} \|\tilde{\mathcal{L}}_h^{-1}\|_h < \infty, \quad \text{wobei } \|\tilde{\mathcal{L}}_h^{-1}\|_h := \sup_{f \neq 0} \frac{\|\tilde{\mathcal{L}}_h^{-1} f\|_h}{\|f\|_h}.$$

*Bem.* Die ind. Matrixnorm ist  $\|\tilde{\mathcal{L}}_h^{-1}\|_h = \|\tilde{\mathcal{L}}_h^{-1}\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |l_{ij}|$ .

**Satz.** Ist das DV  $(RWP)_h$  konsistent und stabil, so auch konvergent. Genauer gilt: Ist  $(RWP)_h$  stabil und konsistent von der Ordnung  $p$  und  $u \in \mathcal{C}^{p+2}(\bar{\Omega})$ , dann ist  $(RWP)_h$  konvergent von der Ordnung  $p$ .

*Beweis.* Setze  $w_h := u_h - R_h u$ . Für  $x \in \partial\Omega_h$  gilt dann  $w_h(x) = 0$  und für  $x \in \Omega_h$  gilt

$$\begin{aligned} \tilde{\mathcal{L}}_h w_h(x) &= \mathcal{L}_h w_h(x) = \mathcal{L}_h u_h(x) - \mathcal{L}_h R_h u(x) \\ &= f_h(x) - \mathcal{L}_h R_h u(x) = R_h f(x) - \mathcal{L}_h R_h u(x) \\ &= R_h \mathcal{L}u(x) - \mathcal{L}_h R_h u(x) \end{aligned}$$

Somit gilt  $w_h = \tilde{\mathcal{L}}_h^{-1} (R_h \mathcal{L}u - \mathcal{L}_h R_h u)$  in  $\Omega_h$ , also

$$\begin{aligned} \|w_h\|_h &= \|\tilde{\mathcal{L}}_h^{-1} (R_h \mathcal{L}u - \mathcal{L}_h R_h u)\| \leq \|\tilde{\mathcal{L}}_h^{-1}\|_h \cdot \|R_h \mathcal{L}u - \mathcal{L}_h R_h u\|_h \\ &\leq c_1 \cdot c_2 \cdot h^p \cdot \|u\|_{\mathcal{C}^{p+2}(\bar{\Omega})} \leq Ch^p \quad \text{für } 0 < h \leq h_0. \end{aligned}$$

**Lem.** Das DV  $(RWP_1)_h$  ist konsistent von der Ordnung 2. Es gilt

$$\|\Delta_h R_h u - R_h \Delta u\|_h \leq \frac{1}{12} \|u\|_{\mathcal{C}^4(\bar{\Omega})} h^2 \quad \forall u \in \mathcal{C}^4(\bar{\Omega}).$$

*Bem.* Um zu zeigen, dass  $(RWP_1)_h$  konvergent ist, müssen wir noch zeigen, dass  $\tilde{\mathcal{L}}_h = -\tilde{\Delta}_h$  invertierbar ist und  $\sup_{0 < h \leq h_0} \|\tilde{\Delta}_h\|_h < \infty$ .

**Def.** Eine Matrix  $A = (a_{ij}) \in \mathbb{R}^{n \times n}$  heißt **M-Matrix**, falls

- a)  $a_{ii} > 0$  für  $i = 1, \dots, n$ , b)  $a_{ij} \leq 0$  für  $i \neq j$ ,  $i, j = 1, \dots, n$ ,
- c)  $A$  invertierbar ist und d) für  $A^{-1} =: B = (b_{ij})$  gilt  $b_{ij} \geq 0$ .

**Lem.** Erfülle  $A \in \mathbb{R}^{n \times n}$  die Bedingungen a) und b). Zerlege  $A = D + L + R$  in eine Diagonalmatrix und strikte untere/obere Dreiecksmatrizen. Dann ist  $A$  genau dann eine M-Matrix wenn

$$\rho(D^{-1}(L + R)) < 1.$$

*Bem.* Es gilt folgende Monotonie-Eigenschaft für M-Matrizen:

$$x \leq y \implies A^{-1}x \leq A^{-1}y.$$

**Def.** Eine Matrix  $A \in \mathbb{R}^{n \times n}$  heißt **reduzibel** (oder zerlegbar), wenn es eine Permutationsmatrix  $P \in \mathbb{R}^{n \times n}$  gibt, sodass

$$PAP^T = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \quad \text{mit } A_{11} \in \mathbb{R}^{k \times k}, 0 < k < n.$$

**Lem** (Gerschgorin). Alle EWe einer Matrix  $A = (a_{ij}) \in \mathbb{C}^{n \times n}$  liegen in der Menge

$$\bigcup_{i=1}^n \overline{B_{r_i}(a_{ii})} \quad \text{mit } r_i := \sum_{j=1}^n |a_{ij}|.$$

Falls  $A$  irreduzibel ist, so liegen sie sogar in

$$\left( \bigcup_{i=1}^n B_{r_i}(a_{ii}) \right) \cup \left( \bigcap_{i=1}^n \partial B_{r_i}(a_{ii}) \right)$$

**Def.** Sei  $A = (a_{ij}) \in \mathbb{R}^{n \times n}$  eine Matrix.

- $A$  heißt **schwach diagonaldominant**, falls

$$\sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \leq |a_{ii}| \quad \text{für } i = 1, \dots, n$$

und ein  $i_0$  existiert, sodass die Ungleichung strikt ist.

- $A$  heißt **diagonaldominant**, falls

$$\sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| < |a_{ii}| \quad \text{für } i = 1, \dots, n$$

- $A$  heißt **irreduzibel diagonaldominant**, falls  $A$  irreduzibel und schwach diagonaldominant ist.

**Lem.** Sei  $A = (a_{ij}) \in \mathbb{R}^{n \times n}$  eine Matrix mit  $a_{ii} > 0$ ,  $i = 1, \dots, n$  und  $a_{ij} \leq 0$ ,  $i, j = 1, \dots, n$ ,  $i \neq j$ , die diagonaldominant oder irreduzibel diagonaldominant ist. Dann ist  $A$  eine M-Matrix.

*Bem.*  $-\tilde{\Delta}_h$  ist irreduzibel diagonaldominant, also eine M-Matrix.

**Lem.** Sei  $A$  eine irreduzible M-Matrix. Dann gilt  $A^{-1} > 0$ .

- **Lem.** Sei  $A \in \mathbb{R}^{n \times n}$  eine M-Matrix und es existiere ein Vektor  $v$ , sodass  $(Av)_j \geq 1$ ,  $j = 1, \dots, n$ . Dann gilt  $\|A^{-1}\|_\infty \leq \|v\|_\infty$ .

**Lem.**  $\|\tilde{\Delta}_h^{-1}\|_\infty \leq \frac{1}{8}$

**Satz.** Das DV  $(RWP_1)_h$  ist konvergent von der Ordnung 2, falls die Lösung von  $(RWP_1)$  zu  $\mathcal{C}^4([0, 1])$  gehört. Es gilt die Abschätzung

$$\|u_h - R_h u\|_\infty \leq \frac{h^2}{96} \|u\|_{\mathcal{C}^4([0, 1])}.$$

## Differenzenverfahren in $(0, 1) \times (0, 1) \subset \mathbb{R}^2$

**Problem.** Wir betrachten nun

$$(\text{RWP}_2) \quad \begin{cases} -\Delta u = f & \text{in } \Omega = (0, 1) \times (0, 1) \\ u = g & \text{auf } \partial\Omega \end{cases}$$

1. Diskretisierung: Setze  $h := \frac{1}{n}$ ,  $n \in \mathbb{N}$  und

$$\Omega_h := \{(x, y) \in \Omega \mid x = ih, y = jh, i, j = 1, \dots, n-1\}$$

$$\partial\Omega_h := \{(x, y) \in \partial\Omega \mid x = ih, y = jh, i, j = 1, \dots, n-1\}$$

2. Approximation der Ableitungen

$$\begin{aligned} -\Delta u(x, y) &= -\frac{\partial^2 u}{\partial x^2}(x, y) - \frac{\partial^2 u}{\partial y^2}(x, y) \\ &\approx -\frac{u(x+h, y) - 2u(x, y) + u(x-h, y)}{h^2} - \frac{u(x, y+h) - 2u(x, y) + u(x, y-h)}{h^2} \\ &= -\frac{u(x+h, y) + u(x-h, y) - 4u(x, y) + u(x, y+h) + u(x, y-h)}{h^2} =: -\Delta_h u \end{aligned}$$

Dabei hat der diskrete Laplace-Operator  $\Delta_h$  die Form eines Differenzensterns. Gesucht ist die Lsg  $u_h : \Omega_h \cup \partial\Omega_h \rightarrow \mathbb{R}$  von

$$(\text{RWP}_2)_h \quad \begin{cases} -\Delta_h u_h = f_h & \text{in } \Omega_h \\ u_h = g & \text{auf } \partial\Omega_h. \end{cases}$$

3. Aufstellen des linearen Gleichungssystems  $-\tilde{\Delta}_h \tilde{u}_h = f_h$ :

$$\tilde{u}_h = \begin{pmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{n-1, n-2} \\ u_{n-1, n-1} \end{pmatrix} \in \mathbb{R}^{(n-1)^2},$$

$$-\tilde{\Delta}_h = \frac{1}{h^2} \begin{pmatrix} A & -I & & 0 \\ -I & A & -I & \\ & \ddots & \ddots & \ddots \\ 0 & & -I & A & -I \\ & & & -I & A \end{pmatrix} \in \mathbb{R}^{(n-1)^2 \times (n-1)^2},$$

$$A = \begin{pmatrix} 4 & -1 & & 0 \\ -1 & 4 & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & 4 & -1 \\ 0 & & & -1 & 4 \end{pmatrix} \in \mathbb{R}^{n-1 \times n-1}$$

**Lem.** Das DV  $(\text{RWP}_2)_h$  ist konsistent von der Ordnung 2. Es gilt

$$\|\Delta_h R_h u - R_h \Delta u\|_h \leq \frac{1}{6} \|u\|_{C^2(\bar{\Omega})} h^2.$$

**Lem.** Das DV  $(\text{RWP}_2)_h$  ist stabil. Es gilt  $\|\tilde{D}_h^{-1}\|_\infty \leq 1/s$ .

**Satz.** Das DV  $(\text{RWP}_2)_h$  ist konvergent von der Ordnung 2, falls die Lösung von  $(\text{RWP}_2)$  zu  $C^4(\bar{\Omega})$  gehört. Es gilt

$$\|u_h - R_h u\|_h \leq 1/48 \|u\|$$

*Bem.* Durch die Einbeziehung weiterer Gitterpunkte zur Approximation des Differentialoperators lässt sich die Konvergenzordnung erhöhen:

$$\begin{aligned} -\Delta_h^{(9)} u(x, y) &= \frac{1}{12h^2} (u(x-2h, y) - 16u(x-h, y) + 30u(x, y) \\ &\quad - 16u(x+h, y) + u(x+2h, y) + u(x, y-2h) - 16u(x, y-h) \\ &\quad + 30u(x, y) - 16u(x, y+h) + u(x, y+2h)) \approx -\Delta u(x, y) \end{aligned}$$

Damit erreicht man die Konsistenzordnung 4.

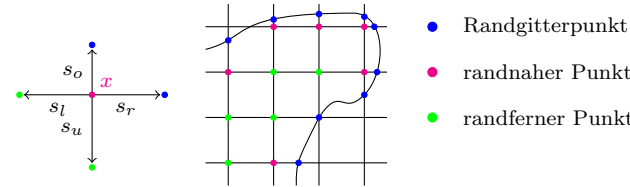
## Differenzenverfahren in allg. Gebieten $\Omega \subset \mathbb{R}^2$

**Situation.** Sei  $\Omega \subset \mathbb{R}^2$  beschränkt.

**Def.** •  $\Omega_h := \{x, y \in \Omega \mid x/h, y/h \in \mathbb{Z}\}$  heißen **innere Gitterpunkte**.

- Ein Punkt  $z_R \in \partial\Omega$  heißt **Randgitterpunkt** (notiert  $z_R \in \partial\Omega_h$ ), falls es einen inneren Gitterpunkt  $z \in \Omega_h$  gibt, sodass  $z_R = r + \alpha h e_1$  oder  $z_R = z + \alpha h e_2$  mit  $|\alpha| \leq 1$ . Die Nachbarn  $N(x, y)$  eines Punktes  $(x, y)$  sind  $(x + s_r h, y)$ ,  $(x - s_l h, y)$ ,  $(x, y + s_o h)$ ,  $(x, y - s_u h)$ , falls  $s_r, s_l, s_o, s_u \in (0, 1]$  und die Verbindungsstrecken zu  $(x, y)$  in  $\Omega$  liegen.
- Ein Punkt  $(x, y) \in \Omega_h$  heißt **randnah**, falls  $(x, y)$  die Nachbarn  $(x - s_l h, y)$ ,  $(x + s_r h, y)$ ,  $(x, y - s_u h)$ ,  $(x, y + s_o h)$  hat mit mindestens einem  $s_i < 1$ . Ansonsten heißt  $(x, y)$  **randfern**.

**Notation.** Wir haben eine Einteilung  $\Omega_h = \Omega_h^{\text{in}} \sqcup \Omega_h^{\text{f}}$  der Gitterpunkte in randnahe und randferne Punkte.



**Lem (Dividierte Differenzen von Newton).**

Für  $u \in C^3([x_l, x_r])$ ,  $x \in (x_l, x_r)$  gilt

$$\begin{aligned} u''(x) &= \frac{2}{x_r - x_l} \left( \frac{u(x_r) - u(x)}{x_r - x} - \frac{u(x) - u(x_l)}{x - x_l} \right) + \mathcal{O}(x_r - x_l) \\ &= \frac{2}{x_r - x_l} \left( \frac{1}{x_r - x} u(x_r) + \frac{1}{x - x_l} u(x_l) \right) - \frac{2}{(x_r - x)(x - x_l)} u(x) \end{aligned}$$

**Verfahren (Shortley-Weller-Diskretisierung).**

Dadurch inspiriert approximieren wir den Laplace-Operator durch

$$\begin{aligned} \mathcal{D}_h u(x, y) &= \frac{1}{h^2} \left( \frac{2u(x - s_l h, y)}{s_l(s_r + s_l)} + \frac{2u(x + s_r h, y)}{s_r(s_r + s_l)} \right. \\ &\quad \left. + \frac{2u(x, y - s_u h)}{s_u(s_o + s_u)} + \frac{2u(x, y + s_o h)}{s_o(s_o + s_u)} - \left( \frac{2}{s_l s_r} + \frac{2}{s_o s_u} \right) u(x, y) \right) \end{aligned}$$

wobei  $x_r - x = s_r h$ ,  $x - x_l = s_l h$ ,  $y_o - y = s_o h$ ,  $y - y_u = s_u h$ . Wir betrachten nun

$$(\text{RWP}_2)'_h \quad \begin{cases} -\mathcal{D}_h u_h = f_h & \text{in } \Omega_h \\ u_h = g & \text{auf } \partial\Omega_h \end{cases}$$

$$(\text{LGS}_2)' \quad \begin{cases} -\tilde{\mathcal{D}}_h \tilde{u}_h = \tilde{f}_h \\ \tilde{f}_h = f_h + g_h \end{cases}$$

$$\text{mit } g_h(x, y) = \frac{1}{h^2} \sum_{(x_N, y_N) \in N(x, y) \cap \partial\Omega_h} S_{x_N, y_N} g(x_N, y_N)$$

wobei

$$S_{x_N, y_N} := \begin{cases} 2/s_l(s_l + s_r) & \text{falls } (x_N, y_N) = (x - s_l h, y), \\ 2/s_r(s_l + s_r) & \text{falls } (x_N, y_N) = (x + s_r h, y), \\ 2/s_o(s_o + s_u) & \text{falls } (x_N, y_N) = (x, y + s_o h), \\ 2/s_u(s_o + s_u) & \text{falls } (x_N, y_N) = (x, y - s_u h), \end{cases}$$

$$-\tilde{\mathcal{D}}_h = (d_{ij}) \quad \text{mit} \quad d_{ii} = 1/h^2 \left( \frac{2}{s_{il} s_{ir}} + \frac{2}{s_{iu} s_{io}} \right) \quad \text{und}$$

$$d_{ij} = 1/h^2 \begin{cases} -2/s_{il}(s_{il} + s_{ir}) & \text{falls } j \text{ der linke Nachbar von } i \text{ ist,} \\ -2/s_{ir}(s_{il} + s_{ir}) & \text{falls } j \text{ der rechte Nachbar von } i \text{ ist,} \\ -2/s_{iu}(s_{iu} + s_{io}) & \text{falls } j \text{ der untere Nachbar von } i \text{ ist,} \\ -2/s_{io}(s_{iu} + s_{io}) & \text{falls } j \text{ der obere Nachbar von } i \text{ ist.} \end{cases}$$

**Lem.** • Die Matrix  $-\tilde{\mathcal{D}}_h$  ist eine M-Matrix.

- Sei  $\Omega \subset \mathbb{R}^2$  beschränkt und gehöre zu dem Streifen  $(x_0, x_0 + d) \times \mathbb{R}$  oder  $\mathbb{R} \times (y_0, y_0 + d)$ . Dann gilt  $\|\tilde{D}_h^{-1}\| \leq d^2/s$ .

*Bem.* Das DV  $(\text{RWP}_2)'_h$  hat in den randnahen Punkten nur die Konsistenzordnung 1. Dennoch gilt:

**Satz.** Sei  $\Omega \subset \mathbb{R}^3$  beschränkt und Teilmenge des Streifens  $(x_0, x_0 + d) \times \mathbb{R}$  oder  $\mathbb{R} \times (y_0, y_0 + d)$ . Dann ist das Verfahren  $(\text{RWP}_2)'_h$  konvergent von der Ordnung 2. Es gilt

$$\|u_h - R_h u\|_h \leq (1/3h^3 + d^2/48h^2) \|u\|_{C^4(\bar{\Omega})}.$$

**Idee.** Bestimme den Wert von  $u$  bei randnahen Punkten  $(x, y)$  durch lineare Interpolation:

- $u(x, y) \approx \frac{s_r}{s_r + s_l} u(x - s_l h, y) + \frac{s_l}{s_r + s_l} u(x + s_r h, y)$
- $u(x, y) \approx \frac{s_o}{s_u + s_o} u(x, y - s_u h) + \frac{s_u}{s_u + s_o} u(x, y + s_o h)$

$$\begin{aligned} (\text{RWP}_2)_h'' &\quad \begin{cases} -\mathcal{D}_h u = f_h & \text{in } \Omega_h \\ u_h = g & \text{auf } \partial\Omega_h \end{cases} \\ (\text{LGS}_2)'' &\quad -\tilde{\mathcal{D}}_h \tilde{u}_h = \tilde{f}_h \end{aligned}$$

**Lem.** Dieses Verfahren besitzt Konsistenzordnung (und somit Konvergenzordnung) 2.

## Allgemeine Differentialoperatoren

**Problem.** Wir betrachten nun

$$(\text{RWP}_3) \quad \begin{cases} -\mathcal{L}u = f & \text{in } \Omega = (0,1) \times (0,1) \\ u = g & \text{auf } \partial\Omega \end{cases}$$

mit

$$-\mathcal{L}u = -(a_{11}(x,y)u_{xx} + 2a_{12}(x,y)u_{xy} + a_{22}(x,y)u_{yy}) + b_1(x,y)u_x + b_2(x,y)u_y + c(x,y)u$$

wobei  $c(x,y) \leq 0$ ,  $\xi^T A(x,y)\xi \geq \lambda_0 \|\xi\|^2$ ,  $\lambda_0 > 0$  und

$$A(x,y) = \begin{pmatrix} a_{11}(x,y) & a_{12}(x,y) \\ a_{21}(x,y) & a_{22}(x,y) \end{pmatrix}$$

**Verfahren.** 1. Diskretisierung:  $h = 1/n$ ,  $\Omega_h$ ,  $\partial\Omega_h$  wie früher.  
2. Approximation:

$$u_x(x,y) \approx \frac{u(x+h,y) - u(x-h,y)}{2h}, \quad u_y(x,y) \approx \dots$$

$$u_{xx}(x,y) \approx \frac{u(x+h,y) - 2u(x,y) + u(x-h,y)}{h^2}, \quad u_{yy}(x,y) \approx \dots$$

Für die Approx. von  $u_{xy}$  haben wir mehrere Möglichkeiten:  
Wir könnten etwa den zentralen DQ in  $x$ - und  $y$ -Richtung verwenden und erhalten

$$u_{xy}(x,y) \approx \frac{1}{4h^2} (u(x+h,y+h) - u(x+h,y-h) - u(x-h,y+h) + u(x-h,y-h))$$

Diese Annäherung hat allerdings den Nachteil, dass sie zu keiner M-Matrix führt. Stattdessen nehmen wir

$$u_{xy}(x,y) \approx \frac{1}{2h^2} \begin{pmatrix} 0 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 0 \end{pmatrix} \quad \frac{1}{2h^2} \begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix}$$

für  $a_{12} \geq 0$                       für  $a_{12} < 0$ .

Wir fassen diese Approx. in folgendem 7-Stern zusammen:

$$-\mathcal{L}_h u := \frac{1}{h^2} \begin{pmatrix} a_{12}^- & |a_{12}| - a_{22} & a_{12}^+ \\ |a_{12}| - a_{11} & 2(a_{11} + a_{22} - |a_{12}|) & |a_{12}| - a_{11} \\ -a_{12}^+ & |a_{12}| - a_{22} & a_{12} \end{pmatrix} + \frac{1}{h} \begin{pmatrix} -b_2 & & \\ -b_1 & 0 & \\ & b_2 & b_1 \end{pmatrix} + \begin{pmatrix} & & \\ & c & \\ & & \end{pmatrix}$$

Dabei ist  $a_{ij}^+ := \max(a_{ij}, 0)$  und  $a_{ij}^- := \min(a_{ij}, 0)$ .

$$(\text{RWP}_3)_h \quad \begin{cases} -\mathcal{L}_h u_h = f_h & \text{in } \Omega_h \\ u_h = g & \text{auf } \partial\Omega_h \end{cases}$$

$$(\text{LGS}_3) \quad -\tilde{\mathcal{L}}_h \tilde{u}_h = \tilde{f}_h$$

**Satz.** Sei  $|a_{12}| \leq \min(a_{11}, a_{22})$ ,  $c \geq 0$  in  $\Omega$ ,  $\mathcal{L}$  gleichmäßig elliptisch.  
Falls  $a_{ii} > |a_{12}| + \frac{h}{2}|b_i|$  für  $i = 1, 2$  in  $\Omega$  und  $u \in \mathcal{C}^4(\bar{\Omega})$ , so ist das DV  $(\text{RWP}_3)_h$  konvergent von der Ordnung 2.

## Differenzenverfahren für parabolische DGLn

**Problem.** **Wärmeleitungsgleichung**

$$(\text{RWP}_4) \quad \begin{cases} u_t(x,t) - \Delta_x u(x,t) = f(x,t) & \text{in } \Omega = (0,1) \times (0,T) \\ u(x,0) = g(x) & \text{für } x \in (0,1) \\ u(0,t) = g_0(t) & \text{für } t \in [0,T] \\ u(1,t) = g_1(t) & \text{für } t \in [0,T] \end{cases}$$

**Verfahren.** 1. Diskretisierung mit  $n$  Raum- und  $m$  Zeitschritten:

$$x_i = ih, \quad h = 1/n, \quad t_k = k\tau, \quad \tau = T/m, \quad u(x_i, t_k) \approx u_i^k$$

2. Approximation der Ableitungen:

$$u_{xx}(x,t) \approx \frac{1}{h^2} (u(x-h,t) - 2u(x,t) + u(x+h,t)) =: \Delta_h u(x,t)$$

Wir wollen nun eine Lösung von

$$\begin{cases} \dot{u}_h(t) - \tilde{\Delta}_h u_h(t) = f_h(t) \\ u_h(0) = g_h \end{cases}$$

für alle Zeiten  $t$  mit

$$u_h(t) = \begin{pmatrix} u_h(h,t) \\ u_h(2h,t) \\ \vdots \\ u_h(1-h,t) \end{pmatrix}, \quad f_h(t) = \begin{pmatrix} f(h,t) + \frac{1}{h^2}g_0(t) \\ f(2h,t) \\ \vdots \\ f(1-h,t) + \frac{1}{h^2}g_1(t) \end{pmatrix}$$

berechnen. Dazu verwenden wir ein Einschrittverfahren, wie das expl./impl. Gauß-Verfahren oder das Crank-Nicolson-Verfahren:

$$(\text{EEV}) \quad \begin{cases} \frac{1}{\tau}(u_i^{k+1} - u_i^k) - \tilde{\Delta}_h u_i^k = f_i^k \\ u_i^0 = g_h \end{cases}$$

$$(\text{IEV}) \quad \begin{cases} \frac{1}{\tau}(u_i^{k+1} - u_i^k) - \tilde{\Delta}_h u_i^{k+1} = f_i^{k+1} \\ u_i^0 = g_h \end{cases}$$

$$(\text{CNV}) \quad \begin{cases} \frac{1}{\tau}(u_i^{k+1} - u_i^k) - \frac{1}{2}\tilde{\Delta}_h(u_i^k + u_i^{k+1}) = f(x_i, t_k + \frac{\tau}{2}) \\ u_i^0 = g_h \end{cases}$$

**Lem.** Sei  $f(x, -) \in \mathcal{C}^1([0, T])$  für alle  $x \in [0, 1]$ .

Dann gilt für die Approximation von  $(\text{RWP}_4)$ :

- Die Verfahren (EEV) und (IEV) besitzen einen Konsistenzfehler von  $\mathcal{O}(h^2 + \tau)$ , falls  $u \in \mathcal{C}^4([0, 1] \times [0, T])$
- Das Verfahren (CNV) besitzt einen Konsistenzfehler von  $\mathcal{O}(h^2 + \tau^2)$ , falls  $u \in \mathcal{C}^4([0, 1] \times [0, T])$ .

**Lem.** Es gelte  $2\tau \leq h^2$  für (EEV). Die Verfahren (EEV), (IEV) und (CNV) sind stabil.

## Differenzenverfahren für hyperbolische DGLn

**Problem.** **Wellengleichung**

$$\begin{cases} \partial_{tt}u - c^2\partial_{xx}u = f(x,t) & \text{in } \Omega = (0,1) \times [0,T] \\ u(0,t) = g_0(t), \quad u(1,t) = g_1(t) & \text{für } t \in [0,T] \\ u(x,0) = q_0(x), \quad u_t(x,0) = q_1(x) & \text{für } x \in (0,1) \end{cases}$$

**Verfahren.** 1. Diskretisierung:  $x_i = ih$ ,  $h = \frac{1}{n}$ ,  $t_k = k\tau$ ,  $\tau = \frac{T}{m}$

2. Approximation:

$$\partial_{xx}u(x_i, t_k) \approx \frac{1}{h^2} (u(x_{i-1}, t_k) - 2u(x_i, t_k) + u(x_{i+1}, t_k))$$

$$\partial_{tt}u(x_i, t_k) \approx \frac{1}{\tau^2} (u(x_i, t_{k-1}) - 2u(x_i, t_k) + u(x_i, t_{k+1}))$$

$$\partial_t u(x_i, 0) \approx \frac{1}{2\tau} (u(x_i, t_1) - u(x_i, t_{-1}))$$

Wir erhalten das lineare Gleichungssystem

$$\begin{cases} \frac{1}{\tau^2}(u_i^{k-1} - 2u_i^k + u_i^{k+1}) - \frac{c^2}{h^2}(U_{i-1}^k - 2u_i^k + u_{i+1}^k) = f_i^k \\ \text{für } i = 1, \dots, n-1 \text{ und } k = 0, \dots, m. \\ u_0^k = g_0^k = g_0(t_k), \quad u_n^k = g_1^k = g_1(t_k), \\ u_i^0 = q_0, i = q_0(x_i), \quad \frac{1}{2\tau}(u_i^1 - u_i^{-1}) = q_1, i = q_1(x_i) \end{cases}$$

*Bem.* Das Differenzenverfahren ...

- ☺ ... ist einfach in der Herleitung und Implementierung.
- ☺ ... besitzt eine gute Konvergenz (z. B. Ordnung 2) bei genügend glatter Lösung.
- ☺ ... ermöglicht Adaptivität bzw. unregelm. Gitter nur schwer.

## Schwache Lsgstheorie für elliptische DGLn

**Def.** Der  **$L^p$ -Raum** ist für  $1 \leq p < \infty$  definiert durch

$$L^p(\Omega) := \{v : \Omega \rightarrow \mathbb{R} \mid \|v\|_p < \infty\} \quad \text{mit} \quad \|v\|_p := \left( \int_{\Omega} |v(x)|^p dx \right)^{1/p},$$

für  $p = \infty$  durch

$$L^\infty(\Omega) := \{v : \Omega \rightarrow \mathbb{R} \mid \|v\|_\infty < \infty\} \quad \text{mit} \quad \|v\|_\infty := \operatorname{ess\,sup}_{x \in \Omega} |v(x)|.$$

*Bem.*  $(L^p(\Omega), \|\cdot\|_p)$  ist ein Banachraum, für  $p = 2$  sogar ein Hilbertraum mit Skalarprodukt  $\langle u, v \rangle_{L^2(\Omega)} := \int_{\Omega} u(x)v(x) dx$ .

**Satz (Höldersche Ungleichung).** Sei  $u \in L^p(\Omega)$  und  $v \in L^q(\Omega)$  mit  $1 \leq p, q, r \leq \infty$  und  $1/p + 1/q = 1/r$ . Dann ist  $uv \in L^r(\Omega)$  mit

$$\|uv\|_r \leq \|u\|_p \cdot \|v\|_q.$$

**Def.** Die Menge aller  $k$ -mal stetig differenzierbaren Funktionen auf  $\Omega$  mit **kompaktem Träger** ist

$$\mathcal{C}_0^k(\Omega) := \{\varphi \in \mathcal{C}^k(\Omega) \mid \operatorname{supp}(\varphi) := \overline{\{x \in \Omega \mid \varphi(x) \neq 0\}} \text{ ist kompakt}\}.$$

**Def.**  $\mathcal{D}(\Omega) := \mathcal{C}_0^\infty(\Omega)$  heißt Raum der **Testfunktionen** in  $\Omega$ .

**Lem** (Partielle Integration). Für  $u, v \in \mathcal{C}^1(\overline{\Omega})$  gilt

$$\int_{\Omega} v(x) \mathcal{D}_i u(x) dx = \int_{\partial\Omega} v(x) u(x) \eta_i(x) dx - \int_{\Omega} \mathcal{D}_i v(x) u(x) dx.$$

Für  $u \in \mathcal{C}^k(\overline{\Omega})$ ,  $\varphi \in \mathcal{C}_0^k(\Omega)$  und  $\alpha = (\alpha_1, \dots, \alpha_n)$ ,  $|\alpha| \leq k$  gilt

$$\int_{\Omega} \varphi(x) \mathcal{D}^\alpha u(x) dx = (-1)^{|\alpha|} \int_{\Omega} \mathcal{D}^\alpha \varphi(x) u(x) dx.$$

**Def.**  $L_{\text{loc}}^1(\Omega) := \{v : \Omega \rightarrow \mathbb{R} \mid v|_K \in L^1(K) \text{ für jedes kpkte } K \subset \Omega\}$  heißt Raum der **lokal integrierbaren Funktionen**.

**Def.** Sei  $u \in L_{\text{loc}}^1(\Omega)$  und  $\alpha \in \mathbb{N}^n$ . Eine Funktion  $v \in L_{\text{loc}}^1(\Omega)$  heißt **schwache (partielle) Ableitung** von  $u$  (oder die Ableitung von  $u$  im distributionellen Sinn) der Ordnung  $\alpha$ , wenn

$$\int_{\Omega} \varphi(x) v(x) dx = (-1)^{|\alpha|} \int_{\Omega} \mathcal{D}^\alpha \varphi(x) u(x) dx \quad \text{für alle } \varphi \in \mathcal{D}(\Omega).$$

*Bem.* Ist eine Funktion im klassischen Sinne differenzierbar, so auch im schwachen mit derselben Ableitung.

**Lem** (Fundamentallemma der Variationsrechnung). Für  $u \in L_{\text{loc}}^1$ :

$$\int_{\Omega} u(x) \varphi(x) dx = 0 \quad \forall \varphi \in \mathcal{D}(\Omega) \implies u \equiv 0 \text{ (fast-überall)}.$$

**Kor.** Die schwache Ableitung ist eindeutig bestimmt, d. h. sind  $v, w \in L_{\text{loc}}^1(\Omega)$  schwache Ableitungen von  $u$ , so gilt  $v \equiv w$  f. ü. in  $\Omega$ .

**Bsp.** Die schw. Abl. von  $u(x) = |x|$  ist  $v(x) = \mathbb{1}_{(0,\infty)} - \mathbb{1}_{(-\infty,0)}$ .

**Lem.** •  $\mathcal{D}^\alpha(u + \lambda v) = \mathcal{D}^\alpha u + \lambda \mathcal{D}^\alpha v$  •  $\mathcal{D}^{\alpha+\beta} u = \mathcal{D}^\alpha(\mathcal{D}^\beta u)$

**Def.** Der **Sobolev-Raum** für  $1 \leq p < \infty$  ist

$$W^{k,p}(\Omega) := \left\{ u \in L^p(\Omega) \mid \begin{array}{l} \forall \alpha \in \mathbb{N}^n \text{ mit } |\alpha| \leq k : \\ \exists \text{ schwache Ableitung } \mathcal{D}^\alpha u \in L^p(\Omega) \end{array} \right\}$$

$$\|u\|_{k,p} := \left( \sum_{|\alpha| \leq k} \|\mathcal{D}^\alpha u\|_p^p \right)^{1/p}.$$

**Notation.**  $H^k(\Omega) := W^{k,2}(\Omega)$

**Satz.**  $(W^{k,p}(\Omega), \|\cdot\|_{k,p})$  ist ein Banachraum.

*Bem.*  $H^k(\Omega)$  ist sogar ein Hilbertraum mit Skalarprodukt

$$\langle u, v \rangle_{H^k(\Omega)} := \int_{\Omega} \sum_{|\alpha| \leq k} \mathcal{D}^\alpha u \mathcal{D}^\alpha v dx.$$

**Satz** („ $H = W$ “).  $W^{k,p}(\Omega) \cap \mathcal{C}^\infty(\Omega)$  liegt dicht in  $W^{k,p}(\Omega)$ , d. h.

$$\overline{W^{k,p}(\Omega) \cap \mathcal{C}^\infty(\Omega)}^{\|\cdot\|_{k,p}} = W^{k,p}(\Omega).$$

**Def.**  $W_0^{k,p}(\Omega) := \overline{\mathcal{D}(\Omega)}^{\|\cdot\|_{k,p}}$ ,  $H_0^1(\Omega) := W_0^{1,2}(\Omega)$

**Satz.** Sei  $\Omega$  ein beschränktes  $\mathcal{C}^1$ -Gebiet und  $1 \leq p < \infty$ . Dann existiert eine lineare stetige Abbildung  $\tau : W^{1,p}(\Omega) \rightarrow L^p(\partial\Omega)$ , sodass für alle  $u \in W^{1,p}(\Omega) \cap \mathcal{C}(\overline{\Omega})$  gilt:  $\tau(u) = u|_{\partial\Omega}$ .

**Def.** Die Abbildung  $\tau$  heißt **Spuroperator**,  $\tau(u)$  heißt die **Spur** von  $u \in W^{1,p}(\Omega)$  auf  $\partial\Omega$ .

**Satz.** Sei  $\Omega$  ein beschränktes  $\mathcal{C}^1$ -Gebiet. Dann gilt

$$W_0^{1,p}(\Omega) = \{v \in W^{1,p}(\Omega) \mid \tau(v) = 0\}.$$

**Def.** Der **Dualraum** eines Banachraums  $(U, \|\cdot\|_U)$  ist

$$U' := \{\text{lineare, stetige Abbildungen } \psi : U \rightarrow \mathbb{R}\} \quad \text{mit}$$

$$\|\psi\|_{U'} := \sup_{u \in U \setminus \{0\}} \frac{|\psi(u)|}{\|u\|_U} = \sup_{u \in U, \|u\|_U = 1} \psi(u).$$

**Bsp.** Gelte  $1/p + 1/q = 1$  mit  $p, q \in (1, \infty)$ . Dann ist die Abb.

$$j : L^q(\Omega) \rightarrow (L^p(\Omega))', \quad f \mapsto (g \mapsto \int_{\Omega} f(x)g(x) dx)$$

ist ein isometrischer Isomorphismus.

**Notation.**  $\langle \psi, u \rangle_{U', U} := \psi(u)$  für  $\psi \in U'$ ,  $u \in U$ .

**Satz (Riesz'scher Darstellungssatz).**

Sei  $(H, \langle \cdot, \cdot \rangle_H)$  ein Hilbertraum. Dann ist

$$j : H \rightarrow H', \quad \psi \mapsto (\phi \mapsto \langle \psi, \phi \rangle_H)$$

ein isometrischer Isomorphismus.

**Def.**  $W^{-1,q} := (W_0^{1,p}(\Omega))'$ , wobei  $\frac{1}{p} + \frac{1}{q} = 1$ ,  
 $H^{-1}(\Omega) := W^{-1,2} = (H_0^1(\Omega))'$ .

## Variationsgleichungen

**Situation.** Sei  $\Omega \subset \mathbb{R}^d$  beschränkt. Wir betrachten nun wieder

$$\begin{aligned} (\text{RWP}_1) \quad & \begin{cases} \mathcal{L}u = f & \text{in } \Omega \\ u = 0 & \text{auf } \partial\Omega \end{cases} \\ \text{mit} \quad & \mathcal{L}u(x) = - \sum_{i=1}^d \mathcal{D}_i \left( \sum_{j=1}^d a_{ij}(x) \mathcal{D}_j u \right) + \sum_{i=1}^d b_i(x) \mathcal{D}_i u + c(x)u \\ & = -\operatorname{div}(A(x) \mathcal{D}u) + b(x) \cdot \mathcal{D}u + c(x)u. \end{aligned}$$

Sei  $u$  eine Lösung von (RWP<sub>1</sub>) und  $\phi \in \mathcal{D}(\Omega)$ . Dann gilt

$$\begin{aligned} \int_{\Omega} f(x) \phi(x) dx &= \int_{\Omega} \mathcal{L}u(x) \phi(x) dx \\ &= - \int_{\Omega} \operatorname{div}(A(x) \mathcal{D}u(x)) \phi(x) dx + \int_{\Omega} (b(x) \cdot \mathcal{D}u(x) + c(x)u(x)) \cdot \phi(x) dx \\ &= \int_{\Omega} A(x) \mathcal{D}u(x) \cdot \mathcal{D}\phi(x) dx + \int_{\Omega} (b(x) \cdot \mathcal{D}u(x) + c(x)u(x)) \cdot \phi(x) dx \end{aligned}$$

**Def.** Eine Funktion  $u \in H_0^1(\Omega)$  heißt **schwache Lösung** von (RWP<sub>1</sub>), wenn  $u$  folgende Variationsgleichung erfüllt:

$$\begin{aligned} \int_{\Omega} A(x) \mathcal{D}u(x) \cdot \mathcal{D}\phi(x) + b(x) \cdot \mathcal{D}u(x) \phi(x) + c(x)u(x) \phi(x) dx \\ = \int_{\Omega} f(x) \phi(x) dx \quad \text{für alle } \phi \in H_0^1(\Omega). \end{aligned} \quad (\text{VGL}_1)$$

**Sprechweise.** Der Raum, in dem man  $u$  sucht, heißt **Lsgsraum** (oder *Ansatzraum*), der Raum von  $\phi$  heißt **Testraum**.

**Problem** (Allg. Variationsproblem). Seien Abb.  $\ell : H_0^1(\Omega) \rightarrow \mathbb{R}$  und  $B : H_0^1(\Omega) \times H_0^1(\Omega) \rightarrow \mathbb{R}$  gegeben. Gesucht ist  $u \in H_0^1(\Omega)$ , sodass

$$(\text{VGL}_1)' \quad B(u, \phi) = \ell(\phi) \quad \forall \phi \in H_0^1(\Omega).$$

*Bem.* Im obigen Setting ist  $\ell(\phi) := \int_{\Omega} f(x) \phi(x) dx$  und

$$B(u, \phi) := \int_{\Omega} A(x) \mathcal{D}u(x) \cdot \mathcal{D}\phi(x) + b(x) \cdot \mathcal{D}u(x) \phi(x) + c(x)u(x) \phi(x) dx$$

**Def.** Sei  $X$  ein Banachraum. Eine bilin. Abb.  $B : X \times X \rightarrow \mathbb{R}$  heißt

- **positiv**, falls  $B(u, u) > 0$  für alle  $u \in X \setminus \{0\}$ ,
- **stark positiv** (oder **koerziv**), falls  $\lambda > 0$  existiert, sodass

$$\forall u \in X : B(u, u) \geq \lambda \|u\|_X^2,$$

- **beschränkt** (oder *stetig*), falls ein  $\mu > 0$  existiert, sodass

$$\forall u, \phi \in X : |B(u, \phi)| \leq \mu \|u\|_X \|\phi\|_X.$$

- Lem.** • Die Abbildung  $B$  in (VGL<sub>1</sub>)' ist bilinear und beschränkt.  
 • Die Abbildung  $\ell$  in (VGL<sub>1</sub>)' ist linear und stetig.

**Satz.** Sei  $\Omega$  ein beschr. Lipschitz-Gebiet. Dann ist jede klassische Lsg  $u \in \mathcal{C}^2 \cap \mathcal{C}^1(\partial\Omega)$  von (RWP<sub>1</sub>) eine schwache Lsg von (VGL<sub>1</sub>)'.



## Eindeutige Lösung elliptischer DGLn

**Satz (Lax-Milgram).** Sei  $H$  ein Hilbertraum und  $B : H \times H \rightarrow \mathbb{R}$  eine beschränkte, koerzitive Bilinearform. Dann gibt es für jedes  $\ell \in H'$  eine eindeutige Lösung  $u \in H$  von  $\forall \phi \in H : B(u, \phi) = \ell(\phi)$ . Es gilt  $\|u\|_H \leq 1/\lambda \|\ell\|_{H'}$  mit der Koerzitivitätskonstante  $\lambda$  von  $B$ .

**Lem (Poincaré-Ungleichung).**

Sei  $\Omega \subset \mathbb{R}^d$  beschränkt. Dann existiert eine Konstante  $C > 0$ , sodass

$$\|u\|_{L^2(\Omega)} \leq C \|\nabla u\|_{L^2(\Omega, \mathbb{R}^d)} = C \left( \int_{\Omega} |\sum_i \mathcal{D}_i u|^2 dx \right)^{1/2} \quad \forall u \in H_0^1(\Omega).$$

**Kor.** Mit  $C_1 := (1 + C^2)^{-1/2}$  und  $C_2 := 1$  gilt für alle  $u \in H_0^1(\Omega)$ :

$$C_1 \|u\|_{H^1(\Omega)} \leq \|\nabla u\|_{L^2(\Omega, \mathbb{R}^d)} \leq C_2 \|u\|_{H^1(\Omega)}.$$

**Lem.** Falls  $b(x) \equiv 0$  und  $c(x) \geq 0$  in  $\Omega$ , so ist  $B$  in  $(VGL_1)'$  koerziv.

**Satz.** Sei  $\Omega \subset \mathbb{R}^d$  beschränkt und sei  $\mathcal{L}u = -\operatorname{div}(A(x)\mathcal{D}u) + c(x)u$  glm. elliptisch,  $c(x) \geq 0$  in  $\Omega$ ,  $a_{ij}, c_j \in L^\infty(\Omega)$  und  $f \in L^2(\Omega)$ .

Dann besitzt  $(VGL_1)'$  eine eindeutige Lösung  $u \in H_0^1(\Omega)$ .

Außerdem existiert ein  $\hat{C} > 0$ , sodass  $\|u\|_{H^1(\Omega)} \leq \hat{C} \cdot \|f\|_{L^2(\Omega)}$ .

**Def.** Sei  $f \in H^{-1}(\Omega)$ . Eine Fktn  $u \in H_0^1(\Omega)$  heißt *schwache Lösung* von  $(RWP_1)$ , falls  $B(u, \phi) = \langle f, \phi \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \quad \forall \phi \in H_0^1(\Omega)$ .

*Bem.* Gelte  $b \equiv 0$ ,  $c \geq 0$ , glm. Elliptizität,  $c, a_{ij} \in L^\infty(\Omega)$ .

Dann existiert nach Lax-Milgram genau eine schwache Lösung.

**Lem.** Sei  $A$  gleichmäßig elliptisch und  $a_{ij}, b_i, c \in L^\infty(\Omega)$ .

Dann existiert ein  $\mu_0 > 0$ , sodass für alle  $\mu > \mu_0$  das RWP

$$\begin{cases} \mathcal{L}u + \mu u &= f & \text{in } \Omega \\ u &= 0 & \text{auf } \partial\Omega \end{cases}$$

für alle  $f \in H^{-1}(\Omega)$  eine eindeutige Lösung  $u \in H_0^1(\Omega)$  besitzt.

## RWPe mit anderen Randbedingungen

**Problem.** Wir untersuchen nun das inhomogene Randwertproblem

$$(RWP_2) \quad \begin{cases} \mathcal{L}u &= f & \text{in } \Omega, \\ u &= g & \text{auf } \partial\Omega. \end{cases}$$

Angenommen,  $g \in \mathcal{C}(\partial\Omega) \cap L^2(\partial\Omega)$  besitzt eine Fortsetzung

$\tilde{g} \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\bar{\Omega})$  mit  $\tilde{g}|_{\partial\Omega} = g$ . Dann ist  $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\bar{\Omega})$  genau dann eine Lösung von  $(RWP_2)$ , wenn  $v := u - \tilde{g}$  eine Lösung von

$$(RWP_2)', \quad \begin{cases} \mathcal{L}v &= f - \mathcal{L}\tilde{g} & \text{in } \Omega, \\ v &= 0 & \text{auf } \partial\Omega \end{cases}$$

ist. Schwache Formulierung von  $(RWP_2)'$ : Ges. ist  $v \in H_0^1(\Omega)$  mit

$$\begin{aligned} & \int_{\Omega} (A(x)\mathcal{D}v(x) \cdot \mathcal{D}\phi(x) + b(x) \cdot \mathcal{D}v(x)\phi(x) + c(x)v(x)\phi(x)) dx \\ &= \int_{\Omega} f \cdot \phi dx + \int_{\Omega} (A(x)\mathcal{D}\tilde{g} \cdot \mathcal{D}\phi + b \cdot \mathcal{D}\tilde{g}\phi + c\tilde{g}\phi) dx \quad \forall \phi \in H_0^1(\Omega). \end{aligned}$$

Voraussetzungen:  $a_{ij}, b_i, c \in L^\infty(\Omega)$ ,  $f \in L^2(\Omega)$ ,  $\tilde{g} \in \mathcal{C}(\bar{\Omega}) \cap H^1(\Omega)$  und  $g \in L^2(\partial\Omega)$ . Ges. ist ein  $u \in U := \{w \in H^1(\Omega) \mid \tau(w) = g\}$  mit

$$\underbrace{\int_{\Omega} A(x)\mathcal{D}u \cdot \mathcal{D}\phi + b\mathcal{D}u\phi + cu\phi dx}_{B(u, \phi) :=} = \underbrace{\int_{\Omega} f\phi dx}_{\ell(\phi) :=} \quad \forall \phi \in H_0^1(\Omega). \quad (VGL_2)$$

Für  $f \in H^{-1}(\Omega)$  verwendet man

$$B(u, \phi) = \ell'(\phi) := \langle f, \phi \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \quad \forall \phi \in H_0^1(\Omega). \quad (VGL_2)',$$

**Satz.** Sei  $B : H_0^1(\Omega) \times H_0^1(\Omega) \rightarrow \mathbb{R}$  beschränkt und koerziv.

Dann besitzt  $(VGL_2)$  genau dann eine eindeutige Lösung  $u \in U$ , wenn ein  $u_0 \in H^1(\Omega)$  existiert, sodass  $\tau(u_0) \equiv g$  f. ü. auf  $\partial\Omega$ .

**Problem.** Wir betrachten nun die Randbedingung

$$(RWP_3) \quad \begin{cases} \mathcal{L}u &= f & \text{in } \Omega, \\ A(x)\mathcal{D}u \cdot \nu + \mu u &= g & \text{auf } \partial\Omega \text{ (glatt)} \end{cases}$$

Falls  $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\bar{\Omega})$  eine Lösung ist und  $\phi \in \mathcal{C}^\infty(\bar{\Omega})$ , so gilt

$$\begin{aligned} \int_{\Omega} f\phi dx &= - \int_{\Omega} \operatorname{div}(A(x)\mathcal{D}u)\phi + b(x)\mathcal{D}u\phi + c(x)u\phi dx \\ &= - \int_{\partial\Omega} A(x)\mathcal{D}u \cdot \nu \phi ds + \int_{\Omega} A(x)\mathcal{D}u\mathcal{D}\phi + b(x)\mathcal{D}u\phi + c(x)u\phi dx. \end{aligned}$$

Aus der Randbedingung bekommen wir

$$\int_{\partial\Omega} A(x)\mathcal{D}u \cdot \nu \phi ds + \mu \int_{\Omega} u\phi dx = \int_{\partial\Omega} g\phi ds.$$

Zusammengesetzt erhalten wir die Variationsgleichung

$$\mu \int_{\Omega} u\phi dx + \int_{\Omega} A\mathcal{D}u\mathcal{D}\phi + b \cdot \mathcal{D}u\phi + cu\phi dx = \int_{\Omega} f\phi dx + \int_{\partial\Omega} g\phi ds.$$

Wegen Dichtheit von  $\mathcal{C}^\infty(\bar{\Omega})$  in  $H^1(\Omega)$  ist diese Gleichung nicht nur für  $\phi \in \mathcal{C}^\infty(\bar{\Omega})$  sondern allgemeiner für  $\phi \in H^1(\Omega)$  erfüllt.

**Def.** Sei  $\mu \in \mathbb{R}$ ,  $f \in L^2(\Omega)$ ,  $g \in L^2(\partial\Omega)$ . Eine Fktn  $u \in H^1(\Omega)$  heißt *schwache Lösung* von  $(RWP_3)$ , falls für alle  $\phi \in H^1(\Omega)$  gilt:

$$\underbrace{\mu \int_{\Omega} u\phi dx + \int_{\Omega} A\mathcal{D}u\mathcal{D}\phi + b\mathcal{D}u\phi + cu\phi dx}_{B(u, \phi) :=} = \underbrace{\int_{\Omega} f\phi dx + \int_{\partial\Omega} g\phi ds}_{\ell(\phi) :=}.$$

## Approximation von Variationsgleichungen

**Verfahren.** Gegeben sei ein Hilbertraum  $H$ , eine beschränkte, koerzive Bilinearform  $B : H \times H \rightarrow \mathbb{R}$  und ein  $\ell \in H'$ .

Gesucht ist eine Lösung  $u \in H$  der Variationsgleichung

$$(VGL) \quad B(u, \phi) = \ell(\phi) \quad \forall \phi \in H.$$

Wir wollen diese Lösung annähern durch die Lösung eines möglichst ähnlichen, aber *endlichdim.* Problems. Dazu wählen wir einen endlichdim. Unterraum  $U_n \subset H$  (dieser ist wieder ein Hilbertraum), eine beschränkte, koerzitive Bilinearform  $B_n : U_n \times U_n \rightarrow \mathbb{R}$  und ein Element  $\ell_n \in U_n'$ . Wir bestimmen dann die Lösung  $u_n$  von

$$(VGL)_n \quad B_n(u_n, \phi) = \ell_n(\phi) \quad \forall \phi \in U_n.$$

**Fragen.** 1. Wie berechnet man die Lösung  $u_n$  von  $(VGL)_n$ ?

2. Wie kann man  $U_n$ ,  $B_n$  und  $\ell_n$  wählen, sodass  $u_n \xrightarrow{n \rightarrow \infty} u$ ?

**Def.** Die Approximation von  $(VGL)$  mittels  $(VGL)_n$  mit  $U_n \subset H$  und  $B_n = B|_{U_n \times U_n}$  heißt **konforme Approximation** von  $(VGL)$ .

Eine solche Methode wird als **Verfahren von Ritz** bezeichnet.

**Vorgehen.** Um die Lösung  $u_n$  von  $(VGL)_n$  zu berechnen, wählen wir zunächst eine Basis  $w_1, \dots, w_{d_n}$  von  $U_n$ . Wir setzen

$$\hat{\ell} := (\ell_n(w_1), \dots, \ell_n(w_{d_n}))^T \in \mathbb{R}^{d_n}, \quad \text{und}$$

$$\hat{B} := (B_{ij}) \in \mathbb{R}^{d_n \times d_n} \quad \text{mit} \quad B_{ij} := B(w_i, w_j).$$

Dann ist  $u_n = \gamma_1 w_1 + \dots + \gamma_{d_n} w_{d_n} \in U_n$  genau dann eine Lösung von  $(VGL)_n$ , wenn  $\gamma = (\gamma_1, \dots, \gamma_{d_n})^T \in \mathbb{R}^{d_n}$  erfüllt:

$$\hat{B}\gamma = \hat{\ell} \quad (\text{Galerkin-Gleichung}).$$

**Lem (Céa).** Für die Lsgn  $u_n$  der konformen Approximation  $(VGL)_n$  und die Lösung  $u$  von  $(VGL)$  gilt

$$\|u_n - u\|_H \leq C \left( \inf_{v \in U_n} \|u - v\|_H + \|\ell_n - \ell\|_{U_n'} \right)$$

**Folgerung.** Für  $u_n \xrightarrow{n \rightarrow \infty} u$  genügt es,  $U_n$  u.  $\ell_n$  so zu wählen, dass

$$\forall u \in H : \inf_{v_n \in U_n} \|u - v_n\|_H \xrightarrow{n \rightarrow \infty} 0, \quad \|\ell_n - \ell\|_{U_n'} \xrightarrow{n \rightarrow \infty} 0.$$

*Bem.* Ist  $\ell(\phi)$  durch Integration einer Funktion gegeben, so kann man für  $\ell_n(\phi)$  eine Annäherung dieses Integrals etwa mittels der summierten Trapezregel verwenden.

*Bem.* Wir betrachten  $(VGL_1)$ . Es gibt mehrere sinnvolle Möglichkeiten, die Basiselemente  $w_i$  zu wählen. Man versucht dabei zu erreichen, dass die Matrix  $\hat{B}$  möglichst einfach (wenige von null verschiedene Einträge) und gut konditioniert ist.

1. Angenommen, es gibt eine Basis von Eigenfunktionen  $w_j$  von  $\mathcal{L}$ ,  $\mathcal{L}w_j = \lambda_j w_j$ , die in  $L^2(\Omega)$  eine Orthonormalbasis bilden.

Dann ist  $\hat{B}$  eine Diagonalmatrix mit Einträgen  $\lambda_j$ . Beispielsweise ist  $w_j(x) := \sin(\pi j x)$  eine EF von  $\mathcal{L} := -\Delta$  auf  $\Omega = (0, 1)$  zum EW  $\pi^2 j^2$  und es gilt  $\langle w_i, w_j \rangle_{L^2(\Omega)} = 0$  für  $i \neq j$ .

2.  $U_n := \operatorname{span}\{w_j(x) = x^j(1-x) \mid j = 1, \dots, n\} \subset H_0^1(\Omega)$

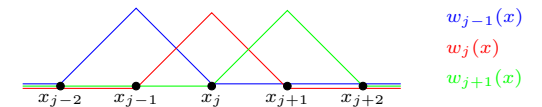
(Das ist eine schlechte Wahl, da dann  $\hat{B}$  vollbesetzt.)

3. Wir unterteilen  $\Omega = (0, 1)$  durch das Gitter

$$0 = x_0 < x_1 < \dots < x_n < x_{n+1} = 1, \quad x_i = ih \text{ mit } h = 1/(n+1),$$

$$U_n := \{v \in \mathcal{C}(0, 1) \mid \forall i : v|_{[x_{i-1}, x_i]} \in \mathbb{P}_1[x], v(0) = v(1) = 0\}$$

Eine Basis von  $U_n$  sind die *Hutfunktionen*



Wir betrachten die VGL zu  $(RWP_1)$  mit  $g_0 = g_1 = 0$ .

Wenn wir  $\ell$  durch die Trapezregel approximieren, erhalten wir

$$\hat{B} = \frac{1}{h} \begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \\ 0 & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{pmatrix}, \quad \ell_n(w_j) = hf(x_j),$$

also das Finite-Differenzen-Verfahren.

# Methode der Finiten Elemente (FEM)

**Ziel.** (VGL) mittels (VGL)<sub>n</sub> approximieren.

**Idee.** Zerlege  $\Omega$  in endlich viele Teilgebiete. Wir wählen  $U_n$  als Raum der Funktionen, die sich als Linearkombination von Basisfktn schreiben lassen, deren Träger nur auf wenige Teilgebiete umfasst.

*Bem.* Als Teilgebiete verwendet man in  $\mathbb{R}^1$  regelmäßige Teilintervalle, in  $\mathbb{R}^2$  Dreiecke oder Rechtecke und in  $\mathbb{R}^3$  Tetraeder. Als lokale Ansatzfunktionen über den Teilgebieten verwenden wir Polynome. Globale Ansatzfktn über  $\Omega$  sind lokale Ansatzfktn mit bestimmten Glattheitsbedingungen am Rand der Teilgebiete.

**Def.** Ein **finites Element** (FE) in  $\mathbb{R}^d$  ist ein Tripel  $(K, P, \Sigma)$  mit

- $K \subset \mathbb{R}^d$  ist kompakt,
- $\partial K$  ist Lipschitz-stetig,
- $P$  ist ein endlichdim. lin. Raum von Funktionen  $p \in C^s(K, \mathbb{R})$
- $\Sigma = \{b_1, \dots, b_m\}$  mit  $b_j \in (C^s(K, \mathbb{R}))'$  ist **P-unisolvant**, d. h.

$$\forall \alpha \in \mathbb{R}^m : \exists! p \in P : b_j(p) = \alpha_j, \quad j = 1, \dots, m$$

*Bem.*  $\Sigma$  ist  $P$ -unisolvant  $\iff \Sigma$  ist Basis von  $P' \subset (C^s(K, \mathbb{R}))'$

**Lem.** • Sei  $\Sigma = \{b_1, \dots, b_m\}$  P-unisolvant.

- $b_1, \dots, b_m$  sind linear unabhängig.
- Sei  $p_j \in P$  so gewählt, dass  $b_i(p_j) = \delta_{ij}$ . Dann ist  $\{p_1, \dots, p_m\}$  eine Basis von  $P$ .
- Sei  $\{p_1, \dots, p_m\}$  eine Basis von  $P$  und seien  $b_i \in (C^s(K, \mathbb{R}))'$  mit  $b_i(p_j) = \delta_{ij}$ ,  $i, j = 1, \dots, m$ . Dann ist  $\{b_1, \dots, b_m\}$  P-unisolvant.

## Finite Elemente vom Lagrange-Typ

**Def.** Der **d-Simplex** mit Ecken  $a_1, \dots, a_{d+1} \in \mathbb{R}^d$  ist

$$K = \{x = \sum_{j=1}^{d+1} \mu_j a_j \mid 0 \leq \mu_j \leq 1, j = 1, \dots, d+1, \sum_{j=1}^{d+1} \mu_j = 1\} \subset \mathbb{R}^d.$$

Dabei heißen  $\mu_1, \dots, \mu_{d+1}$  baryzentrische Koordinaten von  $x$ .  $K$  heißt **nicht entartet**, falls  $a_1, \dots, a_{d+1}$  affin unabhängig sind. Das Simplex mit den Ecken  $a_j = e_j \in \mathbb{R}^d$ ,  $j = 1, \dots, d$ , und  $a_{d+1} = 0$  heißt **d-Einheitssimplex**  $\hat{K}$ .

**Bspe.** Ein nichtentartetes Simplex im  $\mathbb{R}^1$  ist ein geschlossenes Intervall, im  $\mathbb{R}^2$  ein Dreieck und im  $\mathbb{R}^3$  ein Tetraeder.

**Lem.** Jedes nicht entartete  $d$ -Simplex  $K$  ist affin äquivalent zu  $\hat{K}$ : Es gibt genau eine Abb.  $F: \hat{K} \rightarrow K$ ,  $\hat{x} \mapsto A_K \hat{x} + b_K$  mit einer Matrix  $A_K \in \text{GL}(d)$ , sodass  $F(e_j) = a_j$  und  $F(0) = a_{d+1}$ .

**Def.** Ein **simpliziales finites Element vom Lagrange-Typ** der Ordnung  $k$  ist ein Tripel  $(K, P, \Sigma)$  mit

- einem  $d$ -Simplex  $K$ ,
- $P = \mathbb{P}_k := \mathbb{R}[x_1, \dots, x_d]_{\leq k} = \{\text{Polynome vom Grad } \leq k\}$
- $\Sigma = \{b_a : P \rightarrow \mathbb{R}, p \mapsto p(a) \mid a \in \mathcal{K}_k\}$  mit der *Knotenmenge*

$$\mathcal{K}_k = \{x = \sum_{j=1}^{d+1} \mu_j a_j \mid \sum_{j=1}^{d+1} \mu_j = 1, \mu_j \in \{0, \frac{1}{k}, \dots, \frac{k-1}{k}, 1\}\}.$$

*Bem.* • Sei  $K$  ein  $d$ -Simplex. Dann ist  $\dim(\mathbb{P}_k) = |\mathcal{K}_k| = \binom{k+d}{d}$ .

- Die kanonische Basis von  $\mathbb{P}_1$  für  $\hat{K} \subset \mathbb{R}^d$  ist

$$\hat{p}_i(\hat{x}) = \hat{x}_i \text{ für } i = 1, \dots, d, \quad \hat{p}_{d+1}(\hat{x}) = 1 - \hat{x}_1 - \dots - \hat{x}_d.$$

Dann bilden die  $\hat{b}_j \in \Sigma$  eine Dualbasis der  $\hat{p}_i$ , d. h.  $\hat{b}_j(\hat{p}_i) = \delta_{ij}$ .

- Die duale Basis von  $\mathbb{P}_2$  zu  $\Sigma$  für  $\hat{K} \subset \mathbb{R}^d$  ist

$$\begin{aligned} \{\hat{p}_i \mid i \in \{1, \dots, d+1\}\} \cup \{\hat{p}_{ij} \mid i \neq j \in \{1, \dots, d+1\}\} \quad \text{wobei} \\ \hat{p}_i(\hat{x}) = \mu_j(\hat{x})(2\mu_j(\hat{x}) - 1), \quad \hat{p}_{ij}(\hat{x}) = 4\mu_i(\hat{x})\mu_j(\hat{x}), \\ \mu_i(\hat{x}) = \hat{x}_i \text{ für } i = 1, \dots, d, \quad \mu_{d+1}(\hat{x}) = 1 - \hat{x}_1 - \dots - \hat{x}_d \end{aligned}$$

- Kanonische Basiselemente  $p_1, \dots, p_m \in \mathbb{P}_k$  für einen allgemeinen Simplex  $K \subset \mathbb{R}^d$  kann man wie folgt berechnen:

1. Finde eine affin lineare Bijektion  $F: \hat{K} \rightarrow K$  wie früher.
2. Setze  $p_i(x) := \hat{p}_i(F^{-1}(x))$  für  $i = 1, \dots, m$ .

## Räume von Finite-Elemente-Funktionen auf $\Omega$

**Def.** Sei  $\Omega \subset \mathbb{R}^d$  ein beschränktes, polygonal berandetes Gebiet. Eine **Triangulierung** von  $\Omega$  mit simplizialen finiten Elementen vom Lagrange-Typ der Ordnung  $k$  ist eine endliche Menge

$$T(\bar{\Omega}) = \{(K_i, P(K_i), \Sigma(K_i)) \mid i = 1, \dots, N\} \quad \text{mit}$$

- $(K_i, P(K_i), \Sigma(K_i))$  sind simpl. El. vom Lagrange-Typ der Ord.  $k$ ,
- $\bar{\Omega} = K_1 \cup \dots \cup K_N$ , •  $\text{int}(K_i) \cap \text{int}(K_j) = \emptyset$  für  $i \neq j$ ,
- Jede Seite von  $K_i$ , d. h. jedes von  $d$  Eckpunkten von  $K_i$  aufgespannte  $(d-1)$ -dimensionale Simplex, ist entweder Teil des Gebietrandes  $\partial\Omega$  oder gleichzeitig Seite eines anderen Simplex  $K_j$ .

**Def.** Die *Knotenmenge* von  $T(\bar{\Omega})$  ist die Vereinigung der Knotenmengen der simpl. FE, d. h.  $\mathcal{K}_k = \mathcal{K}_k(K_1) \cup \dots \cup \mathcal{K}_k(K_N)$ .

**Def.** Der **Raum der finiten Elemente** zu einer Triang.  $T(\bar{\Omega})$  von  $\Omega$  mit simplizialen FE vom Lagrange-Typ der Ordnung  $k$  und Knotenmenge  $\mathcal{K}_k = \{\tilde{a}_1, \dots, \tilde{a}_n\}$  ist

$$U_n = \{v \in C(\bar{\Omega}) \mid v|_{K_i} \in \mathbb{P}_1(K_i) \text{ für } i = 1, \dots, N\}.$$

**Satz** ( $k=1$ ). Sei  $U_n$  der Raum der *linearen finiten Elemente* zu einer Triang.  $T(\bar{\Omega})$  mit simpl. FE vom Lagrange-Typ der Ord. 1.

- Sei  $K$  ein nichtentart.  $d$ -Simplex mit Ecken  $a_1, \dots, a_{d+1}$ . Dann ist durch  $p(a_j), \dots, p(a_{d+1})$  ein Polynom  $p \in \mathbb{P}_1(K)$  eindeutig bestimmt. Für alle  $p \in \mathbb{P}_1(K)$  und  $x \in K$  gilt

$$p(x) = p(a_1)p_1(x) + \dots + p(a_{d+1})p_{d+1}(x). \quad \text{wobei } p_i(a_j) = \delta_{ij}.$$

- Sind  $\mathcal{K} = \{\tilde{a}_1, \dots, \tilde{a}_n\}$  die Knoten der Triangulierung, so ist eine Funktion  $v \in U_n$  durch die Vorgabe von  $v(\tilde{a}_1), \dots, v(\tilde{a}_n)$  eindeutig definiert. Es gilt  $U_n \subset H^1(\Omega)$ .
- Eine Basis von  $U_n$  ist gegeben durch die Funktionen  $p_j \in U_n$  mit  $p_j(\tilde{a}_i) = \delta_{ij}$  für  $i, j = 1, \dots, n$ . Insbesondere gilt  $\dim U_n = n$ .

**Satz** ( $k \geq 1$ ). Sei  $T(\bar{\Omega})$  eine Triangulierung mit finiten Lagrange-Elementen der Ordnung  $k$ ,  $U_n$  der zugehörige Raum der FE und  $\mathcal{K}_k$  die Knotenmenge von  $T(\bar{\Omega})$ . Dann ist durch Vorgabe von  $v|_{\mathcal{K}_k}$  eindeutig ein  $v \in U_n \subset H^1(\Omega)$  bestimmt. Eine Basis von  $U_n$  ist durch  $p_j \in U_n$  mit  $p_j(\tilde{a}_i) = \delta_{ij}$ ,  $i, j = 1, \dots, n$ , gegeben.

**Verfahren** (Realisierung der Finite-Elemente-Methode).

1. Eingabe und Beschreibung des RWP's
2. Umformulierung in ein Variationsproblem
3. Generierung einer Triangulierung. Entweder *uniforme Zerlegung* oder Zerlegung mit *lokaler Verfeinerung* von  $\Omega$ .
4. Erzeugung eines endlich-dim. Problems, d. h. Berechnen der Koeffizientenmatrix und der rechten Seite der Galerkin-Gleichung
5. Lösung der Galerkin-Gleichung

## Konvergenz der FE-Methode

**Def.** Sei  $(K, P, \Sigma)$  ein finites Element,  $P \subset C^s(K)$ .

Dann heißt  $\Pi_K w$  **P-Interpolierende** einer Fktn  $w \in C^s(K)$ , falls

- $\Pi_K w \in P$ ,
- $b_j(\Pi_K w) = b_j(w)$  für jedes  $b_j \in \Sigma$ .

*Bem.* • Ist  $p_1, \dots, p_m$  eine zu  $\Sigma$  duale Basis von  $P$ , d. h.

$$b_i(p_j) = \delta_{ij}, \text{ so gilt } \Pi_K w = \sum_{i=1}^m b_i(w)p_i.$$

- Für Lagrange-FE gilt  $w(\tilde{a}_j) = b_j(w) = b_j(\Pi_K w) = \Pi_K w(\tilde{a}_j)$
- $\forall p \in P : \Pi_K p = p$

**Def.** Sei  $U_n$  ein FE-Raum zu einer Triangulierung  $T(\bar{\Omega})$  und sei  $\{p_1, \dots, p_n\}$  die kanonische Basis von  $U_n$ , d. h.  $b_i(p_j) = \delta_{ij}$ . Die  **$U_n$ -Interpolierende** einer Funktion  $w \in C^s(\bar{\Omega})$  ist dann

$$\Pi w := \sum_{i=1}^n b_i(w)p_i \in U_n.$$

**Lem.** Für alle  $K_i \in T(\bar{\Omega})$  und  $w \in C^s(\bar{\Omega})$  gilt  $(\Pi w)|_K = \Pi_{K_i}(w|_{K_i})$  und somit  $\|w - \Pi w\|_{H^1(\Omega)} = \sum_{K_i \in T(\bar{\Omega})} \|w - \Pi_{K_i} w\|_{H^1(K_i)}$ .

**Lem.** Sei  $F: \hat{K} \rightarrow K$  mit  $F(\hat{x}) = A\hat{x} + b$ ,  $A \in \text{GL}(d)$ , und  $l \in \mathbb{N}$ .

- Es existiert eine Konstante  $c > 0$ , sodass

$$|v \circ F|_{H^l(\hat{K})} \leq c \cdot \|A\|_2^l \cdot \frac{1}{\sqrt{|\det(A)|}} \cdot |v|_{H^l(K)}$$

$$|\hat{v} \circ F^{-1}|_{H^l(K)} \leq c \cdot \|A^{-1}\|_2^l \cdot \sqrt{|\det(A)|} \cdot |\hat{v}|_{H^l(\hat{K})}$$

für alle  $v \in H^l(K)$  bzw.  $\hat{v} \in H^l(\hat{K})$  gilt, wobei

$$|v|_{H^l(K)} := \left( \int_K \sum_{|\alpha|=l} \|\mathcal{D}^\alpha v\|^2 dx \right)^{1/2}$$

eine Seminorm auf  $H^l(K)$  ist.

- Es gilt  $v \in H^l(K) \iff v \circ F \in H^l(\hat{K})$

**Def.** Sei  $K$  ein  $d$ -Simplex mit Ecken  $a_1, \dots, a_{d+1}$ . Wir definieren:

$$h(K) := \max_{i,j=1}^{d+1} |a_i - a_j| \quad \text{Durchmesser}$$

$$\rho(K) := 2 \sup \{R > 0 \mid \exists x \in K : B_R(x) \subseteq K\} \quad \text{Innendurchmesser}$$

$$\sigma(K) := h(K)/\rho(K) > 1 \quad (\text{misst „Spitzheit“})$$

**Lem.** Sei der  $d$ -Simplex  $K$  affin äquivalent zu  $\hat{K}$  vermöge  $F: \hat{K} \rightarrow K$ ,  $\hat{x} \mapsto A\hat{x} + b$ ,  $A \in \text{GL}(d)$ . Dann gilt:

$$\|A\|_2 \leq h(K)/\rho(\hat{K}), \quad \|A^{-1}\|_2 \leq h(\hat{K})/\rho(K).$$

**Lem.** Sei  $k \geq 0$ . Dann existiert eine Konstante  $c > 0$ , sodass

$$\inf_{\hat{p} \in \mathbb{P}_k(\hat{K})} \|\hat{v} - \hat{p}\|_{H^{k+1}(\hat{K})} \leq c \cdot |\hat{v}|_{H^{k+1}(\hat{K})} \quad \forall \hat{v} \in H^{k+1}(\Omega).$$

**Satz** (Abschätzung des lokalen Interpolationsfehlers).

Seien  $(\hat{K}, P(\hat{K}), \Sigma(\hat{K}))$  ein finites Element vom Lagrange-Typ.

Dann existiert ein  $c_K > 0$ , sodass für alle zu  $(\hat{K}, P(\hat{K}), \Sigma(\hat{K}))$  affin äquivalente FE  $(K, P(K), \Sigma(K))$  und für alle  $v \in H^{k+1}(K)$  gilt:

$$|v - \Pi_K v|_{H^r(K)} \leq c_K \frac{h(K)^{k+1}}{\rho(K)^r} |v|_{H^{k+1}(K)}$$

falls  $0 \leq r \leq k+1$ ,  $H^{k+1}(\hat{K}) \hookrightarrow \mathcal{C}^s(\hat{K})$  und  $\mathbb{P}_k(\hat{K}) \subseteq P(\hat{K}) \subset H^r(\hat{K})$ .

TODO: genauer formulieren

**Kor.** Seien die Voraussetzungen des letzten Satzes für das finite Element  $(\hat{K}, P(\hat{K}), \Sigma(\hat{K}))$  erfüllt. Sei eine Familie von zu diesem affin äquivalenten finiten Elementen  $(K_i, P(K_i), \Sigma(K_i))_{i \in I}$  gegeben. Dann existiert eine Konstante  $\tilde{c}_K > 0$ , sodass für alle Elemente  $K_i$  der Familie mit  $h(K_i) \leq 1$  und für alle  $v \in H^{k+1}(K_i)$  gilt:

$$\begin{aligned} \|v - \Pi_K v\|_{H^r(K_i)} &\leq \tilde{c}_K \frac{h(K_i)^{k+1}}{\rho(K_i)^r} |v|_{H^{k+1}(K_i)} \\ &= \tilde{c}_K \sigma(K_i)^r h(K_i)^{k+1-r} |v|_{H^{k+1}(K_i)}. \end{aligned}$$

TODO: was ist  $r$ ?

**Bspe.** Das letzte Korollar liefert für FE vom Lagrange-Typ:

	Ordnung	$k = 1$	$k = 2$
Voraussetzungen	Regularität für $v$	$H^2(K)$	$H^3(K)$
	Beschränkung für $d$	$d \leq 3$	$d \leq 5$
	Beschränkung für $r$	$0 \leq r \leq 2$	$0 \leq r \leq 3$
Konvergenz	$\ v - \Pi v\ _{H^r(K)}$	$\mathcal{O}(h^{2-r})$	$\mathcal{O}(h^{3-r})$

$H'_0(\Omega) \subseteq U \subseteq H^1(\Omega)$ ,  $U$  ist der Lösungsraum

**Voraussetzungen.** Wir suchen die Lösung von (VGL) im Lösungsraum  $U$  mit  $H_0^1(\Omega) \subseteq U \subseteq H^1(\Omega)$ . Es gelte:

(V<sub>1</sub>)  $\bar{\Omega}$  ist ein Polyeder und  $\mathcal{T} = (T_n(\bar{\Omega}))_{n \in \mathbb{N}}$  ist eine Familie von Triangulierungen von  $\Omega$ . Es sei  $\mathcal{T}$  **regulär**, d. h. es existiert eine Konstante  $\sigma_0 > 0$ , sodass  $\sigma(K) \leq \sigma_0$  für alle  $n \in \mathbb{N}$  und  $K \in T_n(\bar{\Omega})$  und es gilt  $h_n := \max_{K \in T_n(\bar{\Omega})} h(K) \xrightarrow{n \rightarrow \infty} 0$ .

(V<sub>2</sub>) Alle finiten Elemente  $(K, P(K), \Sigma(K))$  der Familie  $\mathcal{T}$  sind affin äquivalent zu einem Referenzelement  $(\hat{K}, P(\hat{K}), \Sigma(\hat{K}))$ .

(V<sub>3</sub>)  $U_n \subset \mathcal{C}(\bar{\Omega})$  ist der FE-Funktionenraum zu  $T_n(\bar{\Omega})$ .

(V<sub>4</sub>) Für  $k+1 \geq r \geq 0$  gilt  $H^{k+1}(\hat{K}) \hookrightarrow \mathcal{C}(\hat{K})$  und  $\mathbb{P}_k(\hat{K}) \subseteq P(\hat{K}) \subseteq H^r(\hat{K})$ .

**Satz.** Seien obige Voraussetzungen erfüllt. Sei  $\Pi_n : U \rightarrow U_n$  der zu  $T_n(\bar{\Omega})$  gehörende  $U_n$ -Interpolationsoperator. Dann existiert ein  $c > 0$ , sodass für alle  $0 \leq l \leq r$ ,  $n \in \mathbb{N}$  und  $v \in H^{k+1}(\Omega) \cap U$  gilt:

$$\left( \sum_{K \in T_n(\bar{\Omega})} \|v - \Pi_n v\|_{H^l(K)}^2 \right)^{1/2} \leq c \cdot h_n^{k+1-l} \cdot |v|_{H^{k+1}(\Omega)}.$$

*Bemn.* Für  $l = 0 \leq r$  gilt  $U_n \subset L^2(\Omega)$  und

$$\|v - \Pi_n v\|_{L^2(\Omega)} = \left( \sum_{K \in T_n(\bar{\Omega})} \|v - \Pi_K v\|_{L^2(K)}^2 \right)^{1/2} \leq c \cdot h_n^{k+1} \cdot |v|_{H^{k+1}(\Omega)}.$$

Für  $l = 1 \leq r$  gilt  $U_n \subset H^1(\Omega)$  und

$$\|v - \Pi_n v\|_{H^1(\Omega)} = \left( \sum_{K \in T_n(\bar{\Omega})} \|v - \Pi_K v\|_{H^1(K)}^2 \right)^{1/2} \leq c \cdot h_n^k \cdot |v|_{H^{k+1}(\Omega)}.$$

**Satz** (Konvergenz der konformen Approximation).

Seien die Voraussetzungen (V<sub>1</sub>) – (V<sub>3</sub>) und (V<sub>4</sub>) mit  $r = 1$  erfüllt. Dann gilt für die Lösung  $u$  von (VGL) und die Lösung  $u_n$  der konformen Approximation (VGL)<sub>n</sub> mit  $\ell_n = \ell|_{U_n}$  die Abschätzung

$$\|u - u_n\|_{H^1(\Omega)} \in \mathcal{O}(h_n^k \cdot |u|_{H^{k+1}(\Omega)}) \quad \text{falls } u \in H^{k+1}(\Omega).$$

**Bsp.** Für  $k = 1$ ,  $u \in H^2(\Omega)$  ist die Konvergenzordnung bloß 1.

*Bem.* Die konforme Approximation von  $u$  hat (nach bisherigem Kenntnisstand) im Vergleich zur Interpolation von  $u$  eine um eins schlechtere Konvergenzordnung:

$$\begin{aligned} \|u - u_n\|_{L^2(\Omega)} &\leq \|u - u_n\|_{H^1(\Omega)} \in \mathcal{O}(h_n^k \cdot |u|_{H^{k+1}(\Omega)}) \\ \|u - \Pi_n u\|_{L^2(\Omega)} &\in \mathcal{O}(h_n^{k+1} \cdot |u|_{H^{k+1}(\Omega)}) \end{aligned}$$

Dieser Missstand lässt sich mit weiteren Voraussetzungen beheben:

**Lem** (Aubin-Nitsche). Seien  $U$  und  $H$  Hilberträume,  $E : U \rightarrow H$  eine stetige inj. Einbettung und  $U_n \subset U$  ein endlichdim. Teilraum. Sei  $B : U \times U \rightarrow \mathbb{R}$  eine stetige, koerzitive Bilinearform und  $\ell \in U'$ . Seien  $u \in U$ ,  $u_n \in U_n$  die Lösungen (VGL) bzw. (VGL)<sub>n</sub>. Dann gilt

$$\|E(u - u_n)\|_H \leq c_B \cdot \|u - u_n\|_U \cdot \sup_{r \in H \setminus \{0\}} \frac{\inf_{w \in U_n} \|w(r) - w\|_U}{\|r\|_H},$$

wobei  $w(r) \in U$  für  $r \in H$  die Lösung des **adjungierten Problems**

$$B(\phi, w(r)) = \langle E(\phi), r \rangle_H \quad \forall \phi \in U \quad \text{ist.}$$

**Kor.** Seien die Voraussetzungen aus dem Satz zur Konvergenz der konformen Approximation erfüllt. Zusätzlich existiere ein  $c_a > 0$ , sodass für alle  $r \in L^2(\Omega)$  die Lösung  $w(r)$  vom adjungierten Problem

$$B(\phi, w(r)) = \langle \phi, r \rangle_{L^2(\Omega)} \quad \forall \phi \in H^1(\Omega)$$

die Abschätzung  $\|w(r)\|_{H^2(\Omega)} \leq c_a \cdot \|r\|_{L^2(\Omega)}$  erfüllt. Dann gilt

$$\|u - u_n\|_{L^2(\Omega)} \in \mathcal{O}(h_n^{k+1} \cdot |u|_{H^{k+1}(\Omega)}) \quad \text{falls } u \in H^{k+1}(\Omega).$$

## Rechteckige finite Elemente

**Def.** Ein **rechteckiges finites Element vom Lagrange-Typ** der Ordnung  $k$  ist ein Tupel  $(K, P(K), \Sigma(K))$  mit

- $K = [c_1, c_1 + r_1] \times \dots \times [c_d, c_d + r_d] \subset \mathbb{R}^d$  ist ein Rechteck,
- $P(K) = \mathbb{Q}_k(K) := \{p(x) = \sum_{\alpha \in \{0, \dots, k\}^d} \lambda_\alpha x_1^{\alpha_1} \dots x_d^{\alpha_d}\} \subset \mathbb{P}_{dk}(K)$
- $\Sigma(K) = \{b : P(K) \rightarrow \mathbb{R}, \quad p \mapsto p(a) \mid a \in \mathcal{K}_k\}$ , wobei  $\mathcal{K}_k = \{(c_1 + i_1 \frac{r_1}{k}, \dots, c_d + i_d \frac{r_d}{k}) \mid i_j \in \{0, \dots, k\}, j = 1, \dots, d\}$ .

**Satz.** Jedes Polynom  $p \in \mathbb{Q}_k(K)$  ist eindeutig durch die Werte auf der Knotenmenge  $\mathcal{K}_k$  definiert.

**Def.**  $T_n(\bar{\Omega}) := \{(K_i, P(K_i), \Sigma(K_i)) \mid i = 1, \dots, N\}$  heißt Triangulierung von  $\bar{\Omega}$  mit rechteckigen FE vom Lagrange-Typ, wenn

- $\bar{\Omega} = K_1 \cup \dots \cup K_N$ ,  $\bullet \text{ int}(K_i) \cap \text{int}(K_j) = \emptyset$  für  $i \neq j$ ,
- Jede Seite von  $K_i$  ist entweder eine Teilmenge von  $\partial\Omega$  oder die Seite von einem anderen  $K_j$

**Def.** Der *Finite-Element-Raum* zur Triangulierung  $T_n(\bar{\Omega})$  ist

$$U_n := \{v \in \mathcal{C}(\bar{\Omega}) \mid v|_{K_i} \in \mathbb{Q}_k(K_i), i = 1, \dots, N\}.$$

**Lem.**  $\bullet U_n \subset H^1(\Omega)$

- Eine Basis von  $U_n$  ist durch Polynome  $p_j \in U_n$  mit  $p_i(a_j) = \delta_{ij}$  für alle  $ia_j \in \mathcal{K}_k$  gegeben.

## Simpliziale Elemente vom Hermite-Typ

**Def.** Ein **simpliziales finites Element vom Hermite-Typ** ist ein Tupel  $(K, P(K), \Sigma(K))$  mit

- $K$  ist ein  $d$ -Simplex mit Eckpunkten  $a_1, \dots, a_d$
- $P(K) := \mathbb{P}_3(K)$
- $\Sigma(K) := ???$   
 $\{p(a_i) \mid i = 1, \dots, d+1\} \cup \{p(a_{ijl}) \mid 1 \leq i < j < l \leq d+1\}$   
 $a_{ijl} := \frac{1}{3}(a_i + a_j + a_l), \{\mathcal{D}p(a_i)(a_j - a_i) \mid 1 \leq i \neq j \leq d+1\}$

Der zugehörige Finite-Elemente-Raum zu einer Triangulierung  $T(\bar{\Omega})$  mit simpl. finiten Elementen vom Hermite-Typ ist

$$U_n := \{v \in \mathcal{C}(\bar{\Omega}) \mid v|_K \in \mathbb{P}_3(K) \forall K \in T(\bar{\Omega})\}$$

*Bemn.*  $\bullet U_n \subset \mathcal{C}^1(\bar{\Omega})$  (nur) für  $d = 1$

- $U_n \subset \mathcal{C}^1(\bar{\Omega})$  für  $d = 2$  erreicht man mit folgenden Elementen:

- Argyris-Dreieck:*  $(K, P(K), \Sigma(K))$  wobei  $K$  ein Simplex mit Ecken  $a_1, a_2, a_3$ ,  $P(K) := \mathbb{P}_5(K)$  und  $\Sigma(K) := \{p(a_i) \mid 1 \leq i \leq 3\} \cup \{\partial_x p(a_i) \mid 1 \leq i \leq 3\} \cup \{\partial_y p(a_i) \mid 1 \leq i \leq 3\} \cup \{\partial_{xy} p(a_i) \mid 1 \leq i \leq 3\} \cup \{\partial_{xx} p(a_i) \mid 1 \leq i \leq 3\} \cup \{\partial_{xy} p(a_i) \mid 1 \leq i \leq 3\} \cup \{\mathcal{D}p(\frac{a_i + a_j}{2}) \cdot \nu_{ij} \mid 1 \leq i < j \leq 3\}$  (wobei  $\nu_{ij} :=$  äußerer Normalenvektor an  $\frac{a_i + a_j}{2}$ ) ist.

- Bell-Dreieck:*  $(K, P(K), \Sigma(K))$  mit  $P(K) := \{p \in \mathbb{P}_5 \mid \frac{\partial p}{\partial \nu_{ij}} \in \mathbb{P}_3(K') \text{ für alle Kanten } K' \text{ von } K\}$ ,  $\Sigma(K) := \{p(a_i) \mid 1 \leq i \leq 3\} \cup \{\partial_x p(a_i) \mid 1 \leq i \leq 3\} \cup \{\partial_y p(a_i) \mid 1 \leq i \leq 3\} \cup \{\partial_{xx} p(a_i) \mid 1 \leq i \leq 3\} \cup \{\partial_{xy} p(a_i) \mid 1 \leq i \leq 3\} \cup \{\partial_{yy} p(a_i) \mid 1 \leq i \leq 3\}$



### Nichtkonforme finite Elemente

Bisher war  $B_n = B_{U_n \times U_n}$ . Manchmal ist es schwierig,  $B$  exakt auszuwerten und muss daher angenähert werden, etwa wenn  $B$  durch Integration definiert ist. Dann hilft folgende Verallgemeinerung des Lemmas von Céa:

**Satz** (1. Lemma von **Strang**). Sei  $B_U \times U \rightarrow \mathbb{R}$  eine beschränkte, koerzitive Bilinearform,  $U_n \subset U$ ,  $B_n : U_n \times U_n \rightarrow \mathbb{R}$  eine beschränkte, *gleichmäßig koerzitive* Bilinearform, d. h.

$$\exists \alpha > 0 : \forall n : B_n(u_n, u_n) \geq \alpha \|u_n\|_U^2 \quad \text{für alle } u \in U_n.$$

Dann existiert eine Konstante  $c > 0$ , sodass

$$\|u - u_n\|_U \leq c \left( \inf_{v_n \in U_n} \Phi(v_n) + \|\ell - \ell_n\|_{U'} \right)$$
$$\Phi(v_n) := \|u - v_n\|_U + \sup_{w_n \in U_n, w_n \neq 0} \frac{|B(v_n, w_n) - B_n(v_n, w_n)|}{\|w_n\|_U}$$

**Def.** Das **Fehlerfunktional der Quadraturformel** mit Stützpunkten  $x_i^K$  und Gewichten  $w_i^K$  ist

$$E_K(\psi) := \int_K \psi(x) \, dx - \sum_{i=1}^q w_i^K \psi(x_i^K).$$

**Satz** (Konvergenzsatz für das Modellproblem).

Seien die Voraussetzungen (V<sub>1</sub>) – (V<sub>3</sub>) erfüllt,  $P(\hat{K}) = \mathbb{P}_k(\hat{K})$  mit  $k > d/2$  (damit  $H^{k+1}(\Omega) \hookrightarrow \mathcal{C}^s(\hat{K})$ ). Sei die Quadraturformel so gewählt, dass  $E_{\hat{K}}(p) = 0$  für alle  $p \in \mathbb{P}_{2k-2}$ . Für die Lösung  $u \in H_0^1(\Omega)$  von (VGL) mit  $B(u, \varphi) := \int_{\Omega} \mathcal{D}u \cdot \mathcal{D}\varphi \, dx$ ,  $\ell(\varphi) := \int_{\Omega} f \varphi \, dx$  mit  $f \in H^k(\Omega)$  gelte  $u \in H^{k+1}(\Omega)$ . Dann gilt:

- Es gibt eine Konstante  $C_1 > 0$ , sodass  $\forall p, q \in \mathbb{P}_k(K)$ :

$$E_K(\mathcal{D}p \cdot \mathcal{D}q) \leq C_1 \cdot h^k(K) \cdot \|p\|_{H^k(K)} \cdot \|q\|_{H^1(K)}$$

- Es gibt eine Konstante  $C_2 > 0$ , sodass  $\forall f \in H^k(K)$ ,  $\varphi \in H^1(K)$ :

$$E_K(f\varphi) \leq C_2 \cdot h^k(K) \cdot \|f\|_{H^k(K)} \cdot \|\varphi\|_{H^1(K)}$$

- Für die Lsg  $u_n \in U_n$  von (VGL)<sub>n</sub> mit der nicht-konformen Approx.  $B_n(u, \varphi)$  und  $\ell_n(\varphi)$  mittels der Quadraturformel gilt:

$$\|u - u_n\|_{H^1(\Omega)} \leq C \cdot (\|u\|_{H^{k+1}(\Omega)} + \|f\|_{H^k(\Omega)})$$

**Def.** Sei  $\Omega \subset \mathbb{R}^d$  ein Gebiet mit glattem Rand. Dann heißt  $(\Omega_n, T(\overline{\Omega}_n))$  **zulässige Gebietsapproximation**, falls  $T_n(\overline{\Omega}_n)$  eine zulässige Triangulierung von  $\Omega_n$  ist und

$$\partial\Omega_n \cap \{ \text{Eckpunkte der Simplexes in } T(\overline{\Omega}_n) \} \subset \partial\Omega.$$

Wir betrachten nun

**Problem.**

$$(VGL_n) \quad \begin{cases} \overbrace{\int_{\Omega_n} \mathcal{D}u_n \cdot \mathcal{D}\varphi_n \, dx}^{B_n(u_n, \varphi_n) :=} = \overbrace{\int_{\Omega_n} \tilde{f} \varphi_n \, dx}^{\ell_n(\varphi_n) :=} & \forall \varphi \in U_n \\ u_n \in U_n \end{cases}$$

wobei  $\tilde{f} \in L^2(\Omega_n \cup \Omega)$  mit  $\tilde{f}|_{\Omega} = f$

**TODO:** etwas ausgelassen; wichtig?

**Def.** Sei  $(\hat{K}, \hat{P}, \hat{\Sigma})$  ein gerades (simpliciales oder rechteckiges) FE vom Lagrange-Typ. Ein finites Element  $(K, P, \Sigma)$  vom Langrange-Typ (k) heißt **isometrisch äquivalent** zu  $(\hat{K}, \hat{P}, \hat{\Sigma})$ , falls eine umkehrbare Abbildung

$$F : \hat{K} \rightarrow K, \quad \hat{x} \mapsto (F_1(\hat{x}), \dots, F_d(\hat{x}))$$

mit  $F_1, \dots, F_d \in \hat{P}$  existiert sodass

- $K = F(\hat{K})$
- $P = \{p = \hat{p} \circ F^{-1} \mid \hat{p} \in \hat{P}\}$
- $\Sigma = \{p \mapsto p(a_i) \mid \text{für Punkte } a_i := F(\hat{a}_i), \hat{a}_i \in \mathcal{K}_k\}$

*Bem.* Die zu  $(K, P, \Sigma)$  isoparametrische Abbildung  $F$  ist für isoparametrische FE vom Langrange-Typ (k) eindeutig durch die Werte an den Knoten  $\hat{a}_i$  definiert.

### Finite Elemente für parabolische Probleme

**Problem.** Wärmeleitungsgleichung:

$$\begin{cases} u_t(x, t) - \Delta_x u(x, t) & = & f(x, t) & \text{in } \Omega \times (0, T) \\ u(x, 0) & = & u_0(x) & \text{für } x \in \Omega \\ u(x, t) = 0 & \text{für } x \in \partial\Omega, t \in [0, T] \end{cases}$$

$$\int_{\Omega} u_t(x, t) \varphi(x) \, dx = \int_{\Omega} \Delta u(x, t) \varphi(x) \, dx + \int_{\Omega} f(x, t) \varphi(x) \, dx - \int_{\Omega} \mathcal{D}u(x, t) \cdot \mathcal{D}\varphi(x) \, dx$$

Anders geschrieben:

$$\frac{d}{dt} \langle u(t, -), \varphi \rangle + B(u, \varphi) - \langle f(t, -), \varphi \rangle$$

mit  $\varphi \in H_0^1(\Omega)$ ,  $u \in \mathcal{C}^1(0, T; H_0^1(\Omega))$ . Anfangsbedingung:  $u(0) = u_0$ .

Idee: 1. Diskretisierung im Ort, 2. Diskretisierung in der Zeit

- Sei  $U_n \subset U = H_0^1(\Omega)$

$$\frac{d}{dt} \langle u_n(t), \varphi_n \rangle + B_n(u_n, \varphi_n) = \langle f(t, -), \varphi \rangle \quad \forall \varphi \in U_n$$

Sei  $\varphi_1, \dots, \varphi_{d_n}$  eine Basis von  $U_n$ ,  $u_n(t) = \sum_{j=1}^{d_n} \gamma_j(t) \varphi_j(x)$ ,

$$u_n^0 = \sum \gamma_j^0 \varphi_j,$$

$$(GGL) \quad \begin{cases} \sum_{j=1}^{d_n} \frac{d}{dt} \gamma_j(t) \langle \varphi_j, \varphi_i \rangle + \sum_{j=1}^{d_n} \gamma_j(t) B_n(\varphi_j, \varphi_i) & = & \langle f(t, -), \varphi_i \rangle \\ \gamma_j(0) & = & \gamma_j^0, j = 1, \dots \end{cases}$$

$$\begin{cases} M \frac{d}{dt} \gamma(t) + B \gamma(t) & = & g(t) \\ \gamma(0) & = & \gamma^0 \end{cases}$$

mit  $M_{ij} = \int_{\Omega} \varphi_j(x) \varphi_i(x) \, dx$  (Massenmatrix),

$B_{ij} = \int_{\Omega} \mathcal{D}\varphi_j \cdot \mathcal{D}\varphi_i \, dx$ ,  $g_j(t) = \int_{\Omega} f(t, x) \varphi_j(x) \, dx$ . Dabei ist

$\gamma \in \mathbb{R}^{d_n}$ . Durch Modellreduktion erhält man eine Gleichung

$$\begin{cases} \tilde{M} \frac{d}{dt} \tilde{\gamma}(t) + \tilde{B} \tilde{\gamma}(t) & = & \tilde{g}(t) \\ \tilde{\gamma}(0) & = & \tilde{\gamma}^0 \end{cases}$$

sodass  $\gamma \in \mathbb{R}^r$ ,  $r < d_n$  und  $\gamma \approx V \tilde{\gamma}$ .

Idee: 1. Diskretisierung in der Zeit, 2. Diskretisierung im Ort

$$\tau = T/q, \quad t_i = i \cdot \tau$$

$$\text{Sei } \psi_i(t) := \begin{cases} t - t_{i-1}/\tau & t \in [t_{i-1}, t_i], \\ t_i - t/\tau & t \in [t_i, t_{i+1}], \\ 0 & \text{sonst} \end{cases}$$

$$u(x, t) \approx u_{\tau}(x, t) = \sum_{j=1}^q \psi_j(t) c_j(x) \quad \text{Somit } u_{\tau}(x, t_i) = c_i(x) \quad \text{Implizites Euler-Verfahren:}$$

$$\frac{d}{dt} u_{\tau}(x, t_{i+1}) \approx \frac{u_{\tau}(x, t_{i+1}) - u_{\tau}(x, t_i)}{\tau} = \frac{c_{i+1}(x) - c_i(x)}{\tau},$$
$$\left\langle \frac{c_{i+1} - c_i}{\tau}, \varphi \right\rangle + B(c_{i+1}, \varphi) = \langle f(t_{i+1}), \varphi \rangle$$
$$\underbrace{\int_{\Omega} c_{i+1} \varphi \, dx + \tau \int_{\Omega} \mathcal{D}c_{i+1} \cdot \mathcal{D}\varphi \, dx}_{\hat{B}(c_{i+1}, \varphi)} = \underbrace{\tau \int_{\Omega} f(t_{i+1}, x) \varphi(x) \, dx + \int_{\Omega} c_i \varphi \, dx}_{\hat{\ell}(\varphi)}$$

2. Schritt: Löse die Variationsgleichung mittels der FEM