Parallele Lösung großer Gleichungssysteme

Peter Bastian
Universität Heidelberg
Interdisziplinäres Zentrum für Wissenschaftliches Rechnen
Im Neuenheimer Feld 368, D-69120 Heidelberg
mail: Peter.Bastian@iwr.uni-heidelberg.de

6. Juli 2009

Inhaltsverzeichnis

Vorwort

Die Lösung linearer Gleichungssysteme $Ax = b$, $A \in \mathbb{R}^{N \times N}$, ist an sich nicht schwierig, ein allgemeines Lösungsverfahren (die Gauß-Elimination) findet sich oft auf den ersten Seiten eines Numerikbuches, siehe z. B. (Deuflhard und Hohmann 1993). Der Rechenaufwand für das allgemeine Verfahren steigt jedoch mit wachsendem N sehr stark an. Für N in der Größenordnung $10^6 \ldots 10^9$ ist das Verfahren vollkommen ungeeignet.

Matrizen dieser Größe treten etwa bei der Diskretisierung partieller Differentialgleichungen auf. Die Größe von N steht dabei in direktem Zusammenhang mit der Genauigkeit der numerischen Approximation der Lösung der Differentialgleichung.

1 Modellproblem, Variationsformulierung

01.11.09
1

VL 1 Variationsformulierung

Literatur: Braess, Finite Elemente, Springer
Hackbusch: Theorie u. Numerik
Brenner/Scott: The math. Theory of FEM, Springer
Ciarlet: The FEM for Ell. Pr., SIAM Classics.

1. Das Modellproblem

Wir sind interessiert an der Lösung der Gleichung

hier $+k_0(x)\cdot u$ einfügen? → Praxis.

$$
\begin{aligned}
-\nabla\cdot\{K(x)\nabla u\} &= f \quad \text{im } \Omega, \qquad (1)\\
u &= g \quad \text{auf } \Gamma_D \subseteq \partial\Omega, \quad \text{Dirichlet RB}\\
-(K(x)\nabla u)\cdot\nu &= j \quad \text{auf } \Gamma_N = \partial\Omega\setminus\Gamma_D. \quad \text{Neumann RB.}
\end{aligned}
$$

- $\Omega \subset \mathbb{R}^d$, $d = 1,2,3$ ist ein Gebiet (offen, zusammenhängend) mit genügend glattem Rand.

$K(x)$ ist für jedes $x\in\Omega$ eine symmetrisch positiv definite $d\times d$ Matrix.

$\text{meas}(\Gamma_D) \neq 0$ ist notwendig für Eindeutigkeit der Lösung ohne weitere Bedingung.

$u \in C^2(\Omega)\cap C^1(\Omega\cup\Gamma_N)\cap C^0(\overline{\Omega})$ heißt klassische Lösung von (1).

- Existenz klassischer Lösung zu beweisen ist aufwändig (siehe Hackbusch) und erfordert unpraktische Bedingung an f.

Eindeutigkeit ist relativ simpel (Maximumprinzip).
Einen eleganteren Zugang bietet die Variationsformulierung.

2. Der eindimensionale Fall

Es sei nun $\Omega = (a,b) \subset \mathbb{R}^1$.

Wir betrachten das Problem

$$-\frac{d}{dx}\left(k(x)\frac{du}{dx}\right) = f(x) \quad \text{in } \Omega=(a,b) \tag{2}$$
$$u(x) = g(x) \quad x \in \{a,b\}$$

(ohne Neumann-RB).

Dies ist eine gewöhnliche Differentialgleichung zweiter Ordnung.
Allerdings kein Anfangswertproblem!

Homogene Dirichlet Randbedingung

Ansatz: $u = w + u'$ wobei $w \in C^2(\Omega) \cap C^0(\bar{\Omega})$ und
$w(x) = g(x)$ für $x \in \{a,b\}$, d.h.
$u'(x) = 0$ für $x \in \{a,b\}$.

Linearität der Differentiation liefert:

$$-\frac{d}{dx}\left(k(x)\frac{d}{dx}(w+u')\right) = -\frac{d}{dx}\left(k(x)\frac{dw}{dx}\right) - \frac{d}{dx}\left(k(x)\frac{du'}{dx}\right) = f \quad \text{in } \Omega$$

Für ein gegebenes w ist (2) also äquivalent zu

$$-\frac{d}{dx}\left(k(x)\frac{du'}{dx}\right) = \underbrace{f + \frac{d}{dx}\left(k(x)\frac{dw}{dx}\right)}_{f'} \quad \text{in } \Omega,$$
$$u'(x) = 0 \quad x \in \{a,b\}.$$

Es genügt daher im Prinzip homogene Dirichlet-Randbedingungen zu betrachten (liegt wesentlich an der Linearität).

Variationsformulierung

01.IV.09
3

Sei $v(x)$ eine genügend oft differenzierbare Funktion (z.B. $v \in C^1(\Omega) \cap C^2(\Omega)$) mit $v(a) = v(b) = 0$.

Mit partieller Integration gilt dann:

$$\underbrace{\int_a^b f(x)\, v(x)\, dx}_{=: (f,v)_\Omega} = \int_a^b -\frac{d}{dx}\Big(\underbrace{k(x)\frac{du}{dx}}_{w}\Big)\, v(x)\, dx$$

„L_2-Skalarprodukt"

$$= \underbrace{\int_a^b \underbrace{k(x)\frac{du}{dx}}_{w}\frac{dv}{dx}\, dx}_{=: a(u,v)} + \underbrace{\Big[\underbrace{-k(x)\frac{du}{dx}}_{w} v(x)\Big]_a^b}_{=0 \text{ da } v(a)=v(b)=0}$$

Für $f \in C^0(\bar{\Omega})$, $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$ Lösung von (2) gilt also

$$a(u,v) = (f,v)_\Omega \quad \text{für alle } v \in C^1(\Omega) \cap C^0(\Omega),\ v(a)=v(b)=0.$$

Nun kann man die Argumentation umdrehen.

Sei V ein geeigneter „Raum von Funktionen" und betrachte das Problem

Finde $u \in V$ so dass $a(u,v) = (f,v)_\Omega$ für alle $v \in V$. (3)

(3) nennt man „Variationsformulierung", da die Funktion v beliebig variieren darf.

Es stellen sich folgende Fragen:

- welche Funktionenräume V sind „geeignet"?
- Unter welchen Umständen hat (3) eine (eindeutige) Lösung.
- Wie kann man das praktisch ausnutzen?

Funktionalräume → Funktionalanalysis.

Betrachte zunächst die rechte Seite

$$\ell(v) := (f, v)_\Omega = \int_a^b f(x)\, v(x)\, dx . \qquad (4)$$

$\ell(v)$ ist wohldefiniert für alle „quadratintegrierbaren" Funktionen

$$V = L_2(\Omega) = \left\{ v : \Omega \to \mathbb{R} \;\middle|\; \int_\Omega |v(x)|^2 dx < \infty \right\}$$

- $(\cdot,\cdot)_\Omega$ definiert ein Skalarprodukt auf $L_2(\Omega)$
- $\|v\|_\Omega = \sqrt{(v,v)_\Omega} = \left(\int_\Omega |v(x)|^2 dx \right)^{1/2}$ ist eine Norm → normierter Raum
- $L_2(\Omega)$ ist vollständig (wie $\mathbb{Q} \to \mathbb{R}$) (jede Cauchy-Folge (erfordert Norm!) konvergiert) → wichtig. $C^2[\Omega] \cap C^0[\bar\Omega]$ ist nicht vollständig bezüglich $\|\cdot\|_2$)
- $L_2(\Omega)$ ist ein Hilbert-Raum. (der $\mathbb{R}^N$ der Funktionenräume).
- Das Integral in (4) ist im Lebesgue-Sinne zu verstehen (kein Riemann-Integral). Änderung der Funktion auf einer „Nullmenge" führt auf dieselbe Äquivalenzklasse
- Eine Forderung $v(a) = v(b) = 0$ ist daher nicht möglich.
- Für festes $f \in L_2(\Omega)$ ist $\ell : L_2(\Omega) \to \mathbb{R}$ ein stetiges, lineares Funktional.

01.IV.09
5

Nun betrachten wir die linke Seite

$$a(u,v) = \int_a^b k(x) \frac{du}{dx} \frac{dv}{dx} \, dx . \qquad (5)$$

- Nun ist über Ableitungen zu integrieren. Im allg. sind Ableitg von L_2-Funktionen nicht quadratintegrierbar. Wir brauchen also andere Räume
- Sobolevraum $H^1(\Omega) := \left\{ v \in L_2(\Omega) \;\middle|\; \int_\Omega |v(x)|^2 + \left|\frac{dv}{dx}(x)\right|^2 dx < \infty \right\}$
  offensichtlich $H^1(\Omega) \subset L_2(\Omega)$
  ↑ „schwache Ableitung"
  ϕ ist Abl. von $v \Leftrightarrow \int_\Omega \phi u \, dx = -\int_\Omega v \frac{du}{dx} dx \quad \forall u \in C_0^\infty(\Omega)$
- $(u,v)_{1,\Omega} := \int_\Omega u(x)v(x) + \frac{du}{dx}(x)\frac{dv}{dx}(x) \, dx$ ist ein Skalarprodukt auf $H^1(\Omega)$
- $\|v\|_{1,\Omega} = \sqrt{(v,v)_{1,\Omega}}$ ist eine Norm auf $H^1(\Omega)$
- 08.V.09: später brauchen wir die Seminorm $|v|_{1,\Omega}$
- $H^1(\Omega)$ ist vollständig bezüglich $\|.\|_{1,\Omega}$
- $H^1(\Omega)$ ist ein Hilbertraum.

bis hier.

- Es ist möglich in einem geeigneten Sinne Randbedingg vorzuschreiben

$$H_0^1(\Omega) = \left\{ v \in H^1(\Omega) \;\middle|\; „v = 0" \text{ auf } \partial\Omega \right\}$$

(„Vervollständigung" von C_0^∞-Funktionen ?)

- $a : H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ nennt man eine Bilinearform.
- a ist hier symmetrisch, dies ist eine Folge der Symmetrie des Differentialoperators.
- Es stellt sich heraus, dass der Sobolevraum $H_0^1(\Omega)$ genau der richtige Raum ist um das Problem (3) zu lösen.

In einem sehr viel allgemeineren Kontext kann man folgende Aussage zeigen:

Lax-Milgram Lemma. Es sei

(i) V ein Hilbertraum (z.B. der $H_0^1(\Omega)$)

(ii) $a : V \times V \to \mathbb{R}$ eine stetige, V-elliptische Bilinearform, d.h.

$\exists C > 0, \forall u, v \in V : \quad a(u,v) \leq C \|u\|_V \|v\|_V$ (Stetigkeit)

$\exists \varepsilon > 0, \forall v \in V : \quad a(v,v) \geq \varepsilon \|v\|_V^2$ (V-Elliptizität)

(iii) $\ell : V \to \mathbb{R}$ eine stetige Linearform

$\exists C', \forall v \in V : \ell(v) \leq C' \|v\|$

Erfordert u.A. Bedingungen an $K(x)$.

dann hat das abstrakte Problem

$$a(u,v) = \ell(v) \quad \forall v \in V$$

genau eine Lösung $u \in V$.

Beweis: Ciarlet, Thm 1.1.3.

Bem 1: Die Symmetrie von a ist hierzu nicht notwendig.

Bem 2: Man kann die Aussage auf konvexe Unterräume verallgemeinern:

Sei $U \subset V$ ein abgeschlossener, konvexer Unterraum von V (d.h. Grenzwerte von Folgen aus U sind wieder in U, $u, u' \in U \to \alpha u + (1-\alpha) u' \in U \quad \forall \alpha \in [0,1]$).

Dann hat das Problem $a(u,v) = \ell(v) \quad \forall v \in U$

genau eine Lösung $u \in U$.

Bem 3: $U \subset V$ kann auch endlichdimensional sein.

01.IV.09
7

Neumann-Randbedingungen

Wir betrachten nun das allgemeinere Problem

$$-\frac{d}{dx}\left(k(x)\frac{du}{dx}\right) = f(x) \quad \text{in } \Omega = (a,b)$$

$$u(a) = 0$$

$$-k(b)\frac{du}{dx}(b) = j$$

Mit partieller Integration erhalten wir für $v \in C^1(\Omega) \cap C^0(\Omega \cup \{a\})$, $v(a) = 0$:

$$\int_a^b f(x)v(x)\,dx = -\int_a^b \frac{d}{dx}\left(k(x)\frac{du}{dx}\right)v(x)\,dx$$

$$= \int_a^b k(x)\frac{du}{dx}(x)\frac{dv}{dx}(x)\,dx \underbrace{- k(b)\frac{du}{dx}(b)}_{=j} v(b) + k(a)\frac{du}{dx}(a)\underbrace{v(a)}_{=0}$$

Das Variationsproblem lautet damit wie folgt.

$$V = \left\{ v \in H^1(\Omega) \;\middle|\; v(a) = 0 \right\}$$

Finde $u \in V$ so dass

$$\underbrace{\int_a^b k(x)\frac{du}{dx}(x)\frac{dv}{dx}(x)\,dx}_{= a(u,v)} = \underbrace{\int_a^b f(x)v(x)\,dx - j\,v(b)}_{l(v)}$$

- Die Dirichlet-Randbedingung nennt man in der Variationsformulierung „essentielle" Randbedingung, da sie explizit in den Funktionenraum eingebaut werden muss.
- Die Neumann-Randbedingung nennt man in der Variationsformulierung „natürliche" Randbedingung, da sie sich automatisch über die partielle Integration ergibt (Vorteil gegenüber Finite Differenzen Methode)

2 Finite Elemente in einer Raumdimension

VL 2 Finite Elemente in 1D

03.11.09
1

Gitterfunktionen

Ziel: Konstruktion eines endlichdimensionalen Teilraums $V_h^1 \subset H_0^1(\Omega)$

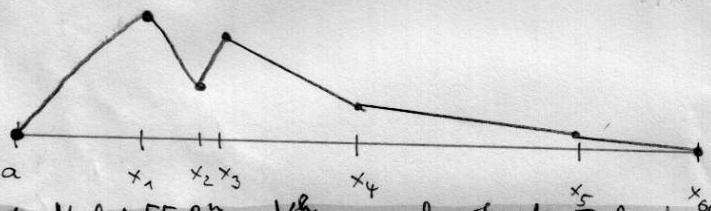
Zentral ist hierfür der folgende

Satz 2.1 Ω sei ein beschränktes Gebiet. Eine stückweise C^∞-Funktion $v: \overline{\Omega} \to \mathbb{R}$ gehört genau dann zu $H^1(\Omega)$, wenn sie stetig ist.

Bew: Braess, Satz 5.2 (dort für beliebiges k).

Minimale Konstruktion:

Unterteile $\Omega = (a,b)$ in N Teilintervalle (x_{i-1}, x_i) $0 < i \leq N$:



$N = 6$

nächstes Mal: FE-Räume V^h nenne damit unten Indexe hinkommt

Polynomgrad brauchen wir eigentlich nicht.

$$V_h^1 = \{ v \in C^0(\overline{\Omega}) \mid v|_{[x_{i-1}, x_i]} \text{ ist Polynom vom Grad 1 und } v(a) = v(b) = 0 \}$$

$0 < i \leq N$

$h := \max_{0 < i \leq N} x_i - x_{i-1}$ „Gitterweite", später: $h \to 0$

$v \in V_h^1$ ist eindeutig festgelegt durch die Werte an den Stellen x_i, $0 < i < N$, die Dimension von V_h^1 ist also

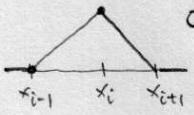
$$\dim V_h^1 = N - 1.$$

V_h^1 ist also ein $N-1$-dimensionaler Vektorraum.

Man nennt V_h^1 einen konformen Finite-Elemente-Raum, da $V_h^1 \subset H_0^1(\Omega)$

Knotenbasis einer Basis:

Es gibt natürlich unendlich viele Basen. Besonders vorteilhaft ist die sog. Knotenbasis oder Lagrangebasis.



$$\phi_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}} & x \in [x_{i-1}, x_i] \\ \frac{x_{i+1} - x}{x_{i+1} - x_i} & x \in (x_i, x_{i+1}] \\ 0 & \text{sonst} \end{cases} \qquad 0 < i < N$$

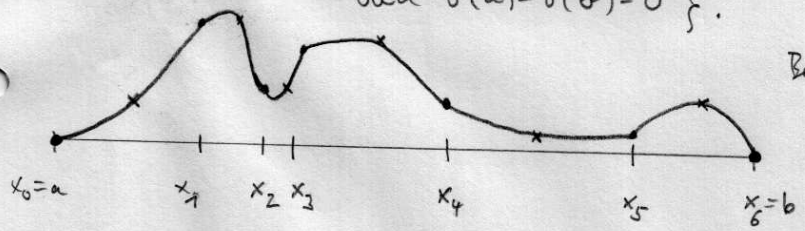
Es gilt offensichtlich

- $\phi_i(x_j) = \begin{cases} 1 & i=j \\ 0 & \text{sonst} \end{cases}$ (wie bei Lagrangepolynom)

- $\phi_i(x) > 0$ für $x \in (x_{i-1}, x_{i+1})$

Höhere Ordnung.

$$V_h^k = \{ v \in C^0(\bar{\Omega}) \mid v|_{[x_{i-1}, x_i]} \text{ ist Polynom vom Grad } k,\ 0 < i \leq N \text{ und } v(a) = v(b) = 0 \}.$$



Beispiel für $k=2$.

$\dim V_h^2 = N-1 + N = 2N-1$

$\dim V_h^k = N-1 + (k-1)N = kN-1$

Man kann wieder eine Knotenbasis aufstellen.

Basisdarstellung

Gegeben eine Basis Φ_h^k für V_h^k dann läßt sich $u \in V_h^k$ darstellen

$u(x) = \sum_{i=1}^{n} z_i \phi_i(x)$, $n = \dim V_h^k$ $\quad FE: \mathbb{R}^n \to V_h^k$ heißt FE-Isomorphismus.

Gleichungssystem

Das Variationsproblem

Finde $u \in H_0^1(\Omega)$ so dass $a(u,v) = \ell(v) \quad \forall v \in H_0^1(\Omega)$

lösen wir nun im endlichdimensionalen Teilraum $V_h^k \subset H_0^1(\Omega)$:

Finde $u_h \in V_h^k$ so dass $a(u_h, v) = \ell(v) \quad \forall v \in V_h^k$.

Zur Lösung nutzen wir nun die Basisdarstellung:

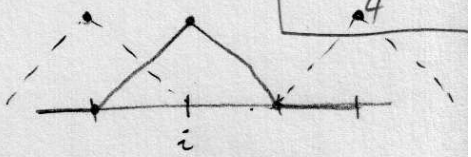
$$
\begin{aligned}
& a(u_h, v) = \ell(v) && \forall v \in V_h^k \\
\Leftrightarrow\ & a(u_h, \phi_i) = \ell(\phi_i) && 0 < i \leq n \quad \text{(Testen auf Basis genügt)} \\
\Leftrightarrow\ & a\Big(\sum_{j=1}^{n} z_j \phi_j, \phi_i\Big) = \ell(\phi_i) && 0 < i \leq n \quad \text{(Basisdarstellung von } u_h\text{)} \\
\Leftrightarrow\ & \sum_{j=1}^{n} z_j\, a(\phi_j, \phi_i) = \ell(\phi_i) && 0 < i \leq n \quad \text{(Linearität)} \\
\Leftrightarrow\ & Az = b
\end{aligned}
$$

mit $A_{ij} = a(\phi_j, \phi_i)$ und $b = \ell(\phi_i)$

A ist symmetrisch und positiv definit

- $A_{ij} = a(\phi_j, \phi_i) = a(\phi_i, \phi_j) = A_{ji}$
- $$
\begin{aligned}
x^T A x &= \sum_{i=1}^{n} x_i \Big(\sum_{j=1}^{n} A_{ij} x_j \Big) = \sum_{i=1}^{n} x_i \Big(\sum_{j=1}^{n} a(\phi_j, \phi_i) x_j \Big) \\
&= \sum_{i=1}^{n} x_i\, a\Big(\sum_{j=1}^{n} x_j \phi_j, \phi_i\Big) = a\Big(\underbrace{\sum_{j=1}^{n} x_j \phi_j}_{=: v_h}, \underbrace{\sum_{i=1}^{n} x_i \phi_i}_{=: v_h}\Big) \\
&= a(v_h, v_h) \geq \varepsilon \|v_h\|_{1,\Omega} > 0 \text{ für } v_h \neq 0 \text{ da } \|\cdot\|_{1,\Omega} \text{ eine Norm}
\end{aligned}
$$
- Somit ist A auch regulär (s.p.d. $\Leftrightarrow$ alle EW reell und positiv).

Besetztheit von A



$$A_{ij} = a(\phi_j, \phi_i) = \int_\Omega k(x) \frac{d\phi_j}{dx} \frac{d\phi_i}{dx} dx$$

$$= \begin{cases} 0 & \text{falls} \quad |i-j| > 1 \\ \neq 0 & \text{falls} \quad j \in \{i-1, i, i+1\} \end{cases}$$

$\Rightarrow$ A ist eine Tridiagonalmatrix.

← bis hier.

Fehlerabschätzungen

Wir benötigen noch weitere Sobolevräume:

$$H^k(\Omega) = \left\{ v \in L_2(\Omega) \;\middle|\; \int_\Omega \sum_{i=0}^{k} \left| \frac{d^k v}{dx^k}(x) \right|^2 dx < \infty \right\}$$

also $H^0(\Omega) = L_2(\Omega) \supset H^1(\Omega) \supset H^2(\Omega) \supset H^3(\Omega) \ldots$

Man zeigt dann

$$\| u - u_h \|_{1,\Omega} \leq C h^k \| u \|_{k+1,\Omega}$$

u: Lösung des Var. Problems in $H^1(\Omega)$; u_h: FE-Lösung in V_h^k; $\| u \|_{k+1,\Omega}$: Norm von H^{k+1}

Bem: h^k Konvergenz erfordert $u \in H^{k+1}(\Omega)$

Man zeigt auch:

$$\| u - u_h \|_{0,\Omega} \leq C h^{k+1} \| u \|_{k+1,\Omega}$$

3 Finite Elemente in mehreren Raumdimensionen

06.IV.09
1

3 Finite Elemente in mehreren Raumdimensionen

Unser Modellproblem:

$$-\nabla\cdot\{K\nabla u\} = f \quad \text{in } \Omega$$
$$u = g \quad \text{auf } \Gamma_D$$
$$-(K\nabla u)\cdot\nu = j \quad \text{auf } \Gamma_N$$

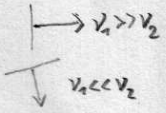
3.1 Schwache Formulierung

Grundlage ist die folgende Green'sche Formel. Für $u, v \in C^1(\Omega)\cap C(\bar\Omega)$ und $\Omega \subset \mathbb{R}^d$ ein Gebiet gilt

$$\int_\Omega \partial_i u\, v\, dx = -\int_\Omega u\, \partial_i v\, dx + \int_{\partial\Omega} u\, v\, \nu_i \underbrace{ds}_{\text{Bogenlänge}} \qquad i = 1,\dots,d$$

$\partial_i u = \frac{\partial u}{\partial x_i}$ ist die Ableitung nach der i-ten Variablen

ν_i: i-te Komponente der äußeren Einheitsnormale.



Damit rechnen wir: $v \in C^1(\bar\Omega)$

$$\int_\Omega f v\, dx = \int_\Omega -\nabla\cdot\{K\nabla u\}\, v\, dx = \int_\Omega \sum_{i=1}^d \partial_i \Big\{-\sum_{j=1}^d K_{ij}(x)\,\partial_j u\Big\}\, v\, dx$$

$$= \sum_{i=1}^d \int_\Omega \partial_i \Big\{\underbrace{-\sum_{j=1}^d K_{ij}(x)\,\partial_j u}_{=w}\Big\}\, v\, dx$$

$$= \sum_{i=1}^d \Bigg[\int_\Omega \underbrace{\sum_{j=1}^d K_{ij}(x)\partial_j u}_{(K\nabla u)_i}\, \partial_i v\, dx + \int_{\partial\Omega} \underbrace{-\sum_{j=1}^d K_{ij}(x)\partial_j u}_{(K\nabla u)_i}\, v\, \nu_i\, ds \Bigg]$$

$$= \int_\Omega (K\nabla u)\cdot\nabla v\, dx + \int_{\partial\Omega} \underbrace{-(K\nabla u)\nu}_{=j \text{ auf } \Gamma_N}\, v\, ds \qquad \text{(}=0 \text{ auf } \Gamma_D\text{)}$$

$$= \int_\Omega (K\nabla u)\cdot\nabla v\, dx + \int_{\Gamma_N} j v\, ds$$

also:

$$\underbrace{\int_\Omega (K\nabla u)\cdot\nabla v\, dx}_{=: a(u,v)} = \underbrace{\int_\Omega f v\, dx - \int_{\Gamma_N} j v\, ds}_{=: \ell(v)}$$

Die entsprechenden Funktionenräume sind dann

$$L_2(\Omega) = \left\{ v : \Omega \to \mathbb{R} \;\middle|\; \int_\Omega v^2 \, dx < \infty \right\},$$

$$H^1(\Omega) = \left\{ v \in L^2(\Omega) \;\middle|\; \int_\Omega v^2 + \nabla v \cdot \nabla v \, dx < \infty \right\},$$

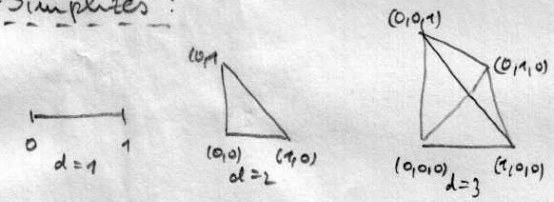
$$V = \left\{ v \in H^1(\Omega) \;\middle|\; "v|_\Gamma = 0" \right\}.$$

Nun gilt es wieder entsprechende endlichdimensionale Teilräume zu konstruieren.

3.2 Triangulierung

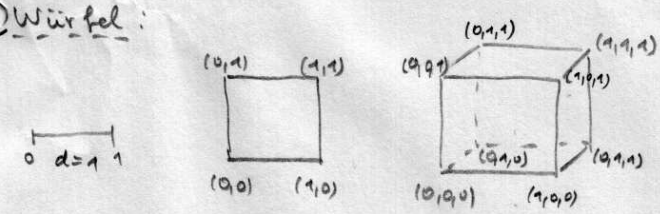
Das Gebiet Ω ist in Teilgebiete „einfacher geometrischer Gestalt" zu zerlegen.

d-Simplizes:



konvexe Hülle der $d+1$ Punkte
$(0, \dots, 0)$
$(1, \dots, 0)$
$(0, \dots, 1)$

d-Würfel:



konvexe Hülle der 2^d Punkte
$(x_1, \dots, x_d) \quad x_i \in \{0,1\}$

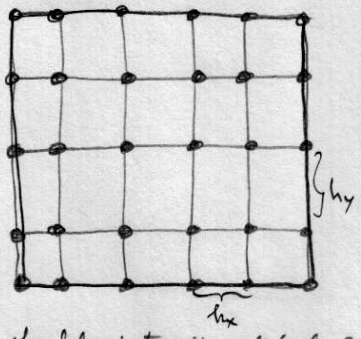
Ω seinen polyedrisch (damit Zerlegung in Simplizes oder Würfel möglich).

Krumm berandete Gebiete sind auch möglich durch
- sukzessive Approximation
- isoparametrische Elemente

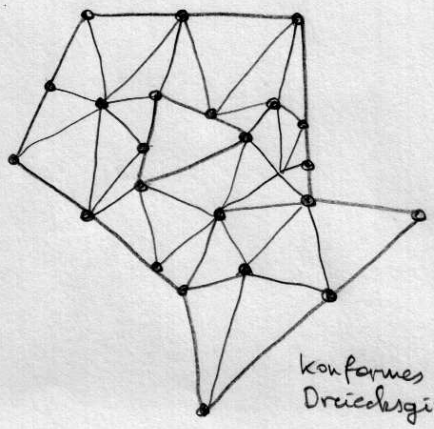
Es gibt auch Pyramiden, Prismen. Notwendig für Mischen von Würfeln u. Simplizes oder Spezialanwendungen.

Beispiele für Triangulierungen: (nur $d=2$)

06.IV.09
3



strukturiertes, äquidistantes Gitter



konformes Dreiecksgitter

Eigenschaften einer Triangulierung.

$\mathcal{T}(\Omega) = \{T_1, T_2, \dots, T_M\}$ sei eine Zerlegung von Ω in abgeschlossene Gebiete ("sog. Elemente") welche entweder Simplizes oder Würfel sind.

↑ Bem 8.IV.09: Mache Elemente offen, das ist sinnvoller für später → Lemma 7.5.

1.) $\mathcal{T}(\Omega)$ heißt konforme Triangulierung falls:

(i) $\bigcup_{i=1}^{M} T_i = \overline{\Omega}$

Bem 8.IV.09: Definiere shape regular extra, siehe Lemma 7.5.

(ii) Für $i \neq j$ ist $T_i \cap T_j$

- eine gemeinsame Ecke von T_i und T_j, oder
- " " Kante
- " " Seite
- leer!

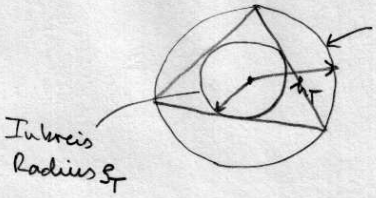
Dimensionsunabhängige Sprechweise: „Entitäten der Codimension k" mit $0 \le k \le d$ (oder, oder, oder)

„Element" ↗ ↖ „Vertex"

Dies ist wesentlich um Teilraume $V_h \subset H^1(\Omega)$ zu konstruieren.

2. $h \in \mathbb{R}^+$ ist die kleinste Zahl so dass jedes Element in einen Kreis mit Durchmesser $2h$ passt. Wir schreiben dann $\mathcal{T}_h(\Omega)$

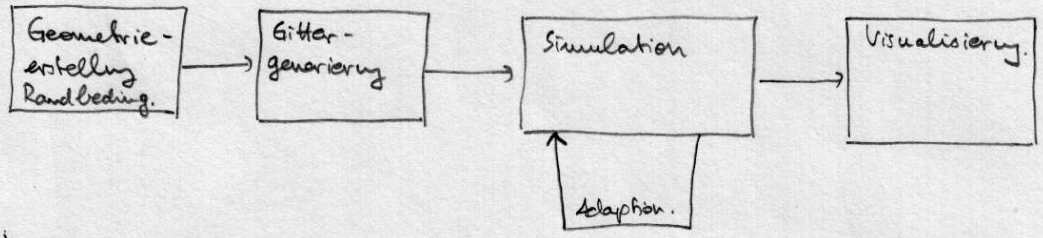
3. $\mathcal{T}_h$ heißt quasiuniform wenn es eine Zahl $K > 0$ gibt so dass jedes $T \in \mathcal{T}_h$ einen Kreis mit Radius $\rho_T \ge \frac{h_T}{K}$ enthält



$$\Rightarrow \quad \frac{h_T}{K} \le \rho_T \le h_T$$

3.3. Workflow ab hier!

06.IV.09
4



VTK, Paraview

Gmsh, Salome

Tetgen
Netgen
Cubit

hierarchische Gitterverfeinerung
Adaptivität und
Fehlerkontrolle

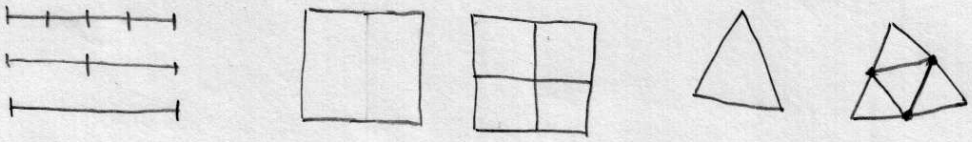
→ Vorlesung Stefan Lang.

3.4 Hierarchische Gitterverfeinerung

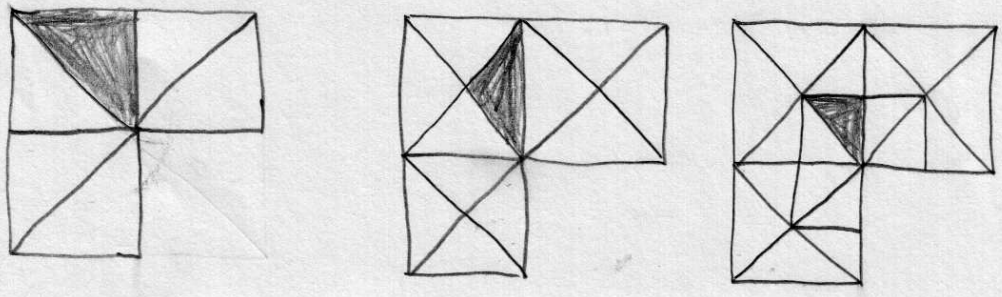
Gitterfeinheit wird durch zwei Kriterien beeinflusst:

1) Komplexe Geometrie muss aufgelöst werden
   Problem: kleine Details, Ecken u. Kanten, anisotrope Objekte
2) Diskretisierungsfehler in der PDE soll kleiner als Toleranz sein.

Gitterverfeinerung ist einfacher als Gittererzeugung:



Lokale Gitterverfeinerung



06.11.09
5

- Baumstruktur erlaubt effiziente Verwaltung
- Erlaubt Verfeinern und Vergröbern (Wegnahme der Verfeinerung)
- Es gibt verschiedene Arten der hierarchischen Verfeinerung (Bisektion, 2^d-Teilung („rote Verfeinerung"), hängende Knoten, ...)
- DUNE bietet sehr viele Möglichkeiten.

3.5 Finite Elemente Räume

Multiindexe Notation:

$\alpha = (\alpha_1, \dots, \alpha_d)$, $\alpha_i \in \mathbb{N}_0$

$|\alpha| = \sum_{i=1}^{d} \alpha_i \qquad |\alpha|_\infty = \max_i \alpha_i$

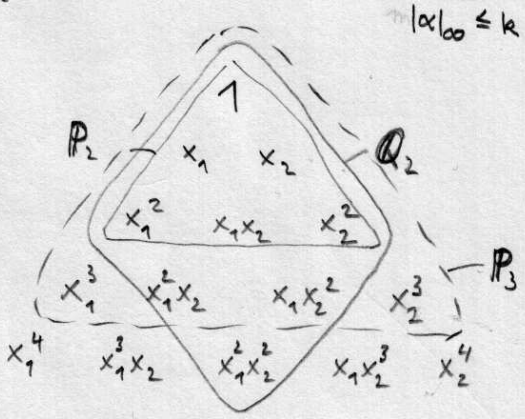
$x \in \mathbb{R}^d$ dann bedeutet $x^\alpha = \prod_{i=1}^{d} x_i^{\alpha_i}$

Polynome in d Raumdimensionen

$$\mathbb{P}_k = \left\{ u : \mathbb{R}^d \to \mathbb{R} \;\middle|\; u(x) = \sum_{|\alpha| \le k} c_\alpha x^\alpha \right\}$$

$$\mathbb{Q}_k = \left\{ u : \mathbb{R}^d \to \mathbb{R} \;\middle|\; u(x) = \sum_{|\alpha|_\infty \le k} c_\alpha x^\alpha \right\}$$

$d=2$



max bis hier!

Konforme Finite Elemente der Ordnung k.

Sei $\mathcal{T}(\Omega)$ eine Triangulierung mit Simplizes:

$$P_k(\mathcal{T}(\Omega)) = \left\{ v \in C^0(\bar{\Omega}) \mid v|_{T_i} \in P_k \right\}.$$

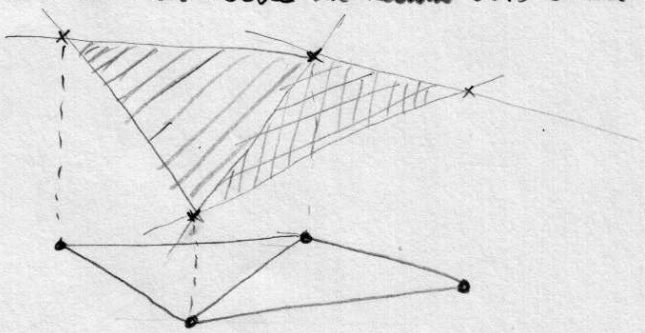
Sei $\mathcal{T}(\Omega)$ eine Triangulierung mit Würfeln:

$$Q_k(\mathcal{T}(\Omega)) = \left\{ v \in C^0(\bar{\Omega}) \mid v|_{T_i} \in Q_k \right\}.$$

Beispiel: Lineare Dreieckselemente, d.h. P_1 für $d=2$:

$$v|_{T_i} = c_{i,0} + c_{i,1} x_1 + c_{i,2} x_2$$

- $v|_{T_i}$ ist durch 3 Koeffizienten $c_{i,0}, c_{i,1}, c_{i,2}$ eindeutig festgelegt, z.B. Werte in den 3 Eckpunkten des Dreiecks
- Kann man sich als Ebene im Raum vorstellen

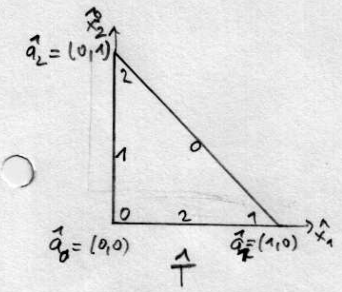


- Stetiger Anschluss an die Nachbardreiecke ist durch lineare Interpolation auf der Kante gewährleistet. Wert auf der Kante hängt nur von den Werten in den beiden anliegenden Ecken ab.
- $\dim P_1(\mathcal{T}(\Omega))$ = Anzahl Knoten der Triangulierung.
- Geht analog für Q_1.
- P_k, Q_k können wieder mit einer Lagrange-Basis ausgestattet werden

3.6 Rückzug auf das Referenzelement

Die Konstruktion der Basisfunktionen gelingt einfach auf dem Referenzelement. Damit lassen sich sog. affine Familien von Finiten Elementen behandeln. (Alle Lagrangeelemente gehören zu dieser Klasse).

Referenzsimplex ($d=2$)



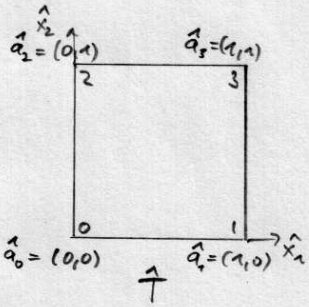
Lagrangebasis für P_1

$$\hat{\varphi}_0(\hat{x}) = 1-\hat{x}_1-\hat{x}_2$$
$$\hat{\varphi}_1(\hat{x}) = \hat{x}_1$$
$$\hat{\varphi}_2(\hat{x}) = \hat{x}_2$$

Es gilt jeweils: $\hat{\varphi}_i(\hat{a}_j)=\delta_{ij}$

$\hat{\varphi}_i$ ist auf der gegenüberliegenden Kante $\emptyset$.

Referenzwürfel ($d=2$)



Für Q_1:

$$\hat{\varphi}_0(\hat{x}) = (1-\hat{x}_1)(1-\hat{x}_2)$$
$$\hat{\varphi}_1(\hat{x}) = \hat{x}_1(1-\hat{x}_2)$$
$$\hat{\varphi}_2(\hat{x}) = (1-\hat{x}_1)\hat{x}_2$$
$$\hat{\varphi}_3(\hat{x}) = \hat{x}_1\hat{x}_2$$

$\hat{\varphi}_i$ ist auf beiden nichtanliegenden Kanten $\emptyset$.

Analog können auch lokale Basisfunktionen für P_k bzw Q_k konstruiert werden.

07.IV.09
8

Transformation.

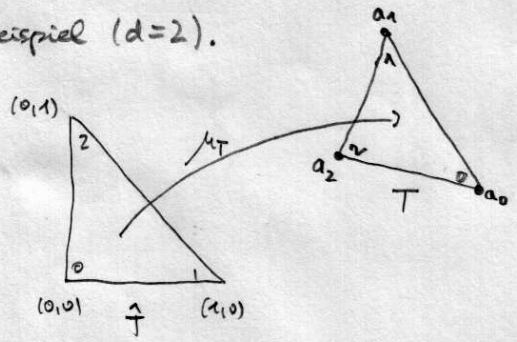
Sei T ein beliebiger d-Simplexe oder d-Würfel. Dann gibt es eine stetig differenzierbare Abbildung $\mu_T : \hat{T} \to T$ vom jeweiligen Referenzelement auf T.

Diese kann leicht angegeben werden. Wir zeigen das für d-Simplizes.

Seien $a_0, \dots, a_d \in \mathbb{R}^d$ die Ecken des d-Simplexe T.

Dann ist $\mu_T(\hat{x}) = \sum_{i=0}^{d} \hat{\varphi}_i(\hat{x})\, a_i$ wobei $\hat{\varphi}_i$ die Lagrangebasis von P_1 auf dem Referenzelement $\hat{T}$ ist.

Beispiel ($d=2$).



$$\mu_T(\hat{x}) = (1-\hat{x}_1-\hat{x}_2)\,a_0 + \hat{x}_1 a_1 + \hat{x}_2 a_2$$
$$= (a_1-a_0)\hat{x}_1 + (a_2-a_0)\hat{x}_2 + a_0$$
$$= \underbrace{\left(a_1-a_0,\ a_2-a_0\right)}_{B_T}\hat{x} + a_0$$

Beobachtung: Für d-Simplizes ist μ_T eine affin-lineare Abbildung

Für d-Würfel geht das ganz genauso, allerdings ist die Abbildung im allg. nicht mehr affin-linear. Sie ist affin-linear für Parallelogramme.

Ist $\{\hat{\varphi}_i\}$ eine Lagrange-Basis für P_k oder Q_k mit Lagrange-Punkten a_i, so ist $\mu_T(\hat{x}) = \sum_{i=0}^{m-1} \hat{\varphi}_i(\hat{x})\, a_i$ eine Abb., die es erlaubt krummlinig berandete Gebiete zu beschreiben. Rechnet man auch mit P_k oder Q_k so bezeichnet man das als isoparametrischen Ansatz.

3.7 Aufstellen der Steifigkeitsmatrix

07.IV.09
9

Sei T ein beliebiger d-Simplex oder d-Würfel mit Referenzelement $\hat{T}$.

Die Abbildung $\mu_T : \hat{T} \to T$ sei affin linear.

$\{\hat{\varphi}_i\}_{i=0}^{n-1}$ sei eine Lagrange-Basis für P_k auf dem Referenzelement zu den Punkten $\{\hat{a}_i\}_{i=0}^{n-1}$.

Dann ist $\Phi^T = \{\varphi_i^T \mid \varphi_i(x) = \hat{\varphi}_i(\mu_T^{-1}(x)),\ i = 0, \dots, n-1\}$

eine Basis für P_k bzw Q_k auf T mit $\varphi_i^T(a_j) = \delta_{ij}$, $a_j = \mu_T(\hat{a}_j)$.

Für eine Funktion $\hat{u} \in C^1(\hat{T})$ gilt mit der Kettenregel:

$$u(x) := \hat{u}(\mu_T^{-1}(x))$$

$$\frac{\partial}{\partial x_i} u(x) = \sum_{j=1}^{d} \frac{\partial \hat{u}}{\partial \hat{x}_j}(\mu_T^{-1}(x)) \; \frac{\partial \mu_T^{-1}, j}{\partial x_i}(x)$$

Damit gilt

$$\nabla u(x) = \begin{bmatrix} \partial_1 u(x) \\ \vdots \\ \partial_d u(x) \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial \mu_T^{-1},1}{\partial x_1}(x) & \cdots & \frac{\partial \mu_T^{-1},d}{\partial x_1}(x) \\ \vdots & & \\ \frac{\partial \mu_T^{-1},1}{\partial x_d}(x) & \cdots & \frac{\partial \mu_T^{-1},d}{\partial x_d}(x) \end{bmatrix}}_{(*)} \begin{bmatrix} \hat{\partial}_1 \hat{u}(\mu_T^{-1}(x)) \\ \vdots \\ \hat{\partial}_d \hat{u}(\mu_T^{-1}(x)) \end{bmatrix}$$

$$\nabla u(x) = \left(\nabla \mu_T^{-1}(x)\right)^T \hat{\nabla} \hat{u}(\mu_T^{-1}(x))$$

$$= \left(\nabla \mu_T(\mu_T^{-1}(x))\right)^{-T} \hat{\nabla} \hat{u}(\mu_T^{-1}(x))$$

$x = \mu_T(\hat{x})$ $\Leftrightarrow$

$$\boxed{\nabla u(\mu_T(\hat{x})) = \left(\nabla \mu_T(\hat{x})\right)^{-T} \hat{\nabla} \hat{u}(\hat{x})}$$

07.IV.09
10

Dies wenden wir nun auf die Basisfunktionen an:

$$A_{ij} = a(\varphi_j, \varphi_i) = \int_\Omega (K(x)\nabla\varphi_j)\cdot\nabla\varphi_i \, dx$$

$$= \sum_{T\in\mathcal{T}(\Omega)} \int_T (K(x)\nabla\varphi_j)\cdot\nabla\varphi_i \, dx$$

Aufsplitten in Elemente

$$= \sum_{T\in\mathcal{T}(\Omega)} \int_{\hat{T}} \left(K(\mu_T(\hat{x}))\nabla\varphi_j(\mu_T(\hat{x}))\right)\cdot\nabla\varphi_i(\mu_T(\hat{x}))\det\nabla\mu_T(\hat{x})\, d\hat{x}$$

Transformation auf das Referenzelement

$$= \sum_{T\in\mathcal{T}(\Omega)} \int_{\hat{T}} \left[\left[K(\mu_T(\hat{x}))\left(\nabla\mu_T(\hat{x})\right)^{-T}\hat{\nabla}\varphi_j(\hat{x})\right]\right]\cdot\left(\nabla\mu_T(\hat{x})\right)^{-T}\hat{\nabla}\varphi_i(\hat{x})\det\nabla\mu_T(\hat{x})\, d\hat{x}$$

Nutze Formel

$$= \sum_{T\in\mathcal{T}(\Omega)} \sum_{s=1}^{m} \left[K(\mu_T(\hat{x}_s))\left(\nabla\mu_T(\hat{x}_s)\right)^{-T}\hat{\nabla}\varphi_j(\hat{x}_s)\right]\cdot\left(\nabla\mu_T(\hat{x}_s)\right)^{-T}\hat{\nabla}\varphi_i(\hat{x}_s)\det\nabla\mu_T(\hat{x}_s)\, w_s$$

Quadratur

↑ Gitterdurchlauf über Elemente

$\hat{\nabla}\varphi_i(\hat{x}_s)$ kann vorab tabelliert werden

$\nabla\mu_T(\hat{x}_s) = B_T$ (unabh. von x_s) im affin-linearen Fall.

Quadratur muss exakt für Polynome vom Grad $2(k-1)$ sein.

Nennt man „Assemblierung"

4 Iterative Lösung schwachbesetzter linearer Gleichungssysteme

Das definive Buch für Iterationsverfahren ist (Hackbusch 1991).

Direkte Verfahren, wie die bekannte Gauß–Elimination, liefern nach endlich vielen Rechenschritten die bis auf Rundungsfehler exakte Lösung x eines linearen Gleichungssystemes $Ax = b$. Die Frage ist nur: Wieviele Rechenoperationen sind notwendig?

Für ein Gleichungssystem der Dimension $n \times n$ benötigt die Gauß–Elimination $\frac{2}{3}n^3 + O(n^2)$ Rechenoperationen (im wesentlichen Additionen und Multiplikationen).

Nun enthalten die Matrizen, die aus der Methode der Finiten Differenzen oder Finiten Elemente resultieren jedoch eine große Zahl von Nullen. So hat die Matrix statt n^2 nur $O(n)$ Einträge. Damit läßt sich einiges sparen. Es zeigt sich, dass man den Aufwand für die Gauß–Elimination in diesem Fall auf $O(n^{1.5})$ reduzieren kann, wenn A die Diskretisierung eines zweidimensionalen Problems darstellt (nested dissection ordering, siehe (Axelsson und Barker 1984)). In drei Raumdimensionen ist der Gewinn nicht so groß: $O(n^2)$.

Unangenehm bei direkten Eliminationsverfahren ist auch, dass im Verlauf der Transformation auf obere Dreiecksgestalt Nullelemente der Matrix ungleich Null werden können, man spricht von „fill-in“. Geschickte Methoden minimieren dieses Auffüllen, können es jedoch nicht vollständig vermeiden. So erhöht sich der Speicherbedarf bei dem oben erwähnten $O(n^{1.5})$-Verfahren von $O(n)$ auf $O(n \log n)$.

4.1 Klassische lineare Iterationsverfahren

Iterative Verfahren zur Lösung von $Ax = b$ gehen von einem beliebigen Startwert $x^0 \in \mathbb{R}^n$ aus und konstruieren eine Folge

$$x^0, x^1, \ldots, x^k, \ldots,$$

die für $k \to \infty$ (hoffentlich) gegen die Lösung x konvergiert. Dies bedeutet, dass jedes x^k nur eine *Näherung* und somit mit einem Fehler, dem *Iterationsfehler*, behaftet ist. Außerdem konvergieren Iterationsverfahren typischerweise nicht für eine beliebige Systemmatrix A. Somit ist zu klären unter welchen Voraussetzungen an A Konvergenz sichergestellt werden kann.

Einige einfache Iterationsverfahren konstruiert man über Defektkorrektur. Sei x^k die Näherung im k-ten Schritt des Verfahrens. Dann ist

$$e^k = x - x^k \tag{4.1}$$

der Fehler im k-ten Schritt. Für diesen Fehler gilt wegen Linearität folgende *Defektgleichung*:

$$Ae^k = A(x - x^k) = Ax - Ax^k = b - Ax^k = d^k. \tag{4.2}$$

Die Größe $d^k = b - Ax^k$ heißt *Defekt* und läßt sich leicht ausrechnen. Nun ist natürlich die Lösung von (4.2) genauso schwierig wie die Lösung des ursprünglichen Gleichungssystemes. Die Idee ist nun (4.2) nur *näherungsweise* zu lösen, indem die Matrix A durch eine leichter zu invertierende Matrix W ersetzt wird. Allerdings bekommen wir dann nurmehr eine Näherung v^k des echten Fehlers e^k, also setze

$$Wv^k = d^k. \tag{4.3}$$

Wegen $x = x^k + e^k \approx x^k + v^k$ ist $x^{k+1} = x^k + v^k$ nun eine neue und hoffentlich genauere Näherungslösung. Wir erhalten damit das *Iterationsverfahren*

$$x^{k+1} = x^k + W^{-1}(b - Ax^k). \tag{4.4}$$

Offensichtlich ist die exakte Lösung x Fixpunkt dieses Iterationsverfahrens.
Geeignete Kandidaten für W sind:

$$W_{Ric} = \frac{1}{\omega} I \qquad \text{Richardson-Iteration} \tag{4.5a}$$

$$W_{Jac} = \operatorname{diag}(A) \qquad \text{Jacobi-Iteration} \tag{4.5b}$$

$$W_{GS} = L(A) \qquad \text{Gauß-Seidel-Iteration} \tag{4.5c}$$

Bei allen drei Varianten erfordert die Auflösung von des Systems $Wv = d$ nur $O(n)$ Operationen falls A nur $O(n)$ Einträge hat. Der Gesamtaufwand für einen Iterationsschritt ist somit $O(n)$.

Sowohl beim Jacobi-Verfahren als auch beim Gauß-Seidel-Verfahrem müssen die Diagonalelmente der Matrix ungleich Null sein.

Wir wollen nun einige Aussagen zum Konvergenzverhalten der Iterationsverfahren machen. Dazu machen wir zunächst die

Bemerkung 4.1 (Fehlerfortpflanzung) Für die Iteration aus (4.4) gilt

$$e^{k+1} = (I - W^{-1}A)e^k. \tag{4.6}$$

$S = (I - W^{-1}A)$ wird als *Iterationsmatrix* bezeichnet.
Beweis: $e^{k+1} = x - x^{k+1} = x - x^k - W^{-1}(b - Ax^k) = x - x^k - W^{-1}A(x - Ax^k) = (I - W^{-1}A)e^k$. □

Iterationsverfahren der Bauart (4.4) werden aus diesem Grund als lineare Iterationsverfahren bezeichnet. Zur Charakterisierung der Konvergenz linearer Iterationen benötigen wir die

Definition 4.1 (Spektrum, Spektralradius) Es bezeichne

$$\sigma(A) = \{\lambda \in \mathbb{C} | \lambda \text{ ist Eigenwert von } A \} \tag{4.7}$$

das *Spektrum* der Matrix A und

$$\varrho(A) = \max\{|\lambda| \, | \lambda \in \sigma(A)\} \tag{4.8}$$

ihren *Spektralradius*.

Dann gilt der

Satz 4.2 (Konvergenz linearer Iterationen) Die durch (4.4) gegebene Iteration konvergiert genau dann wenn

$$\varrho(S) < 1 \tag{4.9}$$

mit der Iterationsmatrix $S = I - W^{-1}A$.
Beweis: Idee: $e^k = S^k e^0$, $S^k \to 0$, genauer siehe (HACKBUSCH 1991). □

Für symmetrisch positiv definite Matrizen ($x^T Ax > 0 \ \forall x \neq 0$) gibt folgender Satz Auskunft über die Konvergenz der Richardson-Iteration:

Satz 4.3 Sei A eine symmetrische und positiv definite Matrix mit kleinstem Eigenwert $\lambda_{\min}(A)$ und größtem Eigenwert $\lambda_{\max}(A)$, so konvergiert die mit $\omega = 1/\lambda_{\max}(A)$ gedämpfte Richardson-Iteration mit der Rate

$$\varrho(I - \omega A) = 1 - \frac{1}{\kappa(A)}.$$

Dabei bezeichnet $\kappa(A) = \lambda_{\max}(A)/\lambda_{\min}(A)$ die *Konditionszahl* von A.
Beweis: Wegen der Voraussetzung sind alle Eigenwerte von A reell und positiv, d. h. $0 < \lambda_{\min}(A) \leq \ldots \leq \lambda_i \leq \ldots \leq \lambda_{\max}(A)$. Für die Eigenwerte μ_i der Iterationsmatrix $S = I - \lambda_{\max}(A)^{-1}A$ gilt daher $\mu_i = 1 - \lambda_i/\lambda_{\max}(A)$ und daher $\max_i |\mu_i| = 1 - \lambda_{\min}(A)/\lambda_{\max}(A)$. □

Beachte, dass die richtige Wahl des Dämpfungsfaktors bei der Richardson-Iteration (und auch beim Jacobi-Verfahren für allgemeines symmetrisch positiv definites A) entscheidend ist. Wählt man ω zu groß, so konvergiert die Iteration nicht. Wählt man ω andererseits zu klein so konvergiert die Iteration zwar aber möglicherweise langsamer als optimal wäre ($\omega = 1/\lambda_{\max}$ ist nicht ganz optimal!). Man benötigt also eine hinreichend genaue Schätzung des größten Eigenwertes von A.

Für Finite-Elemente-Diskretisierungen von elliptischen Problemen zweiter Ordnung kann man $\kappa(A) = O(h^{-2})$ zeigen. Somit besitzt das Richardson-Verfahren die asymptotische Konvergenzrate $\varrho = 1 - \frac{1}{Ch^2}$. Es zeigt sich, dass sich auch die Jacobi- bzw. Gauß-Seidel-Iteration asymptotisch genauso verhalten (mit anderer Konstante C).

Frägt man nach dem Aufwand für die Lösung eines linearen Gleichungssystemes so findet man, dass die Anzahl der Iterationen, die man benötigt um den Fehler um einen Faktor ε (z. B. $\varepsilon = 10^{-10}$) zu reduzieren proportional zur Konditionszahl $\kappa(A)$ ist. Für $d = 2$ gilt $h = 1/\sqrt{n}$, eine Iteration hat Aufwand $O(n)$, somit beträgt die Gesamtkomplexität $O(n^2)$.

Ziel dieser Vorlesung ist die Konstruktion von Iterationsverfahren, deren Konvergenzrate nicht oder nur sehr schwach von der Konditionszahl abhängig ist *und* die sich effizient auf parallelen Rechnerarchtiekturen umsetzen lassen.

4.2 Blockvarianten

Als Vorstufe zu den Gebietszerlegungsverfahren behandeln wir nun Blockvarianten der Jacobi- und Gauß-Seidel-Iteration.

Da Indexmengen in der ganzen Vorlesung eine entscheidende Rolle spielen bringen wir hier erst einige Definitionen, siehe hierzu auch (HACKBUSCH 1991, Abschnitt 2.1.1).

Grundsätzlich bezeichnet eine endliche Menge $I \subseteq \mathbb{N}$ eine Indexmenge. Ein Vektor $x \in \mathbb{R}^I$ enthält für jedes $i \in I$ genau eine Komponente. Hierbei ist wichtig, dass

- die Indizes nicht bei 1 beginnen müssen und nicht konsekutiv sein müssen,
- und keine Reihenfolge der Indizes festgelegt ist (die Indexmenge ist nicht angeordnet).

Die Komponenten des Vektors $x \in \mathbb{R}^I$ werden mit $(x)_i, i \in I$ bezeichnet. Dabei setzen wir immer explizit Klammern damit Komponenten nicht mit Indizes verwechselt werden.

Eine weitere Schreibweise betrifft das Herausschneiden von Teilvektoren. Ist $\tilde{I} \subseteq I$ eine Teilmenge der Indexmenge I und $x \in \mathbb{R}^I$ so bezeichnet $x_{\tilde{I}}$ einen Vektor aus $\mathbb{R}^{\tilde{I}}$ der auf den Komponenten $i \in \tilde{I}$ die selben Werte wie x besitzt.

Dies überträgt sich alles analog auf Matrizen. So besitzen (im allgemeinen rechteckige) Matrizen $A \in \mathbb{R}^{I \times J}$ für jedes Indexpaar $(i, j) \in I \times J$ eine Komponente $(A)_{ij} \in \mathbb{R}$. Für $\tilde{I} \subseteq I, \tilde{J} \subseteq J$ bezeichnet $A_{\tilde{I},\tilde{J}}$ die entsprechende Untermatrix von A.

$A \in \mathbb{R}^{I \times J}$ fassen wir als lineare Abbildung $A : \mathbb{R}^I \to \mathbb{R}^J$ auf. $y = A(x) = Ax$ ist natürlich gegeben durch

$$(y)_i = \sum_{j \in J} (A)_{ij} (x)_j$$

(man beachte, dass keine Anordnung notwendig ist).

Wir weisen noch darauf hin, dass I häufig auch als Symbol für die Einheitsmatrix verwendet wird, jedoch sollte aus dem Kontext klar hervorgehen was gemeint ist.

Will man einen Vektor oder eine Matrix darstellen, so ist eine Anordnung der Indexmenge(n) notwendig. So definiert beispielsweise die lexikographische Anordnung, dass die Indizes der Größe nach geordnet werden.

Damit kommen wir nun zur Definition der Blockvarianten. Dazu sei I die Indexmenge einer quadratischen Matrix A (z. B. wären dies die Nummern der inneren Knoten der Triangulierung in unserer Anwendung). Wir wählen eine Anzahl von Blöcken p und bezeichnen mit

$$B = \{1, \ldots, p\}$$

die Indexmenge der Blöcke. Dann sei die Indexmenge I zerlegt in disjunkte, nichtleere Teilmengen

$$I = \bigcup_{i \in B} I_i, \qquad I_i \cap I_j = \emptyset \quad \forall i \neq j.$$

Das Block-Jacobi und das Block-Gauß-Seidel-Verfahren sind dann definiert durch die Matrizen

$$(W_{BJac})_{ij} = \begin{cases} (A)_{ij} & \text{wenn } i, j \in I_k \text{ für ein } k \in B \\ 0 & \text{sonst} \end{cases} \tag{4.10a}$$

$$(W_{BGS})_{ij} = \begin{cases} (A)_{ij} & \text{wenn } i \in I_k, j \in I_l \wedge l \leq k \\ 0 & \text{sonst} \end{cases} \tag{4.10b}$$

dabei wurde für das Block-Gauß-Seidel-Verfahren eine lexikographische Anordnung der Blöcke verwendet. Unter dieser Anordnung der Blöcke haben die Matrizen W_{BJac}, W_{GS} die Gestalt

$$W_{BJac} = \begin{pmatrix} A_{I_1,I_1} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & A_{I_p,I_p} \end{pmatrix}, \quad W_{BGS} = \begin{pmatrix} A_{I_1,I_1} & 0 & \cdots & 0 \\ A_{I_2,I_1} & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ A_{I_p,I_1} & \cdots & A_{I_p,I_{p-1}} & A_{I_p,I_p} \end{pmatrix} \tag{4.11}$$

Die Anwendung der Blockvarianten erfordert die Lösung von p Gleichungssystemen der Dimensionen $|I_1|, \ldots, |I_p|$.

Wir wollen nun die Blockvarianten noch in *algorithmischer Form* schreiben. Dazu definieren wir für jedes $i \in B$ die lineare Abbildung $R_i : \mathbb{R}^I \to \mathbb{R}^{I_i}$ als

$$R_i x = x_{I_i}$$

oder mit anderen Worten

$$(R_i x)_\alpha = (x)_\alpha \quad \forall \alpha \in I_i.$$

R_i ist eine Rechteckmatrix mit genau einer 1 pro Zeile und maximalem Rang.

Wegen $x^{k+1} = x^k + W_{BJac}^{-1}(b - Ax^k)$ erhalten wir für das Block-Jacobi-Verfahren die Darstellung:

$$x^{k+1} = x^k + \sum_{i \in B} R_i^T A_{I_i,I_i}^{-1} R_i (b - Ax^k).$$

Offensichtlich können alle Korrekturen unabhängig voneinander berechnet werden. Nicht so beim Block-Gauß-Seidel-Verfahren. Hier ergibt sich:

for $i = 1, 2, \ldots, p$
$\quad x^{k+\frac{i}{p}} = x^{k+\frac{i-1}{p}} + R_i^T A_{I_i,I_i}^{-1} R_i (b - Ax^{k+\frac{i-1}{p}})$

d. h. die Korrekturen entsprechend den einzelnen Blöcken werden nacheinander berechnet.

4.3 Abstiegsverfahren

Satz 4.4 Ist A symmetrisch positiv definit, dann nimmt das Funktional

$$F(x) = \frac{1}{2} x^{\mathrm{T}} A x - b^{\mathrm{T}} x$$

sein eindeutiges Minimum in $x^* = A^{-1}b$, der Lösung des lineaeren Gleichungssystems $Ax^* = b$, an.

Beweis: Für beliebiges x setze $x = x^* + v$ und rechne

$$\begin{aligned} F(x) = F(x^* + v) &= \frac{1}{2}(x^* + v)^{\mathrm{T}} A (x^* + v) - b^{\mathrm{T}}(x^* + v) \\ &= \frac{1}{2}\Big[x^{*\mathrm{T}} A x^* + \underbrace{x^{*\mathrm{T}} A v + v^{\mathrm{T}} A x^*}_{2v^{\mathrm{T}} A x^*} + v^{\mathrm{T}} A v\Big] - b^{\mathrm{T}} x^* - b^{\mathrm{T}} v \\ &= \underbrace{\frac{1}{2} x^{*\mathrm{T}} A x^* - b^{\mathrm{T}} x^*}_{F(x^*)} + v^{\mathrm{T}}\big(\underbrace{A x^* - b}_{=0}\big) + \frac{1}{2} v^{\mathrm{T}} A v \\ &= F(x^*) + \frac{1}{2}\underbrace{v^{\mathrm{T}} A v}_{>0 \text{ für } v \neq 0} > F(x^*) \end{aligned}$$

Eindeutigkeit: Sei F in x^* und $x' \neq x^*$ minimal, so gilt für $x' = x^* + v$, $v \neq 0$: $F(x') = F(x^*) + \frac{1}{2} v^{\mathrm{T}} A v > F(x^*)$, also $\lightning$ zu F minimal in x'. □

Gegeben ein $x^{(k)}$, eine Näherungslösung von x^*, so kann man $x^{(k)}$ in Richtung einer ebenfalls gegebenen *Suchrichtung* $p^{(k)} \in \mathbb{R}^N$ verbessern, indem man

$$F\left(x^{(k)} + \alpha p^{(k)}\right) \to \min$$

löst.

Das optimale α lässt sich wie folgt berechnen. Zunächst rechne aus:

$$F\left(x^{(k)} + \alpha p^{(k)}\right) = F\left(x^{(k)}\right) + \alpha \left(p^{(k)}\right)^{\mathrm{T}} \left(A x^{(k)} - b\right) + \frac{\alpha^2}{2}\left(p^{(k)\mathrm{T}} A p^{(k)}\right)$$

(setze $x^* = x^{(k)}$ und $v = \alpha p^{(k)}$ in obiger Rechnung).

Eine notwendige Bedingung für ein Minimum ist das Verschwinden der Ableitung

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x} F\left(x^{(k)} + \alpha p^{(k)}\right) &= \left(p^{(k)}\right)^{\mathrm{T}} \left(A x^{(k)} - b\right) + \alpha p^{(k)\mathrm{T}} A p^{(k)} \stackrel{!}{=} 0 \\ \iff \alpha^{(k)} &= \frac{\left(p^{(k)}\right)^{\mathrm{T}} \overset{\text{Defekt!}}{\overset{\downarrow}{\left(b - A x^{(k)}\right)}}}{\underbrace{\left(p^{(k)}\right)^{\mathrm{T}} A p^{(k)}}_{\neq 0 \text{ wg. pos. definit und } p \neq 0}} \end{aligned}$$

Die zweite Ableitung ist gerade

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} F\left(x^{(k)} + \alpha p^{(k)}\right) = p^{(k)\mathrm{T}} A p^{(k)}$$

also positiv und es liegt tatsächlich ein Minimum vor.

Wähle nun die spezielle Suchrichtung $p^{(k)} = -\nabla F\left(x^{(k)}\right)$ (negative Gradientenrichtung). Ausgeschrieben gilt

$$F(x) = \frac{1}{2}\underbrace{\sum_{i=1}^{N}\sum_{j=1}^{N} x_i a_{ij} x_j} - \sum_{i=1}^{N} b_i x_i$$

und damit

$$\frac{\partial F}{\partial x_m} = \frac{1}{2}\left\{2a_{mm}x_m + \sum_{j\neq m} a_{mj}x_j + \sum_{i\neq m} x_i a_{im}\right\} = (Ax-b)_m.$$

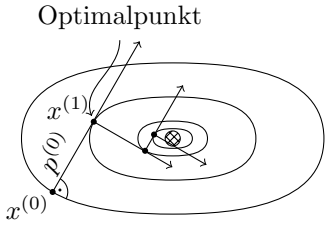
also gilt $p^{(k)} = -\nabla F\left(x^{(k)}\right) = b - Ax^{(k)}$ und es ergibt sich der folgende Algorithmus, das sogenannte „Gradientenverfahren“:

geg. x, b;
$d = b - Ax$;
$d_0 = \|d\|$;
$d_k = d_0$;
while$(d_k > \varepsilon d_0)\{$
    $q = Ad$;
    $\alpha = \frac{d^{\mathrm{T}}d}{d^{\mathrm{T}}q}$;
    $x = x + \alpha d$;
    $d = d - \alpha q$;
$\}$

Problem:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad F(x) = x_1^2 + \varepsilon x_2^2$$

Höhenlinien von F sind Ellipsen um den Ursprung



- „Zickzackkurve = langsame Konvergenz“
- Man zeigt:

$$\|e^{(k)}\|_A \leq \frac{\kappa(A)+1}{\kappa(A)-1}\|e^{(k-1)}\|_A.$$

- „Energienorm“ $\|x\|_A = \sqrt{x^{\mathrm{T}}Ax}$.
- „Energieskalarprodukt“

$$\langle x, y\rangle_A = x^{\mathrm{T}}Ay.$$

Wir lernen zwei Möglichkeiten zur Verbesserung des Gradientenverfahrens kennen, die auch kombiniert werden können.

4.3.1 Vorkonditioniertes Gradientenverfahren

Idee: Wende Gradientenverfahren auf das transformierte System

$$M^{-1}Ax = M^{-1}b \tag{4.12}$$

an.

Problem: $M^{-1}A$ ist im allgemeinen nicht symmetrisch, selbst wenn M^{-1} und A symmetrisch sind.

Ist M symmetrisch positiv definit, so gibt es aber ein T mit $M = TDT^{-1}$ und $D = \text{diag}(d_{ii})$, $d_{ii} > 0$. Man definiert dann formal

$$M^{\frac{1}{2}} = TD^{\frac{1}{2}}T^{-1} \qquad \text{mit } \left(D^{\frac{1}{2}}\right)_{ii} = \sqrt{d_{ii}}$$

denn damit gilt $M^{\frac{1}{2}}M^{\frac{1}{2}} = M$.

Damit ist $M^{-1}A$ ähnlich zu $M^{\frac{1}{2}}M^{-1}AM^{-\frac{1}{2}} = M^{-\frac{1}{2}}AM^{-\frac{1}{2}}$, also

$$\sigma(M^{-1}A) = \sigma(M^{-\frac{1}{2}}AM^{-\frac{1}{2}}).$$

Nun multipliziere (4.12) von links mit $M^{\frac{1}{2}}$ und setze $\hat{x} = M^{\frac{1}{2}}x$

$$\underbrace{M^{-\frac{1}{2}}AM^{-\frac{1}{2}}}_{\hat{A}}\underbrace{M^{\frac{1}{2}}x}_{\hat{x}} = \underbrace{M^{-\frac{1}{2}}b}_{\hat{b}}$$
$$\iff \qquad \hat{A}\hat{x} = \hat{b}.$$

Das transformierte System $\hat{A}\hat{x} = \hat{b}$ is symmetrisch positiv definit und hat die Eigenwerte von $M^{-1}A$.

Das Gradientenverfahren ist formal anwendbar:

geg. $\hat{x}, \hat{b}$;
$\hat{d} = \hat{b} - \hat{A}\hat{x}$;
while(...){
  $\hat{q} = \hat{A}\hat{d}$;
  $\hat{\alpha} = \frac{\hat{d}^T\hat{d}}{\hat{d}\hat{q}}$;
  $\hat{x} = \hat{x} + \hat{\alpha}\hat{d}$;
  $\hat{d} = \hat{d} - \hat{\alpha}\hat{q}$;
}

Allerdings möchte man das Verfahren in dieser Weise nicht praktisch durchführen, da $\hat{A}$ im allgemeinen nicht mehr dünn besetzt ist.

Die Idee ist nun die Transformation *in jedem einzelnen Schritt* des Gradientenvefahrens zu berücksichtigen aber nur die untransformierten Größen zu spreichern.

Gegeben seien also x, b

$$\text{Beachte:} \quad x = M^{-\frac{1}{2}}\hat{x} \quad \text{und} \quad \hat{b} = M^{-\frac{1}{2}}b \quad \text{sowie} \quad \hat{A} = M^{-\frac{1}{2}}AM^{-\frac{1}{2}}$$
$$\hat{x} = M^{\frac{1}{2}}x \qquad\qquad b = M^{\frac{1}{2}}\hat{b}$$
$$\hat{d} = \hat{b} - \hat{A}\hat{x} = M^{-\frac{1}{2}}b - \left(M^{-\frac{1}{2}}AM^{-\frac{1}{2}}\right)\left(M^{\frac{1}{2}}x\right) = M^{-\frac{1}{2}}\Big(\underbrace{b - Ax}_{d}\Big)$$

berechne $d = b - Ax$;
while(...) {

$$\hat{q} = \hat{A}\hat{d} = \left(M^{-\frac{1}{2}} A M^{-\frac{1}{2}}\right)\left(M^{-\frac{1}{2}} d\right) = M^{-\frac{1}{2}} \underbrace{A \underbrace{M^{-1} d}_{v}}_{q};$$

berechne nur $q = Av$; $v = M^{-1}d$;

$$\underset{\substack{\uparrow \\ \alpha \text{ und } \hat{\alpha} \\ \text{sind identisch.}}}{\hat{\alpha}} = \frac{\hat{d}^{\mathrm{T}}\hat{d}}{\hat{d}^{\mathrm{T}}\hat{q}} = \frac{\left(M^{-\frac{1}{2}} d\right)^{\mathrm{T}}\left(M^{-\frac{1}{2}} d\right)}{\left(M^{-\frac{1}{2}} d\right)^{\mathrm{T}}\left(M^{-\frac{1}{2}} q\right)} = \frac{d^{\mathrm{T}}\left(M^{-1} d\right)}{d^{\mathrm{T}}\left(M^{-1} q\right)} = \frac{d^{\mathrm{T}} v}{q^{\mathrm{T}} v}$$

$$\hat{x} = \hat{x} + \hat{\alpha}\hat{d} = M^{\frac{1}{2}} x + \hat{\alpha} \cdot M^{-\frac{1}{2}} d = M^{\frac{1}{2}}\left(x + \hat{\alpha} \underbrace{M^{-1} d}_{v}\right)$$

$$= M^{\frac{1}{2}}(\underbrace{x + \hat{\alpha} v}_{})$$

speichere nur das x und seine updates.

$\hat{x}$ taucht nirgendwo anders auf!

$$\hat{d} = \hat{d} - \hat{\alpha}\hat{q} = M^{-\frac{1}{2}} d - \hat{\alpha} M^{-\frac{1}{2}} q = M^{-\frac{1}{2}}(\underbrace{d - \hat{\alpha} q}_{})$$

speichere nur d und seine updates.

}

Das sogenannte vorkonditionierte Gradientenverfahren lässt sich damit folgendermaßen algorithmisch realisieren:

geg. x, b;
$d = b - Ax$;
$d_0 = \|d\|$;
$d_k = d_0$;
while$(d_k > \varepsilon d_0)\{$
    Löse $v = M^{-1}d$; // M ist sym. pos. definit
    $q = Av$;
    $\alpha = \frac{d^{\mathrm{T}} v}{q^{\mathrm{T}} v}$;
    $x = x + \alpha v$;
    $d = d - \alpha q$;
    $d_k = \|d\|$;
}

4.3.2 Konjugierte Gradienten Verfahren

Die Vorkonditionierung behebt nicht das Problem der langsamen Konvergenz des Gradientenverfahrens, d.h. das Verfahren ist nicht schneller als die Basisiteration $\varrho\left(I - M^{-1}A\right)$.

Die liegt daran, dass das Gradientenverfahren die Optimalität bezüglich einer Suchrichtung wieder verliert.

Sei $p^{(k)}$ eine Suchrichtung. Die Iterierte $x^{(k)}$ ist in Richtung $p^{(k)}$ bereits optimal, falls

$$\alpha = \frac{\left(d^{(k)}\right)^{\mathrm{T}} p^{(k)}}{\left(p^{(k)}\right)^{\mathrm{T}} A p^{(k)}} = 0, \quad \text{was nur für} \quad \left(d^{(k)}\right)^{\mathrm{T}} p^{(k)} = 0 \quad \text{möglich ist.}$$

Nach einem Schritt des Gradientenverfahrens gilt zwar

$$\underbrace{\left(d^{(k)}\right)^{\mathrm{T}}}_{\substack{\text{aktueller}\\ \text{Defekt}}} \underbrace{d^{(k-1)}}_{\substack{\text{letzte}\\ \text{Suchrichtung}}} = 0$$

(nachrechnen!), aber leider im allgemeinen bereits

$$\underbrace{\left(d^{(k)}\right)^{\mathrm{T}}}_{\substack{\text{aktueller}\\ \text{Defekt}}} \underbrace{d^{(k-2)}}_{\substack{\text{vorletzte}\\ \text{Suchrichtung}}} \neq 0,$$

d. h. die Optimalität bezüglich aller Suchrichtungen geht verloren.

Seien $p^{(k)}$ die im Laufe eines Abstiegsverfahrens verwendeten Richtungen (nicht notwendigerweise die Gradientenrichtung).

Minimierung in Richtung $p^{(k)}$ im Schritt k liefert den neuen Defekt

$$d^{(k+1)} = d^{(k)} - \alpha^{(k)} A p^{(k)}.$$

Damit gilt dann natürlich

$$\begin{aligned}\left(d^{(k+1)}\right)^{T} p^{(k)} &= \left(d^{(k)} - \alpha^{(k)} A p^{(k)}\right)^{T} p^{(k)} \\ &= \left(d^{(k)}\right)^{T} p^{(k)} - \frac{\left(d^{(k)}\right)^{\mathrm{T}} p^{(k)}}{\left(p^{(k)}\right)^{\mathrm{T}} A p^{(k)}} \left(p^{(k)}\right)^{T} A p^{(k)} = 0\end{aligned}$$

Damit $x^{(k+1)}$ *zusätzlich* noch optimal bezüglich aller alten Richtungen $p^{(0)}, \ldots, p^{(k-1)}$ ist, muss gelten

$$\left(d^{(k+1)}\right)^{\mathrm{T}} p^{(l)} = 0 \qquad \forall\ 0 \leq l < k$$

also

$$\left(d^{(k)} - \alpha^{(k)} A p^{(k)}\right) p^{(l)} = \underbrace{\left(d^{(k)}\right)^{\mathrm{T}} p^{(l)}}_{\substack{=0\\ \text{per Induktion im Schritt}\\ (k-1) \text{ für } d^{(k)} \text{ sichergestellt}}} - \alpha^{(k)} \underbrace{\left(A p^{(k)}\right)^{\mathrm{T}} p^{(l)}}_{\substack{=0\\ \text{im Schritt } (k)\\ \text{sicherzustellen}}} = 0$$

Wählt man also die $p^{(k)}$ so, dass $\left(p^{(k)}\right)^{\mathrm{T}} A p^{(l)}\ \forall\ 0 \leq l < k$, so bleibt die Optimalität bezüglich aller alten Richtungen erhalten.

Dies nennt man „Verfahren der konjugierten Richtungen“.

Die Verbindung mit Gradientenverfahren liefert das „Verfahren der konjugierten Gradienten“. Für dieses Verfahren zeigt man die Konvergenzrate

$$\|e^{(k)}\|_A \leq \frac{\sqrt{\kappa(A)} + 1}{\sqrt{\kappa(A)} - 1} \|e^{(k-1)}\|_A.$$

Auch hier kann wieder die Technik der Vorkonditionierung eingesetzt werden.

4.4 Parallelisierung des vorkonditionierten Gradientenverfahrens

4.4 Parallelisierung des Vorkonditionierten Gradientenverf. 23.IV.09 1

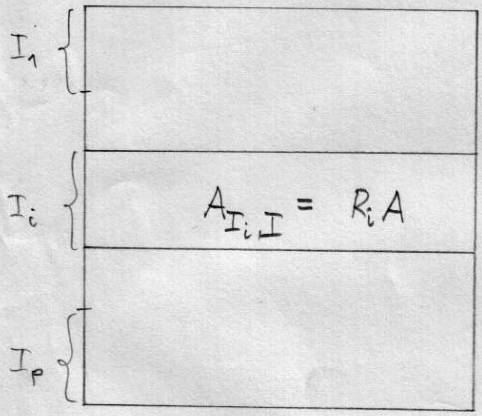
Blockverfahren

Jacobi: $I_i \subset I$ Partitionierung

$$x^{k+1} = x^k + \sum_{i=1}^{p} R_i^T A_i^{-1} R_i (b - Ax^k)$$

mit $A_i = R_i A R_i^T$

Datenaufteilung: Prozess $i \in \{1, \dots, p\}$ erhält alle Zeilen $l \in I_i$ der Matrix A.



A ist dünn besetzt.

Zu $I_i \subset I$ definiere $I_i^1 = \{ j \in I \mid \exists k \in I_i : a_{kj} \neq 0 \}$

allgemein $I_i^0 := I_i$

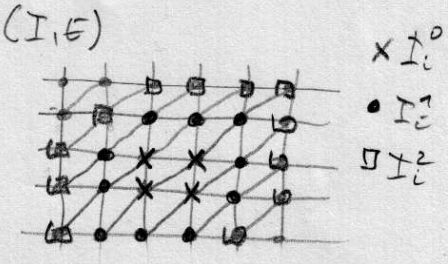
$I_i^{l+1} := \{ j \in I \mid \exists k \in I_i^l : a_{kj} \neq 0 \}$

Wegen Symmetrie von A gilt $I_i = I_i^0 \subseteq I_i^1 \subseteq I_i^2 \subseteq \dots \subseteq I$

Besser: Graph einer Matrix $G(A) = (I, E)$

$E = \{ (i,j) \in I \times I \mid a_{ij} \neq 0 \}$

„algebraischer Überlapp"



Damit genügt es wenn Prozess i die Matrix A_{I_i, I_i^1} speichert.
Der Rest sind ohnehin nur Nullen.

Damit lautet der parallele Block-Vorkonditionierer ausführlich

$\forall i \in \{1, \dots, p\}$: do parallel

$x^0_{I_i^1} = R_{I_i^1} x^0$ Startwert

$b_{I_i} = R_i b$ rechte Seite

$R_{I_i^1} : \mathbb{R}^I \to \mathbb{R}^{I_i^1}$

for ($k = 1, \dots$) {

$d^k_{I_i} = b_{I_i} - A_{I_i, I_i^1} x^k_{I_i^1}$

$R_{I_i^1, I_i} : \mathbb{R}^{I_i^1} \to \mathbb{R}^{I_i}$

$v_{I_i^1} = R^T_{I_i^1 I_i} A^{-1}_{I_i, I_i} d^k_{I_i}$ // A_{I_i, I_i} ist Untermatrix von A_{I_i, I_i^1}

$v_{I_i^1} = R_{I_i^1} \sum_{\{j : I_i^1 \cap I_j^1 \neq \emptyset\}} R^T_{I_j^1} v_{I_j^1}$ // erfordert Kommunikation
es genügt lokale Kommunikation

$x^{k+1}_{I_i^1} = x^k_{I_i^1} + v_{I_i^1}$

}

Bemerkung: - Dies ist nur eine Möglichkeit
- DUNE nimmt eine Aufteilung der codim 0 entities vor

23. IV. 09
3

Gradientenverfahren parallel

$\forall i \in \{1, \dots, p\}$: do parallel

$x^0_{I_i^1} = R_{I_i^1} x^0$;

$b_{I_i} = R_i b$;

$d_{I_i} = b_{I_i} - A_{I_i, I_i^1} x_{I_i}$;

$\delta^0 = \| d_{I_i} \|^2$

$\delta^0 = \sqrt{\sum_{j=1}^{p} \delta_j}$ // globale Norm berechnen

$\delta = \delta^0$;

while ($\delta > \varepsilon \delta^0$) {

$v_{I_i^1} = prec(d_{I_i})$; // eine Kommunikation im Vork.

$q_{I_i} = A_{I_i, I_i^1} v$; // lokal

$z_i = d_{I_i} \cdot (R_{I_i^1, I_i} v_{I_i^1})$;

$n_i = q_{I_i} \cdot (R_{I_i^1, I_i} v_{I_i^1})$;

$z_i = \sum_{j=1}^{p} z_j$

$n_i = \sum_{j=1}^{p} n_j$

} 1 MPI_Reduce

$\alpha = z_i / n_i$

$x_{I_i^1} = x_{I_i^1} + \alpha v_{I_i^1}$;

$d_{I_i} = d_{I_i} - \alpha q_{I_i}$

$\delta = \| d_{I_i} \|^2$;

$\delta = \sqrt{\sum_{j=1}^{p} \delta}$; MPI_Reduce

}

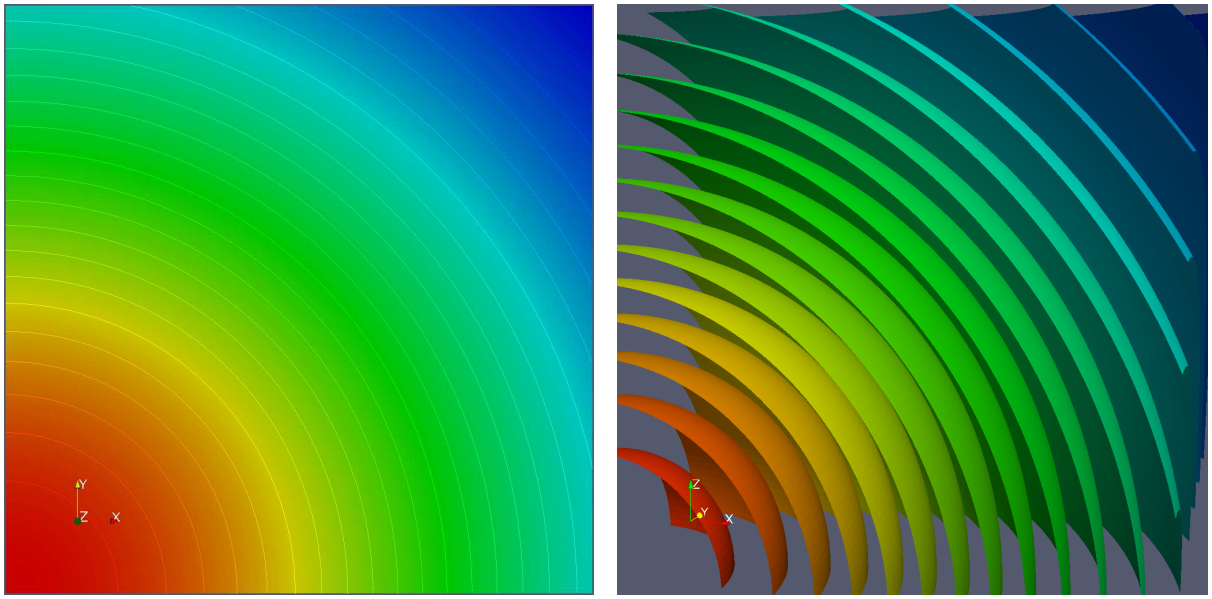


Abbildung 4.1: Lösung des Modellproblems A in 2d und 3d.

4.5 Numerische Resultate

Wir geben nun einige numerische Resultate der bisher behandelten Methoden für verschieden schwierige Testprobleme in zwei und drei Raumdimensionen.

Angegeben sind jeweils die Anzahl der benötigten Iterationen um den anfänglichen Defekt um den Faktor 10^8 zu reduzieren. Die Rechenzeit ist in Sekunden angegeben (Core 2 Duo Prozessor mit 2.5 GHz, gcc-4.2 mit -O2 Optimierung). Es waren maximal 20000 Iterationen erlaubt. Fehlende Einträge bedeuten, dass die Reduktion innerhalb der erlaubten Zahl von Iterationen nicht erreicht wurde.

4.5.1 Modellproblem A

Modellproblem A lautet

$$\begin{aligned} -\Delta u &= (2d - 4\|x\|^2)e^{-\|x\|^2} && \text{in } \Omega = (0,1)^d, \\ u &= e^{-\|x\|^2} && \text{auf } \partial\Omega. \end{aligned}$$

Die exakte Lösung ist

$$u(x) = e^{-\|x\|^2}.$$

Tabelle 4.1: Resultate für Modellproblem A.

Modellproblem A, Q_1, 2d														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8	147		75		113		24		16		10		8	
1/16	562	0.01	282		431		79		35		18		14	
1/32	2113	0.15	1056	0.08	1621	0.06	275	0.03	69		34		25	
1/64	7886	2.18	3939	1.10	6059	0.94	1011	0.43	136	0.03	64	0.03	46	0.01
1/128			14615	16.1			3741	6.42	266	0.18	120	0.22	87	0.10
1/256							13823	115	521	1.94	217	1.89	162	1.23
Modellproblem A, P_1, 2d														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8	218		112		220		51		22		13		13	
1/16	840	0.02	427		854		177		48		26		24	
1/32	3165	0.21	1607	0.11	3230	0.12	645	0.07	98		49		45	
1/64	11820	3.04	6004	1.57	12096	1.74	2403	0.95	193	0.03	95	0.04	88	0.02
1/128							8955	13.9	378	0.24	184	0.30	172	0.20
1/256									739	2.25	359	2.58	336	2.18
Modellproblem A, Q_1, 3d														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8	98	0.01	51		77		18		16		9		8	
1/16	376	0.24	189	0.12	290	0.10	55	0.05	34	0.01	17	0.02	15	0.01
1/32	1416	10.1	708	4.87	1087	4.10	187	1.95	67	0.26	32	0.34	27	0.25
1/64	5287	304.	2641	152.	4063	129.	681	65.6	132	4.43	59	5.86	51	4.18

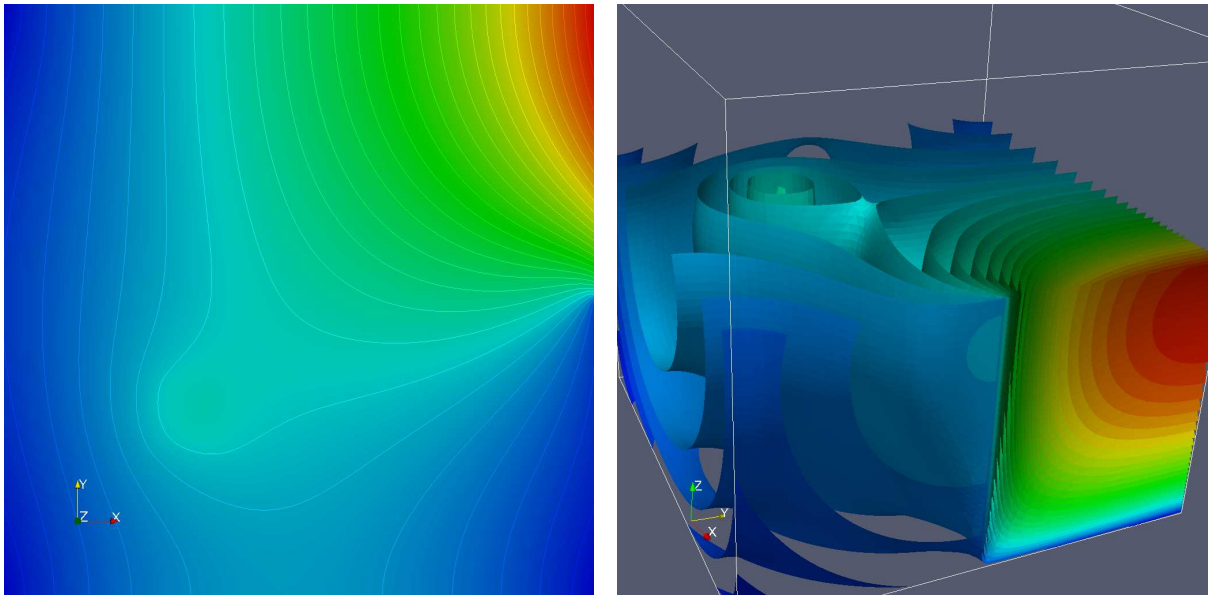


Abbildung 4.2: Lösung des Modellproblems B in 2d und 3d.

4.5.2 Modellproblem B

Modellproblem B lautet

$$\begin{aligned} -\Delta u &= f && \text{in } \Omega = (0,1)^d, \\ u &= g && \text{auf } \Gamma_D, \\ -\nabla u \cdot \nu &= j && \text{auf } \Gamma_N, \end{aligned}$$

mit

$$f(x) = \begin{cases} 50 & 0.25 \leq x_0, x_1 \leq 0.375 \\ 0 & \text{sonst} \end{cases},$$

und

$$\Gamma_N = \{x \,|\, x_1 = 0 \vee x_1 = 1 \vee (x_0 = 1 \wedge x_1 > 1/2)\} \qquad \Gamma_D = \partial\Omega \setminus \Gamma_N,$$

und

$$g(x) = e^{-\|x - x_0\|^2}, \qquad x_0 = (1/2, \ldots, 1/2)^T,$$

sowie

$$j(x) = \begin{cases} -5 & x_0 = 1 \wedge x_1 > 1/2 \\ 0 & \text{sonst} \end{cases}.$$

4.5.3 Modellproblem C

Modellproblem C lautet

$$\begin{aligned} -\nabla \cdot \{k(x)\nabla u\} &= 1 && \text{in } \Omega = (0,1)^d, \\ u &= 0 && \text{auf } \partial\Omega, \end{aligned}$$

mit

$$k(x) = \begin{cases} 20.0 & \lfloor x_0/H \rfloor \text{ gerade, } \lfloor x_1/H \rfloor \text{ gerade, } \lfloor x_2/H \rfloor \text{ gerade} \\ 0.002 & \lfloor x_0/H \rfloor \text{ ungerade, } \lfloor x_1/H \rfloor \text{ gerade, } \lfloor x_2/H \rfloor \text{ gerade} \\ 0.2 & \lfloor x_0/H \rfloor \text{ gerade, } \lfloor x_1/H \rfloor \text{ ungerade, } \lfloor x_2/H \rfloor \text{ gerade} \\ 2000.0 & \lfloor x_0/H \rfloor \text{ ungerade, } \lfloor x_1/H \rfloor \text{ ungerade, } \lfloor x_2/H \rfloor \text{ gerade} \\ 1000.0 & \lfloor x_0/H \rfloor \text{ gerade, } \lfloor x_1/H \rfloor \text{ gerade, } \lfloor x_2/H \rfloor \text{ ungerade} \\ 0.001 & \lfloor x_0/H \rfloor \text{ ungerade, } \lfloor x_1/H \rfloor \text{ gerade, } \lfloor x_2/H \rfloor \text{ ungerade} \\ 0.1 & \lfloor x_0/H \rfloor \text{ gerade, } \lfloor x_1/H \rfloor \text{ ungerade, } \lfloor x_2/H \rfloor \text{ ungerade} \\ 10.0 & \lfloor x_0/H \rfloor \text{ ungerade, } \lfloor x_1/H \rfloor \text{ ungerade, } \lfloor x_2/H \rfloor \text{ ungerade} \end{cases}.$$

4.5.4 Modellproblem D

Modellproblem D lautet

$$\begin{aligned} -\nabla \cdot \{k(x)\nabla u\} &= f && \text{in } \Omega = (0,1)^d, \\ u &= g && \text{auf } \Gamma_D, \\ -\nabla u \cdot \nu &= 0 && \text{auf } \Gamma_N, \end{aligned}$$

Tabelle 4.2: Resultate für Modellproblem B.

Modellproblem A, Q_1, 2d														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8	456		230		424		65		32		14		11	
1/16	1770	0.07	888	0.02	1504	0.01	237		59		24		18	
1/32	6720	0.43	3364	0.21	5436	0.22	877	0.09	112		45		32	
1/64			12614	3.20	19895	3.11	3249	1.28	215	0.04	87	0.04	61	0.02
1/128							12055	18.8	415	0.28	168	0.27	118	0.13
1/256									806	2.88	328	2.63	231	1.71
Modellproblem A, P_1, 2d														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8	667		338		830		138		41		18		16	
1/16	2619	0.04	1327	0.02	2969	0.03	525	0.01	82		35		32	
1/32	10009	0.60	5075	0.32	10778	0.40	2017	0.20	159		68		62	
1/64			19131	4.57			7637	2.81	306	0.05	133	0.05	124	0.04
1/128									590	0.36	259	0.39	244	0.28
1/256									1143	3.45	505	3.47	478	3.08
Modellproblem A, Q_1, 3d														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8	180	0.01	92		176		29		29		12		10	
1/16	694	0.42	349	0.21	596	0.22	95	0.09	54	0.02	22	0.02	19	0.01
1/32	2622	17.6	1313	8.74	2126	7.86	343	3.54	102	0.39	42	0.44	35	0.32
1/64	9813	531.	4908	263.	7747	240.	1269	119.	197	6.42	80	7.70	67	5.40

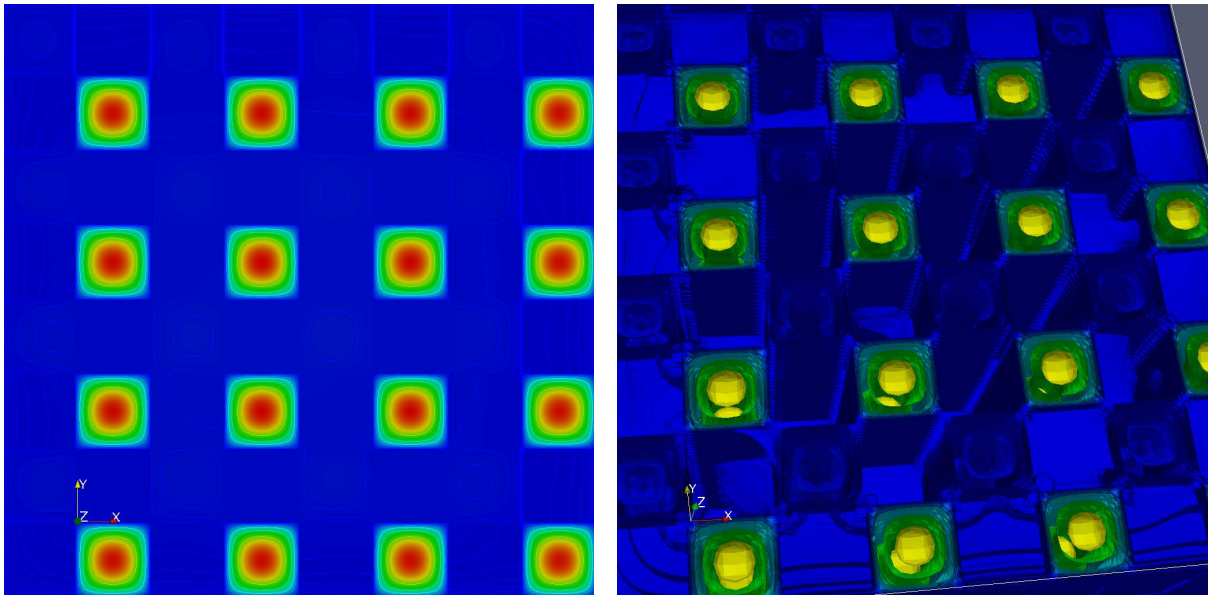


Abbildung 4.3: Lösung des Modellproblems C in 2d und 3d.

Tabelle 4.3: Resultate für Modellproblem C.

Modellproblem C, Q_1, 2d														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8	4665	0.06	2354	0.01	3334	0.01	724		27		17		8	
1/16			13573	0.26			4335	0.12	281		38		27	
1/32							17512	1.91	1761	0.08	73		52	
1/64									8644	1.48	142	0.06	99	0.03
1/128											282	0.49	196	0.22
1/256											577	4.82	405	2.96
Modellproblem C, Q_1, 3d														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8	127	0.01	65		96		22		21		10		8	
1/16	1326	0.83	667	0.42			208	0.20	1179	0.45	32	0.03	23	0.02
1/32	9966	68.2	4996	34.8			1425	14.8	8594	32.9	71	0.76	56	0.51
1/64							8382	792.			151	14.6	124	9.96

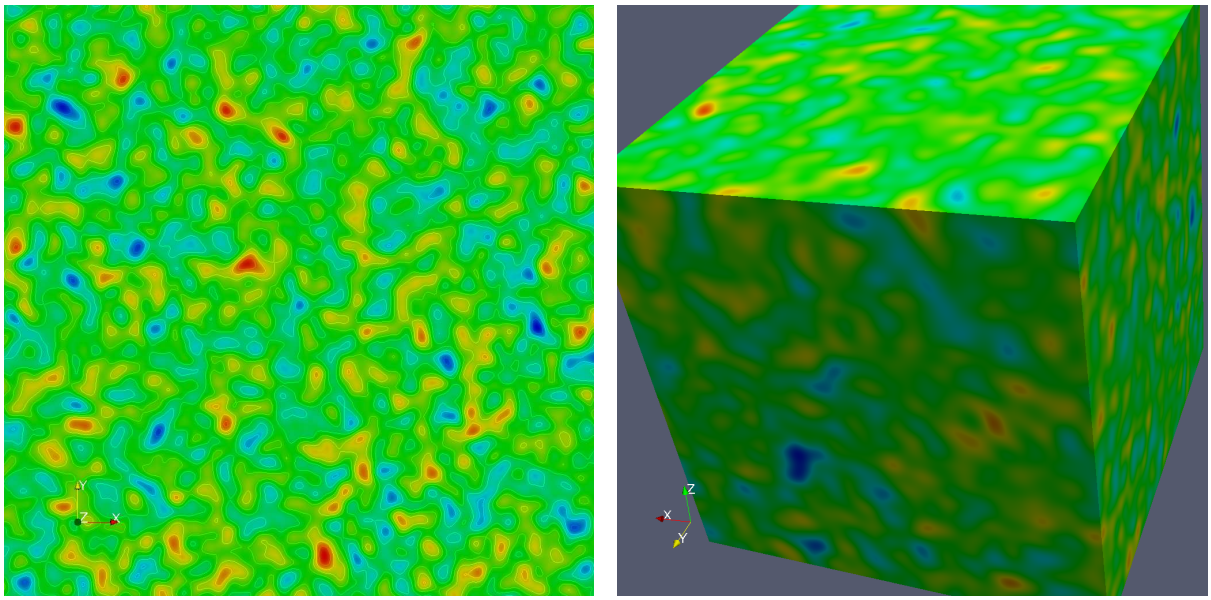


Abbildung 4.4: Log-normalverteilte Permeabilitätsfelder in 2d und 3d.

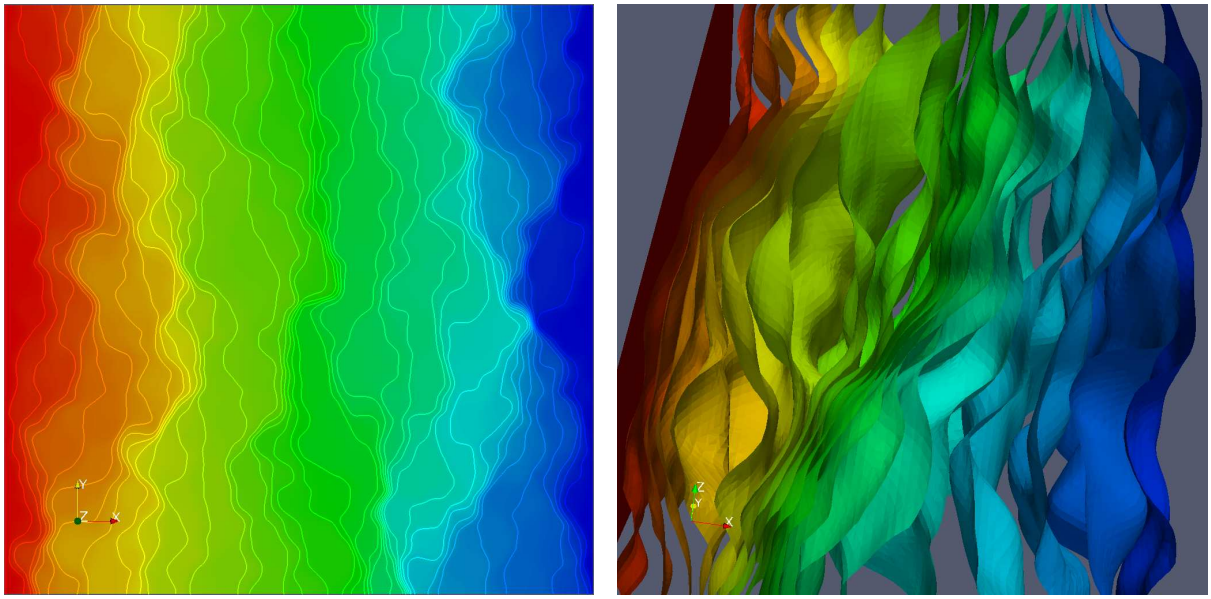


Abbildung 4.5: Lösung des Modellproblems D in 2d und 3d.

Tabelle 4.4: Resultate für Modellproblem D.

Modellproblem D, Q_1, 2d														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/64							11307	4.58	1825	0.31	193	0.08	110	0.03
1/128									5755	3.87	375	0.62	250	0.28
1/256									15489	57.2	707	5.72	492	3.67
1/512										385.	1345	53.6	955	35.2
Modellproblem D, Q_1, 3d														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/16	2538	1.52	1280	0.78			395	0.37	452	0.17	48	0.05	36	0.03
1/32	10096	67.8	5069	34.0			1401	14.6	2190	8.48	88	0.93	73	0.69
1/64			19158	1046			4905	469.	5859	195.	166	16.3	140	11.9

mit

$$\Gamma_D = \{x \,|\, x_0 = 0 \vee x_0 = 1\} \qquad\qquad \Gamma_D = \partial\Omega \setminus \Gamma_N,$$

und

$$g(x) = \begin{cases} 1 & x_0 = 0 \\ 0 & x_0 = 1 \end{cases} .$$

Die Funktion $k(x)$ ist log-normalverteilt mit vorgegebenem Mittelwert, Varianz und Korrelationslänge. Beispiele sieht man in Abbildung 4.4. Die Varianz betrug 3 und Mittelwert war 0, d.h. die Permeabilitäten schwanken zwischen 10^{-3} und 10^3. Als Korrelationslänge wurde 1/64 in 2d und 1/32 in 3d verwendet.

4.5.5 Modellproblem E

$$\begin{aligned} -\nabla \cdot \{K(x)\nabla u\} &= 1 && \text{in } \Omega = (0,1)^d, \\ u &= 0 && \text{auf } \partial\Omega, \end{aligned}$$

mit $K(x)$ einem diagonalen Tensor

$$K_{ij}(x) = \begin{cases} 10^{-6} & i = j = 0 \\ 1 & i = j > 0 \\ 0 & \text{sonst} \end{cases} .$$

Vorsicht: Matrix für Q_1 ist zwar symmetrisch und positiv definit, aber nicht irreduzibel diagonaldominant. Jacobi konvergiert aber nicht für jede s.p.d Matrix ungedämpft, daher die Probleme mit dem Jacobi-Verfahren für Q_1-Elemente. Im 2d-Fall ist die entstehende Matrix nahezu tridiagonal, das ILU_0-Verfahren ist bei richtiger Anordnung exakt für Tridiagonalmatrizen.

Tabelle 4.5: Resultate für Modellproblem E.

Modellproblem E, Q_1, 2d, lexikographische Anordnung														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8			186		524		46		16		20		2	
1/16			756	0.02	2102	0.02	176		75		43		2	
1/32			2949	0.19	8250	0.33	666	0.07	255	0.01	89	0.01	2	
1/64			11423	2.89			2614	1.03	547	0.09	175	0.07	2	
1/128							10102	15.7	1106	0.74	344	0.55	3	
1/256									2188	7.72	664	5.19	3	0.02
Modellproblem E, P_1, 2d, space depth-first Anordnung														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8	233		119		228		50		8		17		9	
1/16	946	0.02	481	0.01	946	0.01	180	0.01	16		36		35	
1/32	3798	0.24	1930	0.12	3834	0.14	638	0.07	32		76	0.01	89	
1/64	15203	3.79	7724	1.94	15422	2.22	2362	0.96	66	0.01	157	0.07	183	0.05
1/128							9020	14.3	173	0.11	318	0.52	373	0.44
1/256									386	1.16	674	4.94	756	4.97
Modellproblem E, Q_1, 3d, lexikographische Anordnung														
h	Jacobi		Gauß-Seidel		Gradient		Grad+SSOR		CG		CG+SSOR		CG+ILU0	
	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit	IT	Zeit
1/8			127	0.01	264	0.01	34		26		18		8	
1/16			505	0.30	1046	0.38	122	0.11	84	0.04	37	0.04	14	0.01
1/32			1952	12.8	4100	15.2	458	4.71	209	0.80	73	0.76	24	0.22
1/64			7582	404.	16014	495.	1796	169.	422	13.8	143	13.8	44	3.72

5 Überlappende Gebietszerlegungsverfahren

5.1 Motivation: Klassische Schwarz-Methode

H.A. Schwarz hat 1890 ein Verfahren vorgestellt, mit dem er die Existenz von Lösungen der Laplace-Gleichung in „komplizierteren“ Gebieten gezeigt hat.

Zu lösen sei

$$\begin{aligned} -\Delta u &= f && \text{in } \Omega \\ u &= g && \text{auf } \partial\Omega \end{aligned}$$

in einem Gebiet $\Omega = \hat{\Omega}_1 \cup \hat{\Omega}_2$ mit $\hat{\Omega}_1 \cap \hat{\Omega}_2 \neq \emptyset$, etwa wie in Abbildung 5.1.

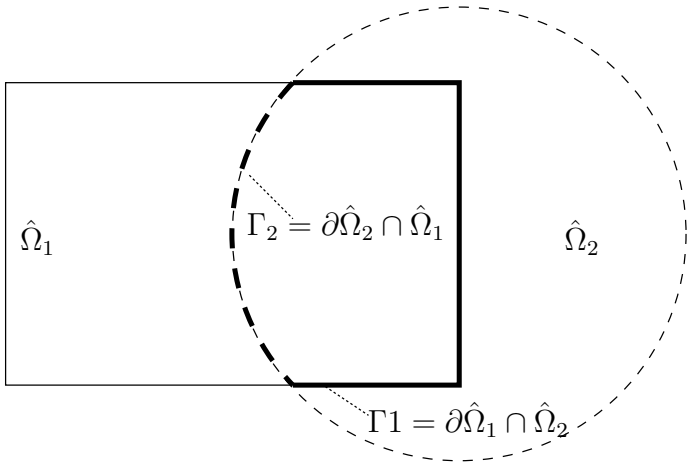


Abbildung 5.1: Zwei überlappende Teilgebiete

Das alternierende Schwarz-Verfahren bestimmt die kontinuierliche Lösung u in ganz Ω durch abwechselndes Lösen in den Teilgebieten $\hat{\Omega}_1$ und $\hat{\Omega}_2$.

Wir definieren (siehe Abb. 5.1):

$$\Gamma_1 = \partial\hat{\Omega}_1 \cap \hat{\Omega}_2, \quad \Gamma_2 = \partial\hat{\Omega}_2 \cap \hat{\Omega}_1$$

und bezeichnen mit

u_i^k	Lösung in $\hat{\Omega}_i$ im Iterationsschritt k
$u_1^k\|_{\Gamma_2}$	u_1^k ausgewertet auf Γ_2,
$u_2^k\|_{\Gamma_1}$	entsprechend

Wir ignorieren im Moment, dass u_i^k eventuell nicht stetig ist).

Dann lautet die alternierende Schwarz-Iteration bei gegebenem Startwert u_0 (in ganz Ω):

for $k = 0, 1, \ldots$ {

Löse erst

$$\begin{cases} -\Delta u_1^{k+1} &= f & \text{in } \hat{\Omega}_1 \\ u_1^{k+1} &= u^k|_{\Gamma_1} & \text{auf } \Gamma_1 \\ u_1^{k+1} &= g & \text{auf } \partial\hat{\Omega}_1 \setminus \Gamma_1 \end{cases} \tag{5.1}$$

und dann

$$\begin{cases} -\Delta u_2^{k+1} &= f & \text{in } \hat{\Omega}_2 \\ u_2^{k+1} &= u^k|_{\Gamma_2} & \text{auf } \Gamma_2 \\ u_2^{k+1} &= g & \text{auf } \partial\hat{\Omega}_2 \setminus \Gamma_2 \end{cases} \tag{5.2}$$

definiere u^{k+1} als

$$u^{k+1}(x,y) = \begin{cases} u_2^{k+1}(x,y) & \text{falls } (x,y) \in \hat{\Omega}_2 \\ u_1^{k+1}(x,y) & \text{sonst} \end{cases} \tag{5.3}$$

}

Man kann zeigen, dass für die entsprechende variationelle Form des Verfahrens

$$\left\|u - u^{k+1}\right\|_{H^1(\Omega)} \leq \rho \left\|u - u^k\right\|_{H^1(\Omega)}$$

gilt, wobei ρ von der Form der Teilgebiete abhängt.

Dazu folgendes

Beispiel 5.1 In *einer* Raumdimension sei $\Omega = (0,1)$, $\hat{\Omega}_1 = (0, \frac{1}{2} + a)$, $\hat{\Omega}_2 = (\frac{1}{2} - a, 1)$, mit $0 < a < \frac{1}{2}$ sowie $f = 0$, $g = 1$, $u^0 = 0$.

Die Iteration lässt sich graphisch durchführen: siehe Abbildung 5.2.

Wir zeigen für dieses Beispiel, dass für den Fehler $e^k = u - u^k$ folgendes Fortpflanzungsgesetz gilt:

$$\left\|e^{k+1}\right\|_\infty = \left(\frac{1-2a}{1+2a}\right)^2 \left\|e^k\right\|_\infty \tag{5.4}$$

Beweis: Laut Zeichnung wird der maximale Fehler jeweils im Punkt $x = \frac{1}{2} - a$ angenommen, also $e^k(\frac{1}{2} - a) = \left\|e^k\right\|_\infty$. Am Rand gilt $e^k(0) = e^k(1) = 0$, dazwischen ist e^k linear.

Bei gegebenem $\left\|e^k\right\|_\infty$ gilt für den Fehler in $\hat{\Omega}_2$:

$$e_2^k(x) = \frac{1-x}{\frac{1}{2}+a} \left\|e^k\right\|_\infty = \begin{cases} 0 & x = 1 \\ \left\|e^k\right\|_\infty & x = \frac{1}{2} - a \end{cases}$$

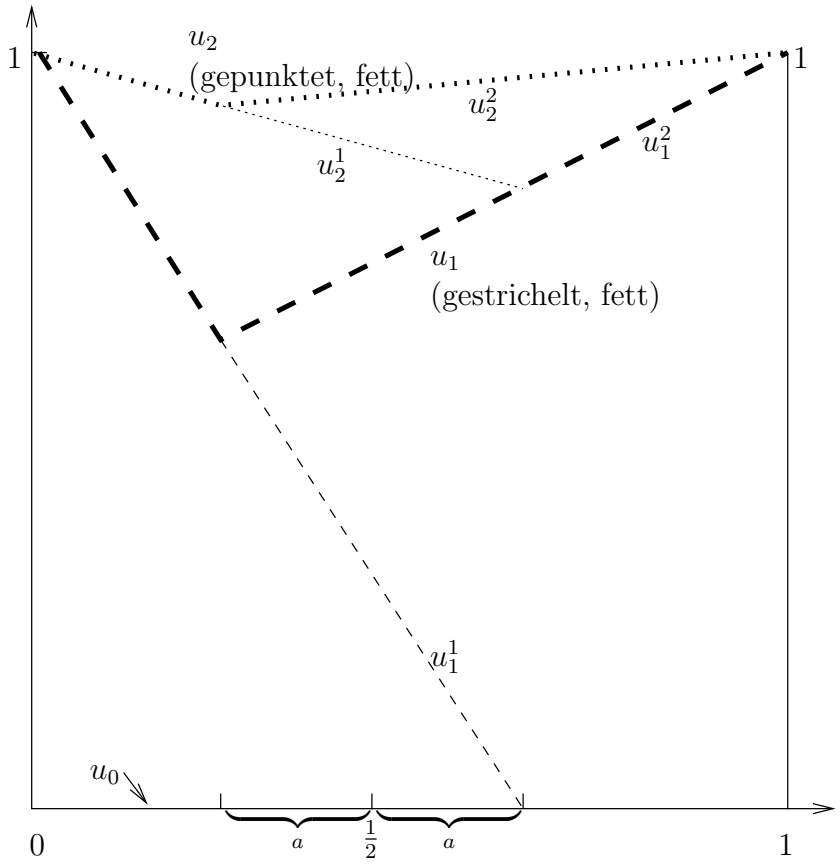


Abbildung 5.2: Alternierende Schwarz-Iteration für Dimension 1

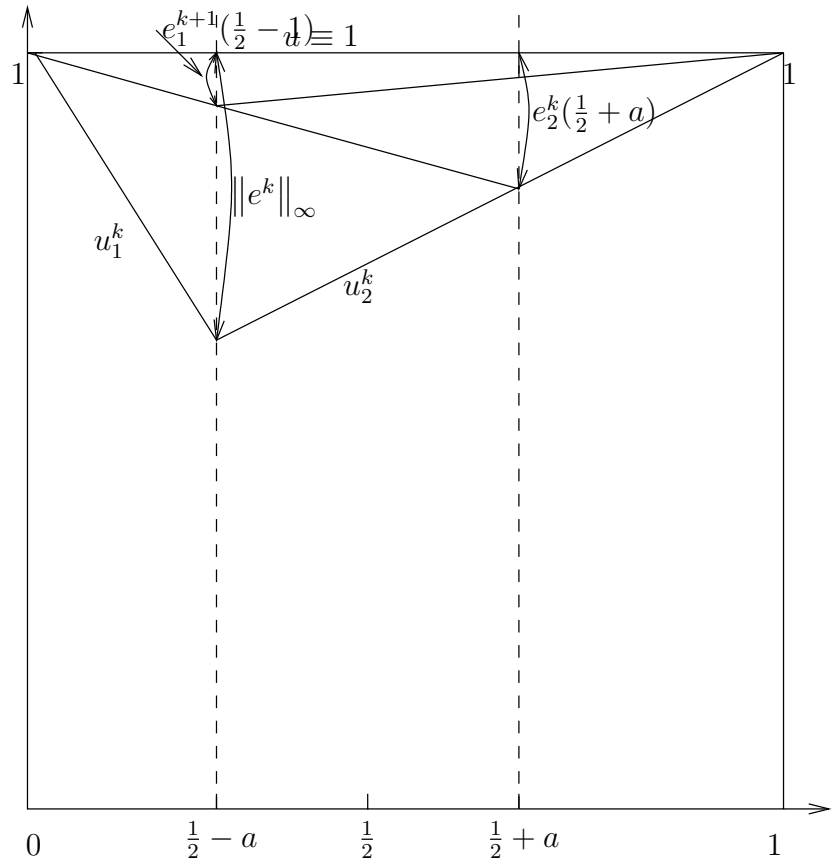


Abbildung 5.3: Fehler der Alternierenden Schwarz-Iteration

Für den Fehler in u_1^{k+1} am Randpunkt $x = \frac{1}{2} + a$ gilt somit

$$e_1^{k+1}(\tfrac{1}{2}+a) = e_2^k(\tfrac{1}{2}+a) = \frac{\frac{1}{2}-a}{\frac{1}{2}+a} \left\|e^k\right\|_\infty$$

Dies ist gleichzeitig der maximale Fehler in u_1^{k+1}, wegen $e_1^{k+1}(0) = 0$ gilt

$$e_1^{k+1} = \frac{x}{\frac{1}{2}+a} \frac{\frac{1}{2}-a}{\frac{1}{2}+a} \left\|e^k\right\|_\infty = \begin{cases} 0 & x = 0 \\ e_2^k(\frac{1}{2}+a) & x = \frac{1}{2}+a. \end{cases}$$

Somit gilt für den Fehler in u_2^{k+1} an der Stelle $x = \frac{1}{2} - a$:

$$\left\|e_\infty^{k+1}\right\| = e_2^{k+1}(\tfrac{1}{2}-a) = e_1^{k+1}(\tfrac{1}{2}-a) = \left(\frac{\frac{1}{2}-a}{\frac{1}{2}+a}\right)^2 \left\|e^k\right\|_\infty .$$

□

5.2 Allgemeine Konstruktion

Nun wollen wir

- die Idee auf $p > 2$ Teilgebiete erweitern
- den Aspekt der Diskretisierung einbringen.

Wir beginnen mit einer systematischen Konstruktion der überlappenden Teilgebiete.

Es sei ein Gebiet $\Omega \subseteq \mathbb{R}^2$ gegeben (wir beschränken uns auf $d = 2$, die Konstruktion geht aber allgemein). Dieses zerlegen wir in *nicht überlappende* Teilgebiete:

$$\Omega_i: \quad \bigcup_{i=1}^{p} \overline{\Omega_i} = \overline{\Omega}, \quad \Omega_i \cap \Omega_j = \emptyset \; \forall i \neq j.$$

Die Teilgebiete seien von dreieckiger oder viereckiger Form wie in Abbildung 5.4. Wir können Ω_i auch als *grobes Gitter* für Ω deuten.

Nun erweitern wir jedes Ω_i um alle Punkte im Abstand $\beta \cdot H$, wobei H die längste Kante eines Ω_i ist;

$$\hat{\Omega}_i = \{x \in \Omega \mid \operatorname{dist}(x, \Omega_i) < \beta \cdot H\}$$

und erhalten somit die *überlappende Zerlegung* von Ω. Die Abbildung 5.5 zeigt diesen Prozess für ausgewählte Teilgebiete. Der Parameter $\beta > 0$ kann frei gewählt werden.

Im Abschnitt 5.1 haben wir die Iteration mit Funktionen $u_i^k \in C^2(\hat{\Omega}_i) \cap C^0\left(\overline{\hat{\Omega}_i}\right)$ durchgeführt (oder $u_i^k \in H_0^1(\hat{\Omega}_i)$, falls man die entsprechende schwache Formulierung bildet).

Um die Idee auf eine numerische Lösung zu übertragen, müssen wir eine Diskretisierung durchführen.

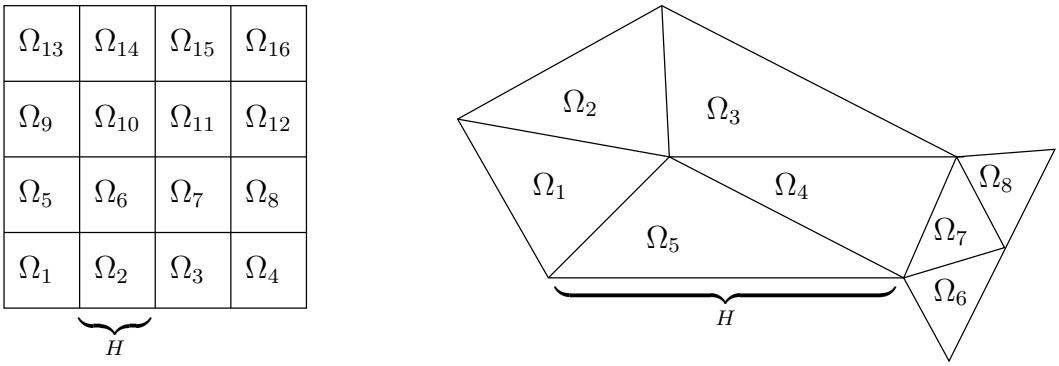


Abbildung 5.4: Zerlegung von Ω in nicht überlappende Teilgebiete Ω_i

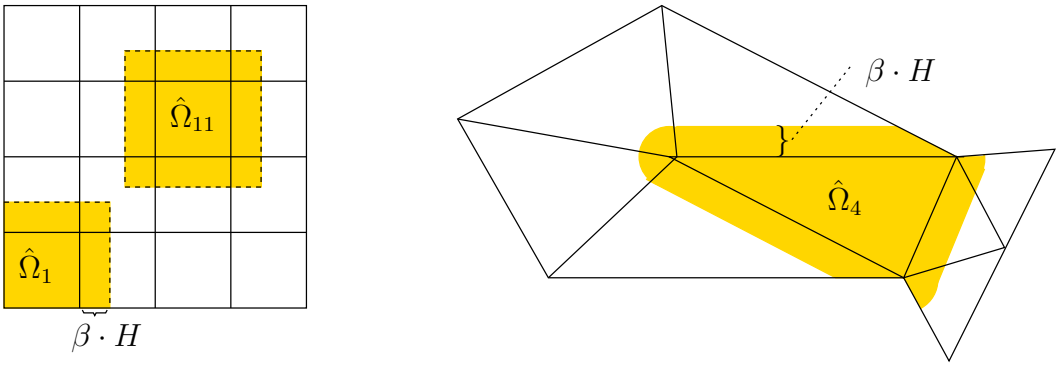


Abbildung 5.5: Konstruktion der überlappenden Zerlegung

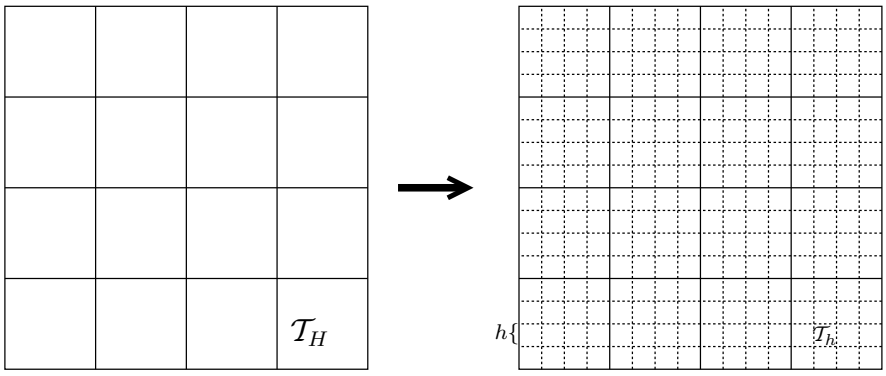


Abbildung 5.6: Konstruktion des feinen Gitters $\mathcal{T}_h$.

Wir benötigen dazu ein Gitter, welches wir durch *regelmäßige Verfeinerung* aus der nichtüberlappenden Zerlegung von Ω erhalten. So ergibt sich das feine Gitter $\mathcal{T}_h$ mit der Gitterweite h. Siehe hierzu Abbildung 5.6

Nun könnten wir die diskrete alternierende Schwarz-Iteration dadurch definieren, dass wir

- alle Teilgebiete in vorgegebener Reihenfolge abarbeiten und
- in jedem Teilgebiet eine neue diskrete Lösung unter Beachtung jeweils neuester Randwerte ausrechnen.

Der Einfachheit halber setzen wir $\beta \cdot H = m \cdot h$ für ein $m \in \mathbb{N}$ voraus.

Wir können diesen Prozess jedoch statt für Funktionen auch auf der Ebene des linearen Gleichungssystemes formulieren.

Dazu sei

I	Indexmenge aller inneren Knoten des feinen Gitters $\mathcal{T}_h$
$Ax = b$	das aus FD- oder FE-Diskretisierung auf $\mathcal{T}_h$ entstandene Gleichungssystem. Es ist $A \in \mathbb{R}^{I \times I}$

Aus der überlappenden Zerlegung in die $\hat{\Omega}_i$ erhalten wir eine überlappende Zerlegung der Indexmenge:

$$\hat{I}_i := \{j \in I \mid \text{Gitterpunkt } (x_j, y_j) \in \hat{\Omega}_i\}.$$

Die Zerlegung ist überlappend, da

$$\underbrace{\overline{\Omega_i} \cap \overline{\Omega_j}}_{\text{Nachbarn}} \neq \emptyset \Rightarrow \hat{I}_i \cap \hat{I}_j \neq \emptyset$$

Es sei $x \in \mathbb{R}^I$ ein beliebiger Vektor. Die Restriktion dieses Vektors auf das Teilgebiet $\hat{\Omega}_i$ besorgt die Matrix $R_i \colon \mathbb{R}^I \to \mathbb{R}^{\hat{I}_i}$

$$(R_i x)_j = (x)_j \quad \forall j \in \hat{I}_i.$$

R_i ist eine Rechteckmatrix mit einer 1 pro Zeile und maximalem Rang. Entsprechend ist $R_i^T \colon \mathbb{R}^{\hat{I}_i} \to \mathbb{R}^I$,

$$(R_i^T x_i) = \begin{cases} (x_i)_j & j \in \hat{I}_i \\ 0 & \text{sonst} \end{cases}$$

die Fortsetzung eines Vektors mit Nullen.

Weiter definieren wir A_i als

$$A_i := R_i A R_i^T = A_{\hat{I}_i, \hat{I}_i}.$$

Beweis: Sei $e_j \in \mathbb{R}^{\hat{I}_i}$, so dass $(e_j)_k = \begin{cases} 1 & k = j \\ 0 & \text{sonst} \end{cases}$. $A_i e_j$ liefert die j-te Spalte von A_i, davon gucken wir die k-te Komponente an ($k \in \hat{I}_i$)

$$(A_i e_j)_k = (R_i A \underbrace{R_i^T}_{\substack{1 \text{ in Komponente } j}})_k = (R_i \underbrace{Az_j}_{\substack{j\text{-te Spalte von } A;\ z_j \in \mathbb{R}^I,\ (z_j)_m = \delta_{jm}}})_k \underset{\text{da } k \in \hat{I}_i}{=} (Az_j)k$$

□

Damit werden wir nun die zweite Varianten der Schwarz-Iteration definieren.

5.3 Multiplikative Schwarz Iteration

Gegeben sei eine beliebige Iterierte $x^{alt} \in \mathbb{R}^I$. Wie üblich gilt die Defektgleichung für $e = x - x^{alt}$:

$$Ae = b - Ax^{alt} \tag{5.5}$$

Statt A nun durch eine andere Matrix gleicher Dimension zu approximieren, wollen wir versuchen, den Fehler nur für die Indizes $\hat{I}_i \subseteq I$ zu berechnen, d.h. wir setzen an

$$e = R_i^T v_i \tag{5.6}$$

mit einer zu bestimmenden Korrektur $v_i \in \mathbb{R}^{\hat{I}_i}$. Einsetzen in (5.5) liefert

$$AR_i^T v_i = b - Ax^{alt}$$

Dieses Gleichungssystem ist überbestimmt ($|I|$ Gleichungen für $\left|\hat{I}_i\right|$ Unbekannte). Multiplizieren wir auf beiden Seiten mit R_i, so ergibt sich ein quadratischens Gleichungssystem:

$$R_i A R_i^T v_i = R_i(b - Ax^{alt}). \tag{5.7}$$

Als zu lösende Matrix erkennen wir $A_i = A_{\hat{I}_i,\hat{I}_i}$ von oben! Als verbesserte Lösung ergibt sich

$$x^{neu} = x^{alt} + R_i^T v_i = x^{alt} + R_i^T A_i^{-1} R_i(b - Ax^{alt}) \tag{5.8}$$

Der Vektor $x^{neu} \in \mathbb{R}^I$ entspricht exakt dem einmaligen Lösen im Teilgebiet $\hat{\Omega}_i$ unter Festhalten der aktuellen Randwerte in der Schwarz-Iteration.

Anwenden der obigen Prozedur auf jedes Teilgebiet ergibt einen Iterationsschritt der multiplikativen Schwarz'schen Methode.

Algorithmus 5.2 (Multiplikativer Schwarz) Mit den Bezeichnungen von oben:

gegeben $x^0 \in \mathbb{R}^I$
for $k = 0, 1, \ldots$
  for $i = 1, 2, \ldots, p$
    $x^{k+\frac{i}{p}} = x^{k+\frac{i-1}{p}} + R_i^T A_i^{-1} R_i(b - Ax^{k+\frac{i-1}{p}})$

Bemerkung 5.3 (Fehlerfortpflanzung im multiplikativen Schwarz) Es gilt für einen Teilschritt

$$e^{k+\frac{i}{p}} = (I - R_i^T A_i^{-1} R_i A) e^{k+\frac{i-1}{p}} \tag{5.9}$$

und somit für eine Iteration:

$$e^{k+1} = (I - P_p) \cdot \cdots \cdot (I - P_2)(I - P_1) e^k, \tag{5.10}$$

wobei $P_i = R_i^T A_i^{-1} R_i A$ abgekürzt wurde.

BEWEIS:

$$\begin{aligned} e^{k+\frac{i}{p}} = x - x^{k+\frac{i}{p}} &= x - x^{k+\frac{i-1}{p}} - R_i^T A_i^{-1} R_i (b - A x^{k+\frac{i-1}{p}}) = \\ &= (x - x^{k+\frac{i-1}{p}}) - R_i^T A_i^{-1} R_i A (x - x^{k+\frac{i-1}{p}}) = \\ &= (I - R_i^T A_i^{-1} R_i A) e^{k+\frac{i-1}{p}}. \end{aligned}$$

Beachte, dass W in der üblichen Formulierung hier nicht invertierbar ist. (5.10) folgt aus p-maliger Anwendung von (5.9). □

Aus (5.10) ergibt sich die Bezeichung „multiplikatives Schwarz-Verfahren".

Bemerkung 5.4 $P_i e$ ist nichts anderes als die Korrektur v_i!

Offensichtlich werden im multiplikativen Schwarz-Verfahren alle Teilkorrekturen v_i sequentiell nacheinander berechnet. In Algorithmus 5.2 benutzt Schritt i das Ergebnis von Schritt $i-1$.

Eine gleichzeitige Berechnung von Teilkorrekturen gelingt mit folgender

Bemerkung 5.5 (unabhängige Teilkorrekturen) Es sei $J \subseteq \{1, \ldots, p\}$ so, dass

$$R_i A R_j^T = 0 \quad \forall i, j \in J, i \neq j \tag{5.11}$$

Dann hängen die Korrekturen v_i, $i \in J$ nicht voneinander ab.

BEWEIS: Betrachte zwei beliebige $i, j \in J$.

$$\begin{aligned} &(I - R_i^T A_i^{-1} R_i A)(I - R_j^T A_j^{-1} R_j A) = \\ &= I - R_j^T A_j^{-1} R_j A - R_i^T A_i^{-1} R_i A + R_i^T A_i^{-1} \underbrace{R_i A R_j^T}_{=0\text{n.v.}} A_j^{-1} R_j A = \\ &= I - P_j - P_i \end{aligned}$$

Mehrfache Anwendung zeigt $\prod_{i \in J}(I - P_i) = I - \sum_{i \in J} P_i$. □

Bedingung (5.11) bedeutet, dass für alle $\eta \in \hat{I}_i$ und $\mu \in \hat{I}_j$ $(A)_{\eta\mu} = 0$ gelten muss (A ist s.p.d.).

Somit besteht J aus den Teilgebieten, „die weit genug auseinander sind". Für unser Beispiel zeigt Abbildung 5.7 eine Zusammenfassung von Teilgebieten, deren Korrekturen sich jeweils gleichzeitig berechnen lassen.

Beachte:

- $\{1, \ldots, p\} = J_1 \cup J_2 \cup J_3 \cup J_4$
- Alle Korrekturen in J_k sind unabhängig voneinander.
- Es genügen immer vier Gruppen für beliebig großes p.
- $\beta < \frac{1}{2}$ muss gelten!
- Die Bearbeitungsreihenfolge – erst J_1, dann J_2, usw. – liefert nicht die identischen Ergebnisse wie $i = 1, 2, \ldots, p$.

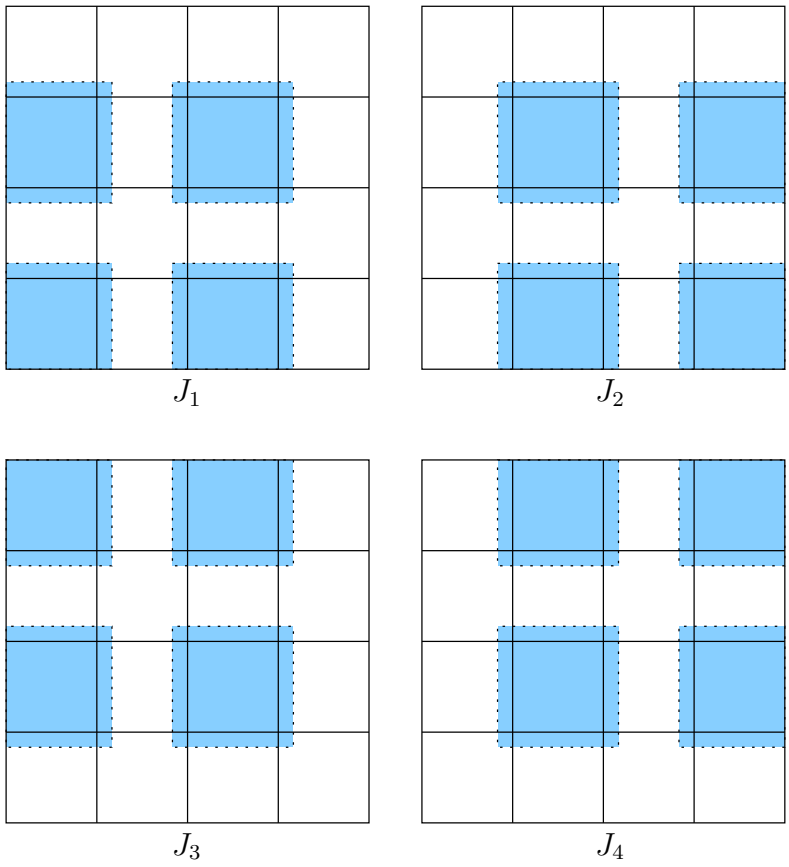


Abbildung 5.7: Gleichzeitig ausführbare Korrekturen im multiplikativen Schwarz

Folgender Satz gibt Auskunft über die Konvergenzeigenschaften des multiplikativen Schwarz-Verfahrens.

Satz 5.6 (Konvergenzrate des mult. Schwarz-Verfahrens) Es gilt die Abschätzung

$$\left\|x - x^{k+1}\right\|_A \leq \rho \left\|x - x^k\right\|_A$$

mit $\rho = 1 - O(H^2)$, $\|x\|_A = \sqrt{x^T A x}$ (A s.p.d.) die Energienorm. Zudem hängt ρ von β und den Koeffizienten k_{ij} der Differentialgleichung ab.

Hingegen hängt ρ *nicht* von h ab.

Beweis: (Smith, Bjørstad und Gropp 1996) oder unten. Vergleiche Satz 5.6 mit dem Beispiel 5.1. □

Für den Fall, dass $\hat{I}_i \cap \hat{I}_j = \emptyset \ \forall i \neq j$ lässt sich die multiplikative Schwarz-Iteration auch als Block-Gauss-Seidel-Ideration schreiben. Vergleiche Abschnitt 4.2.

5.4 Additive Schwarz-Iteration

Statt alle Korrekturen jeweils mit neuesten Werten, kann man auch alle Korrekturen bezüglich x^k berechnen. Dies ergibt die

Algorithmus 5.7 (gedämpfte, additive Schwarz-Iteration) Mit dem Dämpfungsfaktor $\omega \in \mathbb{R}^+$:

gegeben $x^0 \in \mathbb{R}^I$
for $k = 0, 1, 2, \ldots$
$x^{k+1} = x^k + \omega \sum_{i=1}^{p} R_i^T A_i^{-1} R_i (b - Ax^k)$

Für die Fehlerfortpflanzung gilt

Bemerkung 5.8 (Fehlerfortpflanzung im additiven Schwarz) Es gilt

$$e^{k+1} = \left(I - \omega \sum_{i=1}^{p} P_i\right) e^k$$

mit $P_i = R_i^T A_i^{-1} R_i A$ wie oben.

Die Iteration konvergiert nur, wenn ω groß genug gewählt wird, oder man verwendet zusätzlich das Gradientenverfahren. Dies liegt an der mehrfachen Korrektur in den Überlappungsbereichen. Für ω genügt auch die Wahl $\omega = 1/N$ mit N der maximalen Zahl sich überlappender Teilgebiete.

Bezüglich des Konvergenzverhaltens gilt der

Satz 5.9 (Konvergenz der additiven Schwarz-Iteration) Es gilt

$$\kappa\left(\sum_{i=1}^{p} P_i\right) \leq C H^{-2}(1 + \beta^{-1})$$

C ist unabhängig von h und H, hängt aber ab von den Koeffizienten k_{ij} der DGL. Aus dieser Abschätzung folgt, dass das Gradientenverfahren angewandt auf die additive Schwarz-Iteration mit der Rate $\rho = 1 - \frac{1}{\kappa(\sum P_i)}$ konvergiert.

BEWEIS: siehe unten. □

Für den Fall $\hat{I}_i \cup \hat{I}_j = \emptyset \ \forall i \neq j$ entspricht das additive Schwarz'sche Verfahren der Block-Jacobi-Iteration.

5.5 Schwarz-Iteration mit Grobgitterkorrektur

Nach den Sätzen 5.6 und 5.9 hängt die Konvergenzrate des multiplikativen wie auch des additiven Schwarz-Verfahrens von H ab. Mit steigender Prozessorzahl ($p = H^{-2}$) werden immer mehr Iterationen benötigt.

Um diesen Effekt zu verstehen, betrachte ein Teilgebiet Ω_i mit $\partial\Omega_i \cap \partial\Omega = \emptyset$ und den speziellen Fehlervektor $\tilde{e}_i \in \mathbb{R}^I$

$$(\tilde{e}_i)_j = \begin{cases} 1 & \text{Gitterpunkt } (x_j, y_j) \in \overline{\hat{\Omega}_i} \\ \text{beliebig} & \text{sonst} \end{cases}$$

Da A für innere Gitterpunkte die Zeilensumme 0 hat, gilt

$$R_i A \tilde{e}_i = 0.$$

Somit ist auch die für diesen Fehler im Teilgebiet $\hat{\Omega}_i$ berechnete Korrektur $v_i = 0$ (denn $v_i = \tilde{A}_i^{-1} R_i A \tilde{e}_i$): obwohl $\tilde{e}_i = 1$ im Teilgebiet $\hat{\Omega}_i$! Dies gilt allgemein für „glatte" Fehler: Fehler, die in einem Teilgebiet annähernd konstant oder linear sind, werden durch die Teilgebietskorrektur kaum erfasst.

Abhilfe schafft eine *Grobgitterkorrektur*, die wie folgt konstruiert wird. Wir nutzen die zweistufige Gitterkonstruktion aus Abbildung 5.6. Es sei $I_H \subset I$ die Indexmenge der Knoten des feinen Gitters, die auch im groben Gitter vorkommen (Abbildung 5.8).

Die Grobgitterkorrektur v_H wird in $\mathbb{R}^{I_H}$ berechnet. Es sei für $i \in I_H$ der *Einheitsvektor* z_i, also $(z_i)_j = \delta_{ij}$, $i, j \in I_H$. Wie immer bezeichne (x_i, y_i) die Koordinaten des Gitterpunktes $i \in I$. Damit definieren wir die *Prolongation* $R_H^T \colon \mathbb{R}^{I_H} \to \mathbb{R}^I$ mittels

$$(R_H^T z_i)_j = \begin{cases} \left(1 - \frac{|x_j - x_i|}{H}\right)\left(1 - \frac{|y_j - y_i|}{H}\right) & |x_j - x_i|, |y_j - y_i| < H \\ 0 & \text{sonst} \end{cases}$$

Da $v_H = \sum_{i \in I_H} (v_H)_i \, z_i$, ist hiermit die Prolongation für jedes beliebige v_H beschrieben.

Gegeben eine Näherungslösung $x^{alt} \in \mathbb{R}^I$ für $Ax = b$, wird die Grobgitterkorrektur durch Lösen des Gleichungssystems

$$A_H v_H = R_H (b - A x^{alt})$$

mit

$$A_H = R_H A R_H^T$$

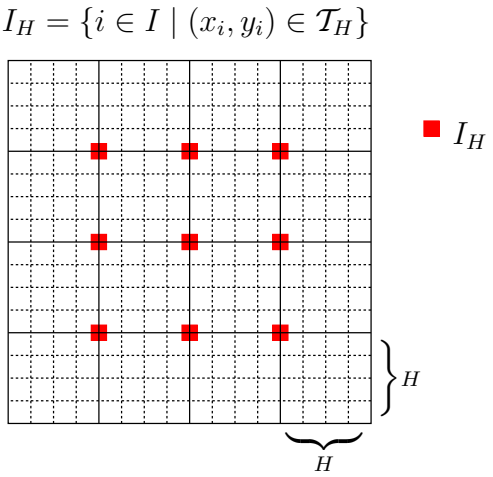


Abbildung 5.8: Zur Definition von I_H

bestimmt. Das Besetztheitsmuster der Matrix A_H ist in diesem Fall ein 9-Punkte-Stern auf dem groben Gitter. Eine weitere Charakterisierung von A_H wird unten gegeben werden.

Als verbesserte Näherung ergibt sich dann

$$x^{neu} = x^{alt} + R_H^T A_H^{-1} R_H (b - A x^{alt}). \tag{5.12}$$

Da (5.12) formal so aussieht wie (5.8) verwenden wir statt dem Index H den Index 0, d.h.

$$R_0 := R_H, \quad A_0 := A_H, \quad \hat{I}_0 = I_H,$$

und erhalten

Algorithmus 5.10 (Multiplikativer Schwarz mit Grobgitterkorrektur) Gegeben $x^0 \in \mathbb{R}^I$.

for $k = 0, 1, \ldots$
  for $i = \mathbf{0}, 1, 2, \ldots, p$
    $x^{k+\frac{i+1}{p}} = x^{k+\frac{i}{p}} + R_i^T A_i^{-1} R_i (b - A x^{k+\frac{i}{p}})$

Algorithmus 5.11 (Additiver Schwarz mit Grobgitterkorrektur) Gegeben $x^0 \in \mathbb{R}^I$.

for $k = 0, 1, 2, \ldots$
  $x^{k+1} = x^k + \omega \sum_{i=\mathbf{0}}^{p} R_i^T A_i^{-1} R_i (b - A x^k)$

Die Konvergenzrate beider Verfahren ist nun unabhängig von H und h, aber weiterhin abhängig von β und den Koeffizienten k_{ij}.

5.6 Hinweise zur praktischen Implementierung

Datenverteilung

Die Vektoren x, b und die Matrix A sind auf die Speicher der (nachrichtengekoppelten) Prozessoren zu verteilen.

Wir gehen aus von der allgemeinen Konstruktion mit $\mathcal{T}_H, \mathcal{T}_h$, nichtüberlappenden Teilgebieten Ω_i, überlappenden Teilgebieten $\hat{\Omega}_i$, $\beta = m \cdot h$ mit $m \in \mathbb{N}$ und den Indexmengen I, I_i, $\hat{I}_i$.

Wenden wir uns zunächst dem Schwarz-Verfahren *ohne* Grobgitterkorrektur zu. Prozessor $i \in \{1, \ldots, p\}$ soll die Korrektur in $\hat{\Omega}_i$ berechnen. Wegen

$$d_i = R_i(b - Ax^{alt})$$

benötigt er dazu alle $(x^{alt})_j$ mit

$$j \in \tilde{I}_i = \{j \in I \mid \text{Gitterpunkt } (x_j, y_j) \in \overline{\hat{\Omega}_i}\}$$

Beachte, dass $\hat{I}_i$ mittels $\hat{\Omega}_i$ und $\tilde{I}_i$ mittels $\overline{\hat{\Omega}_i}$ definiert ist. Es gilt also

$$I_i \subset \hat{I}_i \subset \tilde{I}_i \subset I$$

(wegen $\Omega_i \subset \hat{\Omega}_i \subset \overline{\hat{\Omega}_i}$). $\tilde{I}_i$ unterscheidet sich von $\hat{I}_i$ durch Hinzunahme der Randknoten.

Ein Prozessor $i \in \{1, \ldots, p\}$ speichert folgende Daten:

$$\begin{array}{ll} (x^k)_\alpha & \alpha \in \tilde{I}_i \\ (b)_\alpha & \alpha \in \hat{I}_i \\ (A)_{\alpha\beta} & \alpha \in \hat{I}_i, \beta \in \tilde{I}_i. \end{array}$$

Damit erhalten wir folgende Implementierung für das additive Schwarz-Verfahren:

Algorithmus 5.12 (Implementierung des additiven Schwarz) Gegeben $x^0 \in \mathbb{R}^I$.

Setze $(x_i^0)_\alpha = (x^0)_\alpha$, $\alpha \in \tilde{I}_i$, $(b_i)_\alpha = (b)_\alpha$, $\alpha \in \hat{I}_i$, $\quad i \in \{1, \ldots, p\}$.
In Prozessor $i \in \{1, \ldots, p\}$:
  for $k = 0, 1, 2, \ldots$
    $(d_i)_\alpha = (b)_\alpha - \sum_{\beta \in \tilde{I}_i} (A)_{\alpha\beta}(x_i^k)_\beta, \quad \alpha \in \hat{I}_i$
    $v_i = A_i^{-1} d_i, \quad A_i \colon \mathbb{R}^{\hat{I}_i} \to \mathbb{R}^{\hat{I}_i}$
    $(v_i)_\alpha = 0 \quad \forall \alpha \in \tilde{I}_i \setminus \hat{I}_i$
    $(v_i)_\alpha = (v_i)_\alpha + \sum_{j \in \{l \neq i \mid \alpha \in \hat{I}_l\}} (v_j)_\alpha, \quad \alpha \in \tilde{I}_i$ // Kommunikation
    $(x_i^{k+1})_\alpha = (x_i^k)_\alpha + \omega(v_i)_\alpha \quad \alpha \in \tilde{I}_i$

Inexakte Teilgebietslöser

Bisher haben wir nicht gesagt, wie die Teilgebietsprobleme A_i gelöst werden. Prinzipiell gilt hier dasselbe wir für A selbst, da A_i die Diskretisierung der Differentialgleichung im Teilgebiet $\hat{\Omega}_i$ ist (schwachbesetzt, Kondition $O(h^{-2})$).

Bei der Schwarz-Iteration geht man von einer *exakten* Lösung der Teilgebietsprobleme aus. Iterative, inexakte Lösung ist möglich und führt auf sekundäre Iterationen [(Hackbusch 1991), Lemma 11.3.5]. Die Konvergenz der äusseren Iteration (Schwarz-Verfahren) hängt dann von der Konvergenzrate der inneren Iteration (Teilgebietslöser, z.B. Gauß-Seidel) ab. In der Praxis verwendet man z.B. Mehrgitter als Teilgebietsiteration.

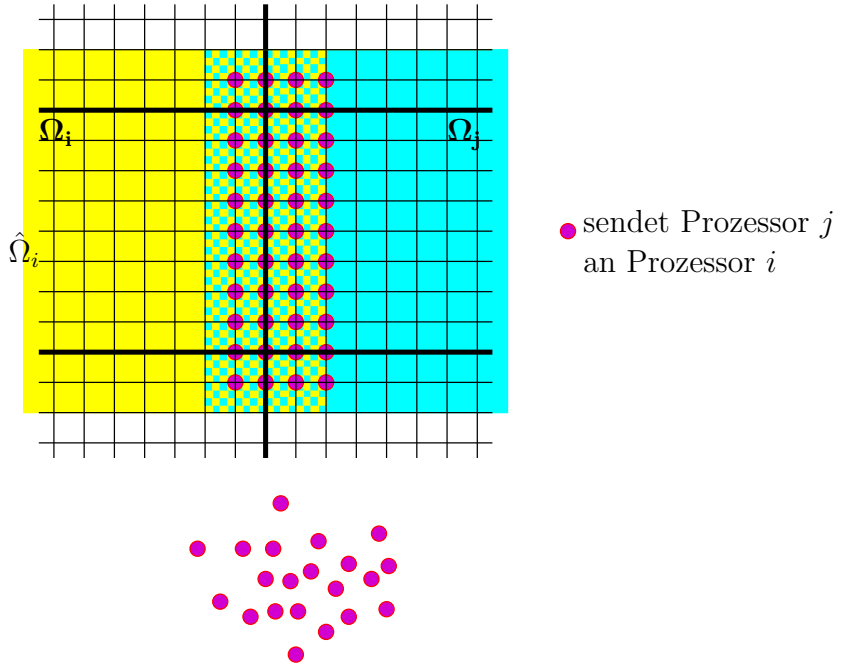


Abbildung 5.9: Kommunikation der Korrekturen im Überlappungsbereich

Praktische LÃ¶sung des Grobgitterproblems

Es werden drei Möglichkeiten diskutiert:

1. Jeder berechnet den Teil von $R_H(b - Ax^k)$ in seinem Teilgebiet, Aufsammeln von d_H in einem Prozessor (alle-an-einen, sequentielle Lösung des Grobgitterproblems, Austeilen der Grobgitterkorrektur (einer-an-alle), verteilte Prolongation.

2. Wie 1, nur dass d_H auf allen Prozessoren gesammelt wird (alle-an-alle-Kommunikation) und jeder das Grobgitterproblem löst. Ist im Prinzip nur alle-an-alle gegen alle-an-einen, einer-an-alle.

3. Verteilte Lösung des Grobgitterproblems mittels eines Iterationverfahrens oder paralleler Gauß-Elimination. Problem: sehr wenige Unbekannte pro Prozessor, nicht sehr effizient.

Reduktion der sequentiellen Komplexität

Wir wenden uns nun Fragen des Rechenaufwandes zu. Zunächst zeigen wir, dass das Schwarz-Verfahren den sequentiellen Aufwand zur Lösung eines Gleichungssystems reduzieren kann.

Dazu bertrachten wir $d = 2$ und ein strukturiertes Gitter mit $h = 1/n$, also n^2 Unbekannten.

Das zu lösende Gleichungssystem hat Bandstruktur, eine exakte Gauß-Elimination, die dies ausnutzt hat Aufwand $O(n^4)$.

Wir verwenden diesen Löser nun als *Teilgebiets- und Grobgitterlöser* im Schwarz-Verfahren. Der Aufwand für *eine* Schwarz-Iteration ist (sequentiell, mit $n_H = 1/H$)

$$T_s(n) = \underbrace{n_H^4}_{\substack{\text{Grob-}\\\text{gitter-}\\\text{problem}}} + \underbrace{n_H^2}_{\substack{\#\text{ Teil-}\\\text{gebiete}}} \cdot \underbrace{\left(\frac{n}{n_H}\right)^4}_{\substack{\text{ein}\\\text{Teilgebiets-}\\\text{problem}}} = n_H^4 + n_H^{-2} n^4$$

Minimieren der Komplexität durch optimale Wahl von n_H:

$$\begin{aligned} &\frac{d}{dn_H}(n_H^4 + n_H^{-2} n^4) = 4n_H^3 - 2n_H^{-3} n^4 = n_H^{-3}(4n_H^6 - 2n^4) \stackrel{!}{=} 0 \\ \iff \quad & 4n_H^6 = 2n^4 \\ \iff \quad & n_H = \left(\tfrac{1}{2}\right)^{\frac{1}{6}} n^{\frac{2}{3}} \end{aligned}$$

Für dieses optimale n_H ergibt sich die Komplexität

$$T_s(n) = n_{H,opt}^4 + n_{H,opt}^{-2} n^4 \approx \boxed{n^{\frac{8}{3}}}$$

Wegen der H, h-Unabhängigkeit der Konvergenz gilt: Das Gleichungssystem lässt sich bis auf Fehler ϵ in $O(n^{\frac{8}{3}})$ Operationen lösen!

Optimales Grobgitter im parallelen Fall

Wir fragen nun, wie groß das grobe Gitter im parallelen Fall gewählt werden sollte. Ausgangspunkt ist:

- Als Teilgebiets- und Grobgitterlöser wird ein Verfahren mit Aufwand n^α für ein Gitter der Größe n^d verwendet
- Jeder Prozessor bearbeitet ein Teilgebiet
- Für das grobe Gitter steht ein separater Prozessor zur Verfügung
- Teil- und Grobgitter können gleichzeitig bearbeitet werden (additives Verfahren)

Die parallele Komplexität ist

$$T_p(n) = \max\left\{ n_H^\alpha, \left(\frac{n}{n_H}\right)^\alpha \right\}$$

optimal ist also

$$n_H^\alpha = \left(\frac{n}{n_H}\right)^\alpha \quad \iff \quad n_H^{2\alpha} = n^\alpha \quad \iff \quad \boxed{n_H = \sqrt{n}}.$$

Die Laufzeit für dieses optimale n_H ist

$$T_p(n) = n^{\frac{\alpha}{2}} \quad \text{bei } n_H^d = n^{\frac{d}{2}} \text{ Prozessoren.}$$

Für gegebenes n haben wir also die minimale Laufzeit (und damit die optimale Prozessorzahl) bestimmt.

Die Wahl $n_H = \sqrt{n}$ ist unabhängig von der Komplexität des zugrundeliegenden Lösungsverfahrens (d.h. von α).

Kommunikation und Überlappung β wurden nicht berücksichtigt.

Speedup einer Iteration

Wir untersuchen nun den Einfluss des Lösers und der Überlappungsbreite auf den Speedup. Ausgangspunkt ist:

$$S(n,p) = \frac{T_{seq}(n)}{T_{par}(n,p)}$$

- Wir betrachten additiven Schwarz *ohne* Grobgitterkorrektur sowohl für T_{seq} als auch für T_{par},
- allgemein in d Raumdimensionen
- der Teilgebiets- und Grobgitterlöser hat den Aufwand n^α für n^d Gitterpunkte ($\alpha \geq d$).

Unter diesen Voraussetzungen gilt (beachte Abbildung 5.10)

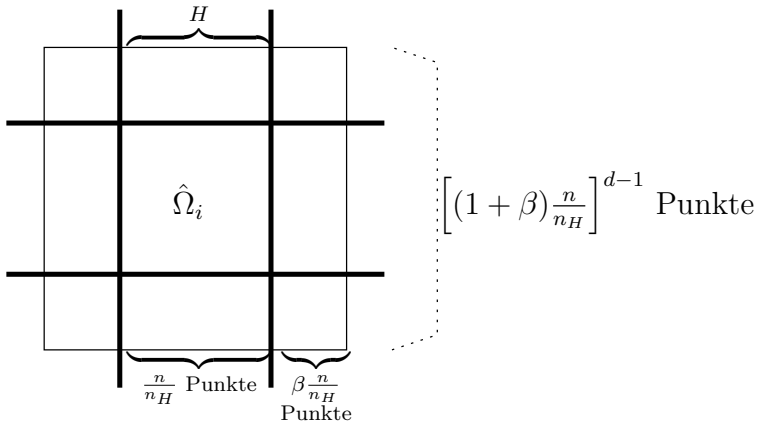


Abbildung 5.10: Illustration zum Speedup des additiven Schwarz

$$\begin{aligned} S(n,p) &= \frac{\left[\frac{n}{n_H}(1+\beta)\right]^\alpha \cdot p \cdot t_f}{\left[\frac{n}{n_H}(1+\beta)\right]^\alpha \cdot t_f + \left\{\beta \cdot \underbrace{\frac{n}{n_H} \cdot \left[(1+\beta)\frac{n}{n_H}\right]^{d-1}}_{\left(\frac{n}{n_H}\right)^d}\right\} \cdot t_w} = \\ &= \frac{p}{1 + \left(\frac{n}{n_H}\right)^{d-\alpha} \cdot \frac{\beta}{1+\beta}(1+\beta)^{d-\alpha} \cdot \frac{t_w}{t_f}}. \end{aligned}$$

Wir unterscheiden zwei Fälle:

$\alpha > d$ (nicht optimaler Löser). In diesem Fall gilt

$$\lim_{n\to\infty} S(n,p) = p,$$

weil dann $\left(\frac{n}{n_H}\right)^{d-\alpha} \to 0$ für $n \to \infty$ mit festem n_H.

$\alpha = d$ (optimaler Löser, z.B. Mehrgitter). Hier hat man

$$S(n,p) = \frac{p}{1 + \frac{\beta}{1+\beta} \cdot \frac{t_w}{t_f}}$$

unabhängig von n. Hier wächst die Kommunikation im selben Maße wie der Rechenaufwand!

Skalierbarkeit

Wie verhält sich das Verfahren bei konstanter Zahl von Unbekannten pro Prozessor, d.h. $\frac{n}{n_H} = K$, d.h. $n = Kp^{\frac{1}{d}\leftarrow\text{geändert}}$, und $p \to \infty$.

Ohne Grobgitter gilt laut dem letzten Abschnitt:

$$S(K \cdot p^{\frac{1}{d}}, p) = \frac{p}{1 + \left(\frac{Kp^{\frac{1}{d}}}{p^{\frac{1}{d}}}\right)^{d-\alpha} \cdot \beta(1+\beta)^{d-1-\alpha} \cdot \frac{t_w}{t_f}} = \frac{p}{1 + K^{d-\alpha} \cdot \beta \cdot (1+\beta)^{d-1-\alpha} \frac{t_w}{t_f}}.$$

also

$$\lim_{p\to\infty} S(Kp^{\frac{1}{d}}, p) = \begin{cases} p & \alpha > d \\ \frac{p}{1+\frac{\beta}{1+\beta}\frac{t_w}{t_f}} & \alpha = d \end{cases}$$

Mit Grobgitter wächst der Aufwand zur *sequentiellen* Lösung des Grobgitterproblems wie $n_H^\alpha = p^{\frac{\alpha}{d}}$ (der Aufwand für die Teilgebietsprobleme bleibt konstant K^α). Für große p wird das Grobgitterproblem zum Flaschenhals und muss verteilt gelöst werden.

6 Abstrakte Schwarz-Theorie

24.4.09
1

6 Abstrakte Schwarz-Theorie

entwickelt seit den 1980er Jahren. „Framework" zur Analyse „aller" Methoden.

- Das vorkonditionierte Gradienten-/CG-Verfahren konvergiert für jeden symmetrisch positiv definiten Vorkonditionierer

$$BAx = Bb.$$

Die Konvergenz(geschwindigkeit) wird durch $\kappa(BA) = \frac{\lambda_{max}(BA)}{\lambda_{min}(BA)}$ charakterisiert.

- Multiplikative Verfahren können auch ohne Abstiegsverfahren eingesetzt werden, dann ist $\varrho(\prod_{i=0}^{\ell}(I-P_i))$ die relevante Größe.

○ Wir behandeln nur den additiven Fall.

Wie bestimmt man $\kappa(C)$ für C s.p.d.?

Lemma 6.1 (Raleigh Quotienten)

Sei C eine s.p.d. Matrix, dann gilt

$$\lambda_{min}(C) = \min_{x\neq 0} \frac{\langle Cx,x\rangle}{\langle x,x\rangle}, \quad \lambda_{max}(C) = \max_{x\neq 0} \frac{\langle Cx,x\rangle}{\langle x,x\rangle}.$$

$\langle\cdot,\cdot\rangle$: euklidisches Skalarprodukt, $\|\cdot\|$ euklidische Norm.

Beweis: Zu C s.p.d. gibt es (genau dann) ein unitäres Q ($Q^TQ=I$) mit $Q^TDQ=C$, $D=\mathrm{diag}\{\lambda_1,\dots,\lambda_n\}\subset\mathbb{R}^+$ (Hackbusch, Satz 2.8.7).

○

$$\text{Somit } \min_{x\neq 0}\frac{\langle Cx,x\rangle}{\langle x,x\rangle} = \min_{Qy\neq 0}\frac{\langle CQy,Qy\rangle}{\langle Qy,Qy\rangle} = \min_{y\neq 0}\frac{\langle Q^TCQy,y\rangle}{\langle y,y\rangle}$$

$$\overset{\text{Unitärer Invarianz}}{=} \min_{x\neq 0}\frac{\langle Dx,x\rangle}{\langle x,x\rangle} = \min_{\|x\|=1}\sum_{i=1}^{n}\lambda_i x_i^2 = \lambda_{min}(C)$$

alles in die λ_{min}-Komponente stecken.

Ebenso für $\max_{x\neq 0}\frac{\langle Cx,x\rangle}{\langle x,x\rangle} = \lambda_{max}(C)$ ☒

Korollar 6.2 Für A,B s.p.d. gilt

$$\lambda_{min}(BA) = \min_{x}\frac{\langle BAx,x\rangle}{\langle x,x\rangle}, \quad \lambda_{max}(BA) = \max_{x\neq 0}\frac{\langle BAx,x\rangle}{\langle x,x\rangle}$$

Beweis: BA ist nicht notwendig symmetrisch, aber $\sigma(BA) = \sigma(A^{1/2}BA^{1/2})$ und $C := A^{1/2}BA^{1/2}$ ist dann symmetrisch positiv definit. ☒

Korollar 6.3 In der Raleigh-Quotienten darf das euklidische Skalarprodukt durch ein beliebiges Skalarprodukt $\langle x,y\rangle_M = x^T M y$ ersetzt werden.

Beweis:

$$\min_{x\neq 0} \frac{\langle Cx,x\rangle}{\langle x,x\rangle} = \min_{0\neq x = M^{1/2}y} \frac{\langle CM^{1/2}y, M^{1/2}y\rangle}{\langle M^{1/2}y, M^{1/2}y\rangle}$$

$$\sigma(C) = \sigma(M^{1/2}CM^{-1/2}) \quad \min_{y\neq 0} \frac{\langle \overbrace{M^{1/2}CM^{-1/2}}^{=I} M^{1/2}y, M^{1/2}y\rangle}{\langle My, y\rangle}$$

$$= \min_{y\neq 0} \frac{\langle Cy,y\rangle_M}{\langle y,y\rangle_M}. \qquad \square$$

Im allgemeinen genügen uns Abschätzungen der Eigenwerte.

Lemma 6.4 A, C seien symmetrisch positiv definite Matrizen. Aus den beiden Abschätzungen (nach Korollar 6.3 ist das Skalarprodukt egal).

$$\gamma\langle x,x\rangle_A \leq \langle Cx,x\rangle_A \leq \Gamma\langle x,x\rangle_A \qquad \forall x\in\mathbb{R}^I \quad (*)$$

folgt

$$\lambda_{\min}(C) \geq \gamma, \quad \lambda_{\max}(C) \leq \Gamma \text{ und } \kappa(C) \leq \frac{\Gamma}{\gamma}.$$

Beweis
Aus $(*)$ (links) folgt $\gamma \leq \frac{\langle Cx,x\rangle_A}{\langle x,x\rangle_A}$ $\forall x\neq 0$, also $\gamma \leq \lambda_{\min}(C) = \min_{x\neq 0} \frac{\langle Cx,x\rangle_A}{\langle x,x\rangle}$

ebenso $\frac{\langle Cx,x\rangle_A}{\langle x,x\rangle} \leq \Gamma$ $\forall x\neq 0$, also $\lambda_{\max}(C) \leq \Gamma$.

nun ist $\kappa(C) = \frac{\lambda_{\max}(C)}{\lambda_{\min}(C)} \leq \frac{\Gamma}{\gamma}$. $\square$

24.4.09
3

Die gedämpfte additive Schwarz-Iteration lautet

$$x^{k+1} = x^k + \omega \underbrace{\sum_{i=0}^{p} R_i^T A_i^{-1} R_i}_{=:B} (b - Ax^k)$$

mit $A_i = R_i A R_i^T$.

Wir betrachten hier nur <u>exakte</u> Teilgebietslöser. Inexakte Teilgebietslöser $\tilde{A}_i \approx A_i$ können ebenfalls ohne viel Aufwand behandelt werden (Toselli/Widlund 2004).

Zu analysieren ist nun $\kappa(BA) = \kappa\left(\sum_{i=0}^{p} \underbrace{R_i^T A_i^{-1} R_i A}_{P_i}\right) = \kappa\left(\sum_{i=0}^{p} P_i\right)$.

Nach Lemma 6.4. sind möglichst gute Zahlen γ, Γ in der Abschätzung

$$\gamma \langle x, x\rangle_A \leq \left\langle \sum_{i=0}^{p} P_i x, x \right\rangle_A \leq \Gamma \langle x, x\rangle_A$$

zu finden.
Es zeigt sich, dass das Skalarprodukt bezüglich der Matrix A besonders vorteilhaft ist.

Zur Analyse des additiven Verfahrens mit exakten Teilgebietslösern genügen die beiden folgenden Voraussetzungen.

<u>Voraussetzung A1 (Stabile Zerlegung)</u>

Es gibt eine Zahl $C_0 > 0$ so dass zu jedem $x \in \mathbb{R}^I$ eine Zerlegung $x = \sum_{i=0}^{p} R_i^T x_i$ mit $x_i \in \mathbb{R}^{\hat{I}_i}$ existiert für die gilt

$$\sum_{i=0}^{p} \langle R_i^T x_i, R_i^T x_i\rangle_A \leq C_0 \langle x, x\rangle_A .$$

<u>Voraussetzung A2 (Verschärfte Cauchy-Schwarz-Ungleichung)</u>

Es gibt Konstanten $0 \leq \varepsilon_{ij} \leq 1$, $1 \leq i, j \leq p$ so dass

$$\left|\langle R_i^T x_i, R_j^T x_j\rangle_A\right| \leq \varepsilon_{ij} \langle R_i^T x_i, R_i^T x_i\rangle_A^{1/2} \langle R_j^T x_j, R_j^T x_j\rangle_A^{1/2} \quad \forall x_i \in \mathbb{R}^{\hat{I}_i},\ x_j \in \mathbb{R}^{\hat{I}_j}.$$

$E \in \mathbb{R}^{p \times p}$ sei die Matrix mit $(E)_{ij} = \varepsilon_{ij}$ und $\rho(E)$ ihr Spektralradius.

Vorsicht: E beinhaltet nicht $\mathbb{R}^{\hat{I}_0}$ von dem wir annehmen, dass es der Grobgitterraum ist. $\varepsilon_{ij} = 1$ ist die Cauchy-Schwarz Ungleichung

Lemma 6.5 (Eigenschaften der P_i)

24.4.09
4

Die Matrizen $P_i = R_i^T A_i^{-1} R_i A$ sind orthogonale Projektionen bezüglich des A-Skalarproduktes und es gelten folgende Rechenregeln

(i) $P_i^2 = P_i$

(ii) $AP_i = P_i^T A$

(iii) $\langle P_i x, P_i y\rangle_A = \langle x, P_i y\rangle_A \quad \forall x,y \in \mathbb{R}^I$

(iv) $\langle P_i x, (I-P_i)y\rangle_A = 0 \quad \forall x,y \in \mathbb{R}^I$

(v) $\|x\|_A^2 = \|P_i x\|_A^2 + \|(I-P_i)x\|_A^2 \quad \forall x \in \mathbb{R}^I$

(vi) $\|P_i x\|_A \leq \|x\|_A$

Beweis:

(i) $P_i^2 = \underbrace{R_i^T A_i^{-1} \overbrace{R_i A}^{A_i} R_i^T A_i^{-1} R_i A}_{=I} = R_i A_i^{-1} R_i A = P_i$

(ii) $AP_i = \underbrace{A R_i^T A_i^{-1} R_i} A = (R_i^T A_i^{-1} R_i A)^T A = P_i^T A$

(iii) $\langle P_i x, P_i y\rangle_A = x^T \underbrace{P_i^T A} P_i y \overset{(ii)}{=} x^T A P_i^2 y = x^T A P_i y = \langle x, P_i y\rangle_A$

(iv) $\langle P_i x, (I-P_i)y\rangle_A = \langle P_i x, y\rangle_A - \langle P_i x, P_i y\rangle_A$

$\overset{(iii)}{=} \langle P_i x, P_i y\rangle_A - \langle P_i x, P_i y\rangle_A = 0$

(v) $\|x\|_A^2 = \|P_i x + (I-P_i)x\|_A^2 = \langle P_i x + (I-P_i)x, P_i x, (I-P_i)x\rangle_A$

$= \langle P_i x, P_i x\rangle_A + \langle (I-P_i)x, (I-P_i)x\rangle_A$

$= \|P_i x\|_A^2 + \|(I-P_i)x\|_A^2$

(vi) Aus (v) folgt

$\|P_i x\|_A^2 = \|x\|_A^2 - \underbrace{\|(I-P_i)x\|_A^2}_{\geq 0} \leq \|x\|_A^2$

Wurzelziehen liefert die Behauptung. ∎

24.4.09
5

Lemma 6.6 (Abschätzung nach oben)

Mit Voraussetzung A2 folgt

$$\Big\langle \sum_{i=0}^{p} P_i x, x \Big\rangle_A \le (\varrho(\mathcal{E})+1)\,\langle x, x\rangle_A .$$

Beweis:

1. $\Big\langle \sum_{i=0}^{p} P_i x, x \Big\rangle_A = \langle P_0 x, x\rangle_A + \Big\langle \sum_{i=1}^{p} P_i x, x \Big\rangle_A$

   - $\langle P_0 x, x\rangle_A \overset{(iii)}{=} \langle P_0 x, P_0 x\rangle_A = \|P_0 x\|_A^2 \overset{(vi)}{\le} \|x\|_A^2 = \langle x, x\rangle_A$

   - $\hat{P} := \sum_{i=1}^{p} P_i$

2. $\underbrace{\langle \hat{P}x, \hat{P}x\rangle_A} = \Big\langle \sum_{i=1}^{p} P_i x, \sum_{i=1}^{p} P_i x \Big\rangle_A = \sum_{1 \le i,j \le p} \langle P_i x, P_j x\rangle_A$

   $\overset{\text{Vor. A2}}{\le} \sum_{1 \le i,j \le p} \varepsilon_{ij} \underbrace{\langle P_i x, P_i x\rangle_A^{1/2}}_{=:(z)_i} \underbrace{\langle P_j x, P_j x\rangle_A^{1/2}}_{=:(z)_j} \qquad z \in \mathbb{R}^p$

   $= z^T \mathcal{E} z$

   $\le \underbrace{\|\mathcal{E}\|_2}_{\text{Spektralnorm von } \mathcal{E}} \underbrace{\langle z, z\rangle}_{\text{euklidisches SP!}}$

   $= \varrho(\mathcal{E})$ da $\mathcal{E}$ sym. — d.h. diezugeordnete Matrixnorm zur euklidischen Norm.

   $= \varrho(\mathcal{E}) \sum_{i=1}^{p} (z)_i^2$

   $= \varrho(\mathcal{E}) \sum_{i=1}^{p} \langle P_i x, P_i x\rangle_A$

   $\overset{(iii)}{=} \varrho(\mathcal{E}) \Big\langle \sum_{i=1}^{p} P_i x, x \Big\rangle_A \overset{C.S.}{\le} \varrho(\mathcal{E}) \underbrace{\langle \hat{P}x, \hat{P}x\rangle_A^{1/2}} \langle x, x\rangle_A^{1/2}$

   also $\langle \hat{P}x, \hat{P}x\rangle_A^{1/2} \le \varrho(\mathcal{E}) \langle x, x\rangle_A^{1/2}$

3. $\langle \hat{P}x, x\rangle_A \overset{C.S.}{\le} \langle \hat{P}x, \hat{P}x\rangle_A^{1/2} \langle x, x\rangle_A^{1/2} \overset{2.}{\le} \varrho(\mathcal{E}) \langle x, x\rangle_A^{1/2} \langle x, x\rangle_A^{1/2} = \varrho(\mathcal{E}) \langle x, x\rangle_A$

also zusamm

$$\Big\langle \sum_{i=0}^{p} P_i x, x \Big\rangle_A \le \langle P_0 x, x\rangle_A + \langle \hat{P}x, x\rangle \le \langle x, x\rangle_A + \varrho(\mathcal{E}) \langle x, x\rangle_A = (\varrho(\mathcal{E})+1) \langle x, x\rangle_A$$

Hackbusch Lemma 2.9.3:

$\|\mathcal{E}\|_2 = \sup_{x \ne 0} \frac{\|\mathcal{E}x\|_2}{\|x\|_2}$ zugeord. Matrixnorm

$\|\mathcal{E}x\|_2 = \sup_{y \ne 0} \frac{|\langle \mathcal{E}x, y\rangle|}{\|y\|_2}$ „Dualitätsaussage"

Zusamm

$\|\mathcal{E}\|_2 = \sup_{x,y \ne 0} \frac{|\langle \mathcal{E}x, y\rangle|}{\|x\|_2 \|y\|_2}$

$\Rightarrow |\langle \mathcal{E}x, y\rangle| \le \|\mathcal{E}\|_2 \|x\|_2 \|y\|_2$

hier weiter 14.5.

06.V.09
6

Bemerkung 6.7 (unabhängige Korrekturen)

Zur Abschätzung des Spektralradius $\varrho(E)$ der Matrix aus A2, und damit in Lemma 6.6.

Jede Matrixnorm überschätzt den Spektralradius (Hackbusch, ..).
Für die Zeilensummennorm gilt:

$$\varrho(E) \leq \|E\|_\infty = \max_i \sum_{j=1}^{p} |\varepsilon_{ij}| .$$

Stammen die Räume V_i aus einer Gebietszerlegung so nutzen wir

$$\varepsilon_{ij} = \begin{cases} 0 & R_i A R_j^T = 0 \quad (\text{wenn } \overline{\Omega}_i \cap \overline{\Omega}_j = \phi) \\ 1 & \text{sonst} \quad (\text{Cauchy-Schwarz Ungl.}) \end{cases}$$

Bei überlappenden Gebietszerlegungsverfahren überlappt ein Teilgebiet üblicherweise nur mit einer maximalen Zahl N benachbarter Teilgebiete unabhängig von p. Also gilt $\varrho(E) \leq N$.

06.V.09
7

Lemma 6.8 (Zerlegungslemma, auch Lemma von Lions)

Aus Voraussetzung A1 (stabile Zerlegung) folgt

$$C_0^{-1} \langle x, x \rangle_A \le \Big\langle \sum_{i=0}^{p} P_i x, x \Big\rangle_A ,$$

d.h. C_0^{-1} liefert eine Abschätzung für den kleinsten Eigenwert.

Beweis:

$$\langle x, x \rangle_A \overset{\text{Zerlegg von } x}{=} \Big\langle x, \sum_{i=0}^{p} R_i^T x_i \Big\rangle_A = \sum_{i=0}^{p} \langle x, R_i^T x_i \rangle_A$$

(folgt doch schon aus $P_i^2 = P_i$ und $R_i^T x_i \in \operatorname{range}(P_i)$)

$$= \sum_{i=0}^{p} \langle x, \underbrace{\underbrace{R_i^T A_i^{-1}}_{P_i} \underbrace{R_i A R_i^T}_{A}}_{I} x_i \rangle_A = \sum_{i=0}^{p} \langle x, P_i R_i^T x_i \rangle_A$$

$$\overset{\text{ausschreiben}}{=} \sum_{i=0}^{p} x^T A P_i R_i^T x_i \overset{6.5\,(iii)}{=} \sum_{i=0}^{p} (P_i x)^T A R_i^T x_i = \sum_{i=0}^{p} \langle P_i x, R_i^T x_i \rangle_A$$

$$\overset{\text{Cauchy-Schwarz}}{\le} \sum_{i=0}^{p} \underbrace{\|P_i x\|_A}_{=:a_i} \underbrace{\|R_i^T x_i\|_A}_{=:b_i}$$

$$\overset{\text{Cauchy-Schwarz im } \mathbb{R}^{p+1}}{\le} \Big(\sum_{i=0}^{p} \|P_i x\|_A^2 \Big)^{1/2} \Big(\sum_{i=0}^{p} \|R_i^T x_i\|_A \Big)^{1/2} .$$

Quadrieren und Einsetzen von A1:

$$\langle x, x \rangle_A^2 \le \Big(\sum_{i=0}^{p} \|P_i x\|_A^2 \Big) \underbrace{\Big(\sum_{i=0}^{p} \|R_i^T x_i\|_A \Big)}_{\le C_0 \langle x, x \rangle_A} \le C_0 \langle x, x \rangle_A \sum_{i=0}^{p} \|P_i x\|_A^2$$

Kürzen liefert

$$\frac{1}{C_0} \langle x, x \rangle_A \le \sum_{i=0}^{p} \langle P_i x, P_i x \rangle_A \overset{6.5\,(iii)}{=} \sum \langle P_i x, x \rangle_A \qquad \blacksquare$$

10.V.09
8

Satz 6.9 (Konvergenz des additiven Verfahrens)

A1 und A2 liefern

$$\varkappa\left(\sum_{i=0}^{p} P_i\right) \leq C_0 (s(\mathcal{E})+1)$$

Beweis: Lemma 6.4, Definition der P_i, Lemma 6.6, Lemma 6.7.

7 Konvergenz des überlappenden Zweigitter-Schwarz-Verfahrens

06.V.09
1

7 Konvergenztheorie für überlappende Schwarz-Verfahren mit Grobgitterkorrektur

- folgt Toselli/Widlund
- Nur exakte Teilgebietslöser
- Nur I^h Verfeinerung von I^H

7.1 Technische Hilfsmittel

Es werden viele Resultate aus der Finite-Elemente-Theorie und der Funktionalanalysis verwendet, die wir hier nicht im Einzelnen beweisen können.

Wir beschränken uns daher auf die spezifischen Dinge.

Poincaré'sche und Friedrich'sche Ungleichung werden extensiv in einer spezifischen Form gebraucht. Daher wiederholen wir diese.

Didaktik: Vertausche 7.1 und 7.2., da man das eher kennt!

Lemma 7.1 (Poincaré Ungleichung)

$\Omega \subset \mathbb{R}^d$ beschränktes Gebiet mit Lipschitz-stetigem Rand.
(zu $x \in \partial\Omega$ gibt es eine Umgebung in der sich der Rand als Graph einer L-stetigen Funktion darstellen lässt).
Für $v \in H^1(\Omega)$ gilt dann

$$\|u\|^2_{L^2(\Omega)} \leq C_1 |u|^2_{H^1(\Omega)} + C_2 \left(\int_\Omega u \, dx \right)^2 .$$

C_1 und C_2 hängen nur von Ω ab.
Hat u Mittelwert $\emptyset$ so fällt der zweite Term weg! ∎

(Tos/Wid Lemma A.13)

Lemma 7.1 (Friedrichsche Ungleichung)

Ω beschränktes Gebiet mit L-stetigem Rand. $\Gamma \subseteq \partial\Omega$ habe nicht verschwindendes Maß. Dann gilt für $u \in H^1(\Omega)$

$$\|u\|^2_{L^2(\Omega)} \leq C_1 |u|^2_{H^1(\Omega)} + C_2 \|u\|^2_{L^2(\Gamma)}$$

mit Konstanten C_1, C_2 die nur von Ω und Γ abhängen.

Gilt $u = 0$ auf Γ (also insbesondere $u \in H^1_0(\Omega)$) so gilt

$$\|u\|^2_{L^2(\Omega)} \leq C_1 |u|^2_{H^1(\Omega)}.$$

Typische Anwendg: Für $u \in V = \{v \in H^1(\Omega) \mid v|_\Gamma = 0\}$ gilt

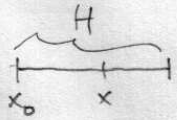
$$\|u\|^2_{H^1(\Omega)} \overset{\text{Def. der Norm}}{=} \|u\|^2_{L^2(\Omega)} + |u|^2_{H^1(\Omega)} \overset{\text{Poincaré}}{\leq} (C_1+1)\,|u|^2_{H^1(\Omega)}$$

und damit

$$\frac{1}{\sqrt{1+C_1}} \|u\|_{H^1(\Omega)} \leq |u|_{H^1(\Omega)} \leq \|u\|_{H^1(\Omega)}$$

(Semi-Norm äquivalent zur vollen Norm, Stetigkeit und Elliptizität der BLF) ∎.

Zum Beweis:



$$\int_{x_0}^{x} f'(\xi)\,d\xi = f(x) - f(\xi)$$

$$\Leftrightarrow |f(x)| \leq |f(x_0)| + \int_{x_0}^{x} |f'(\xi)|\,d\xi \leq$$

Fktionswert im Inneren — Fktwert am Rand (für H^1-Semi-Norm …)

$$\|f\|^2_{L^2(\Omega)} = \int_\Omega |f(x)|^2\,dx \leq |f(x_0)|^2 \underbrace{\int_\Omega 1\,dx}_{H} + \int_\Omega \left(\int_{x_0}^{x} |f'(\xi)|\,d\xi\right)^2 dx$$

$$\overset{\text{C.S.}}{\leq} |f(x_0)|^2 H + \int_\Omega \left[\left(\int_{x_0}^{x} |f'(\xi)|^2\,d\xi\right)^{1/2} \left(\int_{x_0}^{x} 1\,d\xi\right)^{1/2}\right]^2 dx$$

$$\leq |f(x_0)|^2 H + \int_{x_0}^{x} |f'(\xi)|^2\,d\xi \cdot \int_\Omega \int_{x_0}^{x} 1\,d\xi \leq |f(x_0)|^2 H + |f|^2_{H^1} H^2$$

In der Poincaré bzw Friedrich-Ungleichung hängen die Konstanten von der Gebietsgröße ab. Diese Abhängigkeit kann man explizit machen.

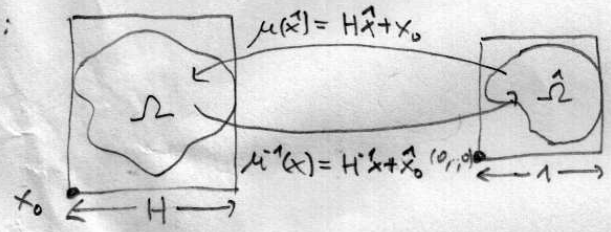
Bemerkung 7.3 (Skalierungsargument)

Es sei $\int_\Omega u\,dx = 0$ oder $u=0$ auf Γ in Lemma 7.1 bzw 7.2. und Ω passt in einen Würfel mit Kantenlänge H. Dann gilt

$$\|u\|^2_{L^2(\Omega)} \le \hat{C}\, H^2\, |u|^2_{H^1(\Omega)}$$

wobei $\hat{C}$ nur noch von der Gebietsform aber nicht der Größe abhängt.

Beweis:



$\hat{u}(\hat{x}) := u(\mu(\hat{x}))$

$(*)\quad \frac{\partial \hat{u}}{\partial \hat{x}_i}(\hat{x}) = \frac{\partial u(H\hat{x}+x_0)}{\partial \hat{x}_i} = \frac{\partial u}{\partial x_i}(\mu(\hat{x}))\cdot H$

$\nabla \mu(\hat{x}) = H\cdot I \quad \det H\cdot I = H^d$

$\nabla \mu^{-1}(x) = H^{-1} I$

$$\|u\|^2_{L^2(\Omega)} \overset{\text{Def}}{=} \int_\Omega |u(x)|^2\,dx \overset{\text{Trafo auf } \hat{\Omega}}{=} \int_{\hat{\Omega}} \underbrace{|u(\mu(\hat{x}))|^2}_{=:\hat{u}(\hat{x})} H^d\,d\hat{x}$$

$$= H^d \|\hat{u}\|^2_{L^2(\hat{\Omega})} \overset{\text{Poincaré/Friedrich}}{\le} \hat{C}\, H^d\, |\hat{u}|^2_{H^1(\hat{\Omega})}$$

$$= \hat{C}\, H^d \int_{\hat{\Omega}} \sum_{i=1}^d \left|\frac{\partial \hat{u}}{\partial \hat{x}}(\hat{x})\right|^2 d\hat{x}$$

Trafo $\hat{\Omega}\to\Omega$

$$\overset{(*)}{=} \hat{C}\, H^d \int_\Omega \sum_{i=1}^d \left| H \frac{\partial u}{\partial x_i}(\mu(\mu^{-1}(x))\right|^2 H^{-d}\,dx$$

$$= \hat{C}\, H^2 \int_\Omega \sum_{i=1}^d \left|\frac{\partial u}{\partial x_i}(x)\right| dx = \hat{C}\, H^2\, |u|^2_{H^1(\Omega)}$$

Dieses Vorgehen nennt man Skalierungsargument. Benutzt man auch in den FE-Approximationssätzen.

→ besser noch hierher!

08.V.09
4

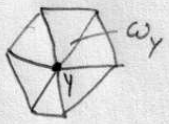
Definition 7.4 (Quasi-Interpolationsoperator) ablhier 14.5.09

V^H : Raum P_1/Q_1-Raum auf $\mathcal{T}^H$. Definiere

$$\tilde{I}^H : H_0^1(\Omega) \to V^H$$

mittels

$$(\tilde{I}^H u)(y) = \begin{cases} 0 & y \in \partial\Omega \\ |\omega_y|^{-1} \int_{\omega_y} u(x)\,dx & \text{sonst} \end{cases}$$



wobei: - y ein Knoten der Triangulierung $\mathcal{T}^H$,

- $\overline{\omega}_y = \bigcup_{\substack{K \in \mathcal{T}^H \\ y \text{ ist Ecke von } K}} K$
- $|\omega_y| = \int_{\omega_y} 1\,dx$.

Plan: Unten ist Voraussetzung A1 nachzuweisen. Hierfür ist zu jedem $u_h \in V^h$ eine Zerlegung $u_h = \sum_{i=0}^{p} u_i$, $u_i \in V_i$ anzugeben so dass $\sum_{i=0}^{p} a(u_i, u_i) \leq C_0\, a(u_h, u_h)$ mit C_0 möglichst unabhängig von H, h.

$V_0 = V^H$ ist der Grobgitterraum, V_i sind die Teilräume zu den $\hat{\lambda}_i$.

Idee der Zerlegung: $u_h \longrightarrow$

$u_0 = \tilde{I}^H u$ (nur notwendig wenn $V^H \not\subset V^h$)

$w = u_h - I^h u_0$ „Rest"

$w = \sum_{i=1}^{p} \theta_i w$ θ_i: „Partition der 1"

$\forall x \in \Omega: \sum_{i=1}^{p} \theta_i(x) = 1$

$u_i = I^h(\theta_i w)$

↑ Lagrange-Interpolation.

$\rightsquigarrow$ Zerlegung $V^h \ni u_h = I^h u_0 + \sum_{i=1}^{p} u_i$

08. V. 09
5

Zur Analyse der Eigenschaften von $\tilde{I}^H$ führen wir ein:

Zu $K \in \mathcal{T}^H$ definiere

$$\overline{\omega}_K = \bigcup_{\substack{K' \in \mathcal{T}^H \\ K' \cap K \neq \emptyset}} K'$$

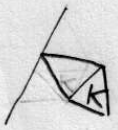
(Elemente woran abgeschlossen, das ist etwas, nicht so gleichlich).

ω_K selbst ist wieder offen!

Fall I: $\partial\omega_K \cap \partial\Omega = \emptyset$ → ω_K ist echt im Innern und bleibt so

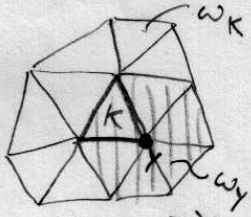
Fall II: $\partial\omega_K \cap \partial\Omega$ ist $(d-1)$-dimensional → o.k.

Fall II': $\partial\omega_K \cap \partial\Omega \neq \emptyset$ aber nicht $(d-1)$ dimensional



⇒ Erweitere ω_K um endlich viele (quasi-uniform) $K' \in \mathcal{T}^H$ so dass Fall II eintritt.

d.h. es gilt entweder Fall I oder Fall II.



Für $\tilde{K} \subseteq \mathcal{T}^H$ setze $\overline{\omega}_{\tilde{K}} = \bigcup_{K \in \tilde{K}} \overline{\omega}_K$.

⇒ $\omega_y \subseteq \omega_K$!

Lemma 7.5 (Stabilität der Quasi-Interpolation)

$\mathcal{T}^H$ sei quasiuniform und es sei $u \in H_0^1(\Omega)$. Dann gibt es ein $C > 0$ (und unabhängig von der Größe von K) so dass

„wie Fehlerabschätzung"

$$\| u - \tilde{I}^H u \|_{L^2(K)} \leq C H_K |u|_{H^1(\omega_K)}, \quad (7.1)$$

$$|\tilde{I}^H u|_{H^1(K)} \leq C |u|_{H^1(\omega_K)}. \quad (7.2)$$

Beweis: Wir beschränken uns auf P_1 in 3D, ist aber übertragbar.

i) $\{\varphi_i^H \mid i \in \mathcal{I}^H\}$ Lagrange-Basis. Es gilt für i Ecke von $K \subset \mathcal{T}^H$:

$$\|\Phi_i\|^2_{L^2(K)} = \int_K |\Phi_i(x)|^2 dx \underset{\uparrow \Phi_i(x) \le 1}{\le} \int_K dx \le C H_K^3 \quad \text{(quasi-uniform)}.$$

ii) $K \in \mathcal{T}^H$, y Ecke von K:

$$\left|(\tilde{I}^H u)(y)\right| = \left| |\omega_y|^{-1} \int_{\omega_y} u(x)\,dx \right| = |\omega_y|^{-1} \left| \underbrace{\int_{\omega_y} u(x) \cdot 1\,dx}_{L^2\text{-Skalarprodukt}} \right|$$

Cauchy-Schwarz $\rightarrow \le |\omega_y|^{-1} \left(\int_{\omega_y} |u(x)|^2 dx \right)^{1/2} \left(\underbrace{\int_{\omega_y} 1\,dx}_{=|\omega_y| \sim H_K^3} \right)^{1/2}$

$$= \|u\|_{L^2(\omega_y)} |\omega_y|^{-1/2}$$

$$\le \|u\|_{L^2(\omega_K)} \; C\, H_K^{-3/2} \qquad (\omega_y \subseteq \omega_K)$$

Das gilt für Fall I und Fall II von oben.

(iii)
$$\|\tilde{I}^H u\|_{L^2(K)} \underset{\uparrow \text{Def.}}{=} \left\| \sum_{i=1}^4 (\tilde{I}^H u)(y_i)\, \Phi_i \right\|_{L^2(K)}$$

Norm-eigensch. Dreiecksungl. $\rightarrow \le \sum_{i=1}^4 \underbrace{|(\tilde{I}^H u)(y_i)|}_{\text{siehe (ii) von oben!}} \|\Phi_i\|_{L^2(K)}$

(i), (ii) $\rightarrow \le \|u\|_{L^2(\omega_K)}\, C\, H_K^{-3/2}\, H_K^{3/2} = C \|u\|_{L^2(\omega_K)}.$

gilt ebenfalls für I und II.

Bemerkung: Konstanten „C" sind hier immer „generisch", d.h. nicht notwendigerweise gleich an verschiedenen Stellen!

(iv) Fall I: Sei $\partial\omega_K \cap \partial\Omega = \emptyset$.

Setze $\hat{u}(x) = u(x) - |\omega_K|^{-1} \int_{\omega_K} u\,dx$.

$\omega_K \leftarrow$ NEU! nicht ω_K!

Konstante Funktion, Mittelwert von u.

Nun gilt

$$\|u - \tilde{I}^H u\|^2_{L^2(K)} = \Big\| \underbrace{u - |\omega_K|^{-1}\int_{\omega_K} u\,dx}_{=\hat{u}} + \underbrace{|\omega_K|^{-1}\int_{\omega_K} u\,dx - \tilde{I}^H u}_{=\tilde{I}^H\left(|\omega_K|^{-1}\int_{\omega_K} u\,dx\right)} \Big\|^2_{L^2(K)}$$

da $\tilde{I}^H$ exakt für Funktionen, die konstant auf ω_K sind.

$$= \|\hat{u} - \tilde{I}^H \hat{u}\|^2_{L^2(K)}$$

Dreiecks-ungl. $\rightarrow \leq \left(\|\hat{u}\|_{L^2(K)} + \|\tilde{I}^H \hat{u}\|_{L^2(K)}\right)^2$

$\leq C\,\|\hat{u}\|_{L^2(\omega_K)}$ nach (iii)

(iii) $\rightarrow \leq C\,\|\hat{u}\|^2_{L^2(\omega_K)}$ größer!

Poincaré-Ungleichung $\rightarrow \leq C H_K^2\,|\hat{u}|^2_{H^1(\omega_K)} \leq C H_K^2 \left(|u|_{H^1(\omega_K)} + |\tilde{I}^H(|\omega_K|^{-1}\int_{\omega_K} u\,dx)|_{H^1(\omega_K)}\right)^2$

$= 0$ da konst!

$$\leq C H_K^2\,|u|^2_{H^1(\omega_K)}$$

Fall II: $\partial\omega_K \cap \partial\Omega \neq \emptyset$ mit Maß ungleich Null.

Sowohl u, als auch $\tilde{I}^H u$ sind $\emptyset$ auf $\partial\omega_K \cap \partial\Omega$ für $u \in H^1_0(\Omega)$.

Somit ist Friedrich-Ungl. anwendbar.

$$\|u - \tilde{I}^H u\|^2_{L^2(K)} \leq C\,\|u\|^2_{L^2(\omega_K)} \leq C H_K^2\,|u|^2_{H^1(\omega_K)}$$

Dreiecksungl., (iii) wie oben — Friedrich

Damit ist (7.1) bewiesen.

08.V.09
8

(V) Beweis von (7.2).

Fall I : ω_K im Inneren

$$|\tilde{I}^H u|^2_{H^1(K)} = \Big| \underbrace{-|\omega_K|^{-1}\int_{\omega_K} u\,dx}_{\text{Konstante ändert nix an Seminorm.}} + \tilde{I}^H u\Big|^2_{H^1(K)}$$

$\tilde{I}^H$ exakt auf Konst. $\rightarrow$ $= |\tilde{I}^H \hat{u}|^2_{H^1(K)}$

sog. inverse Ungleichg. $|u_h|_1 \le Ch^{-1}\|u_h\|_0$ gilt nur für FE-Funktionen $\rightarrow \le C H_K^{-2} \|\tilde{I}^H \hat{u}\|^2_{L^2(K)}$

$$= C H_K^{-2} \|\tilde{I}^H \hat{u} - \hat{u} + \hat{u}\|^2_{L^2(K)}$$

Dreiecksungl $\rightarrow \le C H_K^{-2} \left(\|\hat{u} - \tilde{I}^H \hat{u}\|_{L^2(K)} + \|\hat{u}\|_{L^2(K)}\right)^2$

$0 \le (a-b)^2 = a^2 - 2ab + b^2 \Rightarrow 2ab \le a^2 + b^2$; $(a+b)^2 \le a^2 + 2ab + b^2 \le 2a^2 + 2b^2$

$\rightarrow \le C H_K^{-2} 2 \left(\|\hat{u} - \tilde{I}^H \hat{u}\|^2_{L^2(K)} + \|\hat{u}\|^2_{L^2(K)}\right)$

(7.1) $\downarrow$; $\downarrow$ Poincaré-Ungl. Fall I

$$\le C H_K^{-2} \left(C' H_K^2 |\hat{u}|^2_{H^1(\omega_K)} + C'' H_K^2 |\hat{u}|^2_{H^1(\omega_K)}\right)$$

($H^1(\omega_K)$ vergrößert)

$|\hat{u}|_1 = |u|_1$ $\rightarrow \le C |u|^2_{H^1(\omega_K)}$

Fall II : ω_K am Rand

$$|\tilde{I}^H u|^2_{H^1(K)} \le C H_K^{-2} \|\tilde{I}^H u\|^2_{L^2(K)}$$

inverse Ungl.

(iii) $\rightarrow \le C H_K^{-2} \|\tilde{u}\|^2_{L^2(\omega_K)}$

Friedrich-Ungleichg. $u = 0$ auf $\partial\omega_K \cap \partial\Omega$. $\rightarrow \le C H_K^{-2} H_K^2 |u|^2_{H^1(\omega_K)}$

$$= C |u|^2_{H^1(\omega_K)}$$

∎

09.V.09
9

The Interpolation von $V^H \to V^h$ geschieht mit dem nodalen Interpolationsoperator I^h

Sei $\{\varphi_i^h \mid i \in \mathcal{I}^h\}$ die Lagrangebasis für V^h und $u \in C^0(\overline{\Omega})$, dann sei ($\varphi_i^h(x_j) = \delta_{ij}$)

$$I^h u = \sum_{i \in \mathcal{I}^h} u(x_i)\, \varphi_i^h .$$

I^h heißt Knoteninterpolationsoperator.

Für die Knoteninterpolation von V^H nach V^h zeigen wir:

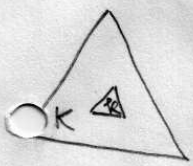
Lemma 7.6 (Stabilität der Knoteninterpolation für V^H).
Es gibt ein $C > 0$ unabhängig von h und H so dass

$$|u_H - I^h u_H|^2_{H^s(k)} \leq C h_k^{2(1-s)} |u_H|^2_{H^1(\omega_k)} \qquad (7.3)$$

für $k \in \mathcal{T}^h$, $u_H \in V^H$, $s = 0, 1$.

Beweis:

(i) Es sei $k \subseteq K$ für ein $K \in \mathcal{T}^H$. u_H ist linear auf K, somit ist die Knoteninterpolation exakt, d.h. $I^h u_H = u_H$ auf k.



$$\Rightarrow |u_H - I^h u_H|^2_{H^s(k)} = 0.$$

Ist $V^H \subseteq V^h$ (z.B. $\mathcal{T}^h$ Verfeinerung von $\mathcal{T}^H$) so ist $u_H - I^h u_H = 0$ und die Behauptung ist trivial erfüllt.

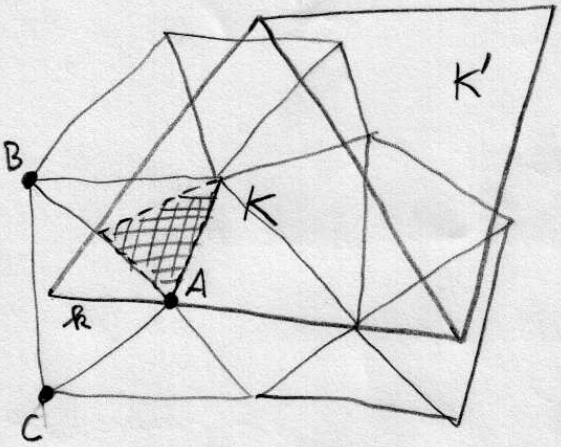
Für unsere Konstruktion sind wir damit fertig!

Der Rest des Beweises behandelt den Fall allgemeinerer Grobgitterräume $V^H \not\subseteq V^h$.

bis hier gekommen 19.5.09

(ii) Wir betrachten nun den Fall $k \cap K \neq 0 \wedge k \cap K' \neq 0$ | 09.V.09 10

für $K, K' \in \mathcal{T}^H$ und $K \neq K'$.



Wir betrachten nun $\mathbb{P}_1$ in $d=3$, der Beweis ist auf $d \neq 3$ übertragbar.

(i) Für die Basisfunktionen $\phi_i \in V^h$ gilt:

$$|\phi_i|^2_{H^1(k)} = \int_k \sum_{j=1}^{d} \underbrace{(\partial_j \phi_i)^2}_{\leq C \frac{1}{h_k}} dx \leq C h_k^{-2} h_k^3 = C h_k.$$

(shape-regular: $\rho \geq c h_k$)

(iii) $$|I^h u_H|^2_{H^1(k)} \overset{\text{Def.}}{=} \left| \sum_{i \in \{A,B,C,D\}} u_H(x_i) \phi_i \right|^2_{H^1(k)}$$

Nummern der vier Ecken.

$$\leq C \left(\sum_{i \in \{A,B,C,D\}} |u_H(x_i)| \, |\phi_i|_{H^1(k)} \right)^2$$

$$\rightarrow \leq C \sum_{i \in \{A,B,C,D\}} |u_H(x_i)|^2 \underbrace{|\phi_i|^2_{H^1(k)}}_{\leq C h_k}$$

$$\leq C h_k \sum_{i \in \{A,B,C,D\}} |u_H(x_i)|^2$$

"$(\sum_{i=1}^{n} z_i)^2 \leq \sum^{n} \sum_{i=1}^{n} z_i^2$"

besser: $\leq n \cdot \sum_{i=1}^{n} z_i^2$

$(z_1 + \dots z_n)^2$

$\leq 2\left[(z_1 + \dots z_{n/2-1})^2 + (z_{n/2} + \dots + z_n)^2\right]$

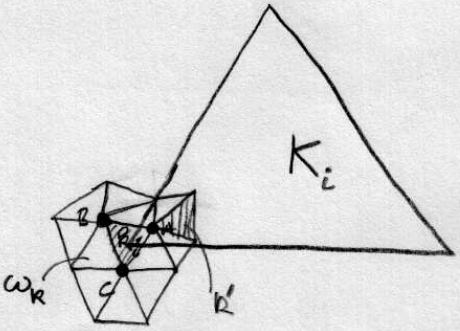
09.V.09
11

(iv) Abschätzen von $|u_H(x_i)|^2$ für $i \in \{A, B, C, D\}$

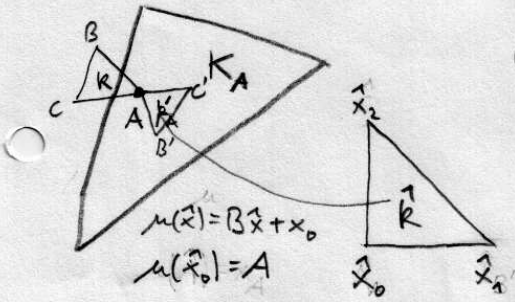
$\mathcal{T}^H$ ist shape-regular. Deshalb gibt es zu jeder Ecke $x_i \in \{A, B, C, D\}$ von $K \in \mathcal{T}^h$ einen zweiten Tetraeder k_i' so dass:

- k_i' hat x_i als eine der vier Ecken,
- k_i' hat Durchmesser $h_{k'} \leq C h_K$,
- $k_i' \subseteq \omega_K$,
- $k_i' \subseteq K_i$ für ein $K_i \in \mathcal{T}^H$.

Achtung: k_i' ist nicht notwendigerweise ein Element von $\mathcal{T}^h$!



O.B.d.A. betrachte $i = A$. Zu k_A wähle wie oben beschrieben ein k_A' mit den Ecken $\{x_A, x_{B'}, x_{C'}, x_D\}$. Transformiere k_A' auf das Referenz-Tetraeder und setze $\hat{u}(\hat{x}) = u_H(\mu(\hat{x}))$ (wie immer).



Auf dem Referenzelement gilt dann:

$$|u_H(A)| = |\hat{u}(\hat{x}_0)| \leq \left(\sum_{i=0}^{3} |\hat{u}(\hat{x}_i)|^2 \right)^{1/2}$$

(nur was dazu)

Äquivalenz von $\|\cdot\|_2$ und $\|\cdot\|_1$-Norm $\rightarrow \leq C \left(\sum_{i=0}^{3} |\hat{u}(\hat{x}_i)| \right)$

Mittelpunktsregel exakt für lineare Fkt. $= C \frac{4}{|\hat{k}|} \left(|\hat{k}| \sum_{i=0}^{3} \frac{1}{4} |\hat{u}(\hat{x}_i)| \right)$

$\left(1 \, d\hat{x} \rightarrow |\hat{k}|\right)$

$$= C \int_{\hat{k}} |\hat{u}(\hat{x})| \, d\hat{x} \leq C \left(\int_{\hat{k}} |\hat{u}(\hat{x})|^2 d\hat{x} \right)^{1/2} \left(\left(\int_{\hat{k}} 1 \, d\hat{x} \right)^{1/2} \leq C \|\hat{u}\|_{L^2(\hat{k})} \right.$$

(Cauchy-)

09. V. 09
12

Nun Trafo auf das echte Tetraeder K'_A

$$\|\hat{u}\|^2_{L^2(\hat{K})} \overset{\text{def.}}{=} \int_{\hat{K}} |\hat{u}(\hat{x})|^2 d\hat{x}$$

Trafo auf K'_A $\searrow$

$$= \int_K \underbrace{|\hat{u}(\mu^{-1}(x))|^2}_{= u_H(x)} \underbrace{\det B^{-1}}_{\leq C h_K^{-3} \leftarrow \text{Größe von } K'_A} dx$$

$$\leq C h_K^{-3} \int_K |u_H(x)|^2 dx = C h_K^{-3} \|u_H\|^2_{L^2(K'_A)}$$

Also insgesamt: $|u_H(A)|^2 \leq C h_K^{-3} \|u_H\|^2_{L^2(K'_A)} \leq C h_K^{-3} \|u_H\|^2_{L^2(\omega_K)}$

$K'_A \subseteq \omega_K$

(v) Nun weiter mit (iii)

Nach (iii) gilt

$$|I^h u_H|^2_{H^1(K)} \leq C h_K \sum_{i \in \{A,B,C,D\}} |u_H(x_i)|^2$$

$$\text{(iv)} \rightarrow \leq C h_K \sum_{i \in \{A,B,C,D\}} h_K^{-3} \|u_H\|^2_{L^2(\omega_K)}$$

$$= C h_K^{-2} \|u_H\|^2_{L^2(\omega_K)}$$

(vi) Sei $\bar{u}_H$ der Mittelwert von u_H in ω_K.

I^h ist exakt für konstante Funktionen. Also gilt:

$$|u_H - I^h u_H|^2_{H^1(K)} = |u_H - \bar{u}_H + \bar{u}_H - I^h u_H|^2_{H^1(K)}$$

$$= |(u_H - \bar{u}_H) + I^h(u_H - \bar{u}_H)|^2_{H^1(K)}$$

$$\leq 2\left(|u_H - \bar{u}_H|^2_{H^1(K)} + \underbrace{|I^h(u_H - \bar{u}_H)|^2_{H^1(K)}}_{\text{(v)}} \right)$$

$$\leq C |u_H - \bar{u}_H|^2_{H^1(K)} + C' h_K^{-2} \|u_H - \bar{u}_H\|^2_{L^2(\omega_K)}$$

Poincaré-Ungl. $u_H - \bar{u}_H$ hat MW $\emptyset$. $\rightarrow$

$$\leq C |u_H - \bar{u}_H|^2_{H^1(K)} + C' h_K^{-2} h_K^2 |u_H - \bar{u}_H|^2_{H^1(\omega_K)}$$

$$\leq C |u_H - \bar{u}_H|^2_{H^1(K)} = C |u_H|^2_{H^1(K)}$$

Das war (7.3) für $s = 1$.

09. V. 09
13

(vii) Der Fall $s=0$.

Wie in (iii) nur mit der L_2-Norm erhält man:

$$\| I^h u_H \|^2_{L^2(k)} = \| \sum_{i \in \{A,B,C,D\}} u_H(x_i)\, \phi_i \|^2_{L^2(k)}$$

Dreiecksungl. $\rightarrow \leq C \sum_{i \in \{A,B,C,D\}} |u_H(x_i)|^2 \, \| \phi_i \|^2_{L^2(k)}$

$\| \phi_i \|_{L^2} \leq h_k^3 \rightarrow \leq C\, h_k^3 \sum_{i \in \{A,B,C,D\}} |u_H(x_i)|^2$

(iv) $\rightarrow \leq C \, \| u_H \|^2_{L^2(\omega_k)}$

(Dies ist mit (v) zu vergleichen. Links $L^2(k)$-Norm, dafür fehlt h_k^{-2}).

Mit $\bar{u}_H$ dem Mittelwert auf ω_k und der Poincaré-Ungl. erhält man:

$$\| u_H - I^h u_H \|^2_{L^2(k)} = \| u_H - \bar{u}_H + I^h(u_H - \bar{u}_H) \|^2_{L^2(k)}$$

$$\leq 2 \, \| u_H - \bar{u}_H \|^2_{L^2(k)} + \| I^h(u_H - \bar{u}_H) \|^2_{L^2(k)}$$

$$\leq 2 \, \| u_H - \bar{u}_H \|^2_{L^2(\omega_k)} + C \, \| u_H - \bar{u}_H \|^2_{L^2(\omega_k)}$$

(größer; (vii))

$$\leq C \, \| u_H - \bar{u}_H \|^2_{L^2(\omega_k)}$$

Poincare $\rightarrow \leq C\, h_k^2 \, | u_H - \bar{u}_H |^2_{H^1(\omega_k)}$

$\bar{u}_H$ konst. $\rightarrow = C\, h_k^2 \, | u_H |^2_{H^1(\omega_k)}$

Betr. $u_h, v_h \in V^h$, stückweise linear (P_1). Damit ist $u_h v_h$ eine stückweise quadratische Funktion. Dafür brauchen wir.

09.V.09
14

Lemma 7.7 Sei u_h eine stückweise quadratische Funktion (also P_2). und I^h sei wie oben die Knoteninterpolation auf stückweise lineare (P_1) Funktionen auf dem selben Gitter. Dann gibt es ein C unabh. von h so dass:

$$|I^h u_h|_{H^1(K)} \leq C |u_h|_{H^1(K)} \quad \text{für alle } K \in \mathcal{T}^h.$$

Beweis:

$$|I^h u_h|^2_{H^1(K)} = |I^h u_h - u_h + u_h|^2_{H^1(K)}$$

$$\leq 2 \underbrace{|u_h - I^h u_h|^2_{H^1(K)}}_{\text{Approximations-fehler } u_h|_K \in H^2(K)} + 2|u_h|^2_{H^1(K)}$$

$$\leq C h_K^2 |u_h|^2_{H^2(K)} + 2|u_h|^2_{H^1(K)}$$

inverse Ungl. / u_h ist FE-Fkt. $\rightarrow$

$$\leq C h_K^2 h_K^{-2} |u_h|^2_{H^1(K)} + 2|u_h|^2_{H^1(K)}$$

$$\leq C |u_h|^2_{H^1(K)}$$

∎

09.V.09
15

Der Overlap wird in folgendem Lemma berücksichtigt.

Lemma 7.8

Sei $\hat{\Omega}_i$ ein Teilgebiet und δ_i der Overlap für dieses Teilgebiet (in der Praxis wählt man für alle denselben). $H_i = \operatorname{diam}(\hat{\Omega}_i)$ und $\hat{\Omega}_{i,\delta_i} = \{ x \in \hat{\Omega}_i \mid \operatorname{dist}(x, \partial\hat{\Omega}_i \setminus \partial\Omega) < \delta_i \}$. Dann gibt es ein $C > 0$ so dass für alle $u \in H^1(\hat{\Omega}_i)$:

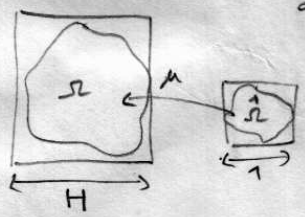
$$\|u\|^2_{L^2(\hat{\Omega}_{i,\delta_i})} \leq C\delta_i^2 \left(\left(1 + \frac{H_i}{\delta_i}\right) |u|^2_{H^1(\hat{\Omega}_i)} + \frac{1}{H_i \delta_i} \|u\|^2_{L^2(\hat{\Omega}_i)} \right).$$

Beweis:

i) Der Spursatz (Funktionalanalysis) sagt $\|\underbrace{u}_{\text{eigentlich } \gamma}\|_{L^2(\partial\Omega)} \leq c \|u\|_{H^1(\Omega)}$. Durch ein Skalierungsargument machen wir die Abhängigkeit von der Größe explizit.

$$\|u\|^2_{L^2(\partial\Omega)} = \int_{\partial\Omega} |u(s)|^2 ds \overset{\text{Trafo } \Omega \to \hat{\Omega}}{=} \int_{\partial\hat{\Omega}} \underbrace{\left| u(\mu(\hat{s})) \right|^2}_{=:\hat{u}(\hat{s})} H^{d-1} d\hat{s} = H^{d-1} \|\hat{u}\|^2_{L^2(\partial\hat{\Omega})}$$

(Oberflächenintegral; anderer Transformationssatz, Oberflächenintegral extra Art.)



Spursatz auf $\hat{\Omega}$

$$\leq \hat{C} H^{d-1} \|\hat{u}\|^2_{H^1(\hat{\Omega})}$$

$$= \hat{C} H^{d-1} \left(\int_{\hat{\Omega}} |\hat{u}|^2 d\hat{x} + \int_{\hat{\Omega}} \sum_{i=1}^{d} |\partial_i \hat{u}|^2 d\hat{x} \right)$$

Trafo $\hat{\Omega} \to \Omega$

$$= \hat{C} H^{d-1} \left(\int_{\Omega} |u|^2 H^{-d} dx + \int_{\Omega} \sum_{i=1}^{d} |H \partial_i u|^2 H^{-d} dx \right)$$

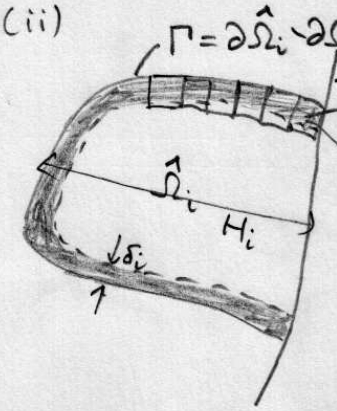
$$= \hat{C} \left(H^{d-1} H^{-d} \|u\|^2_{L^2(\Omega)} + H^{d-1} H^2 H^{-d} |u|^2_{H^1(\Omega)} \right)$$

also

$$\|u\|^2_{L^2(\partial\Omega)} \leq \hat{C} \left(\frac{1}{H} \|u\|^2_{L^2(\Omega)} + H |u|^2_{H^1(\Omega)} \right)$$

09. V. 09
16

(ii)



Triangulierung $\mathcal{T}_{i,\delta_i}$ von $\hat{\Omega}_{i,\delta_i}$ im Patches der Größe δ_i, shape regular.

$$\|u\|^2_{L^2(\hat{\Omega}_{i,\delta_i})} = \sum_{k \in \mathcal{T}_{i,\delta_i}} \|u\|^2_{L^2(k)}$$

Skalierungsargument für die ganze Friedrichsche Ungl. habe ich nicht gemacht

Friedrich'sche Ungleichg $\rightarrow$

$$\leq C \sum_{k \in \mathcal{T}_{i,\delta_i}} \left(\delta_i^2 |u|^2_{H^1(k)} + \delta_i \|u\|^2_{L^2(\partial k \cap \Gamma)} \right)$$

δ_i^2: durchlaufen das k

Summe über alle $\partial k \cap \Gamma$ und $\partial\Omega \cap \partial\hat{\Omega}_i$ dann

$$= C \left(\delta_i^2 |u|^2_{H^1(\hat{\Omega}_{i,\delta_i})} + \delta_i \|u\|^2_{L^2(\partial\hat{\Omega}_i)} \right)$$

Spursatz (i) $\rightarrow$

$$\leq C \left(\delta_i^2 |u|^2_{H^1(\hat{\Omega}_{i,\delta_i})} + \delta_i \hat{C} \left(\frac{1}{H_i} \|u\|^2_{L^2(\hat{\Omega}_i)} + H_i |u|^2_{H^1(\hat{\Omega}_i)} \right) \right)$$

$\downarrow$ vergrößern

$$\leq C \delta_i^2 \left(\left(1 + \frac{H_i}{\delta_i}\right) |u|^2_{H^1(\hat{\Omega}_i)} + \frac{1}{H_i \delta_i} \|u\|^2_{L^2(\hat{\Omega}_i)} \right)$$

10. V. 09
17

Zum Lokalisieren einer Funktion $u_h \in V^h$ auf ein Teilgebiet $\hat{\Omega}_i$ benötigen wir eine sog. „Partition der Eins".

Um diese zu konstruieren sind zwei Voraussetzungen für $\{\hat{\Omega}_i\}$ nötig.

Voraussetzung 01 (Mindestabstand)

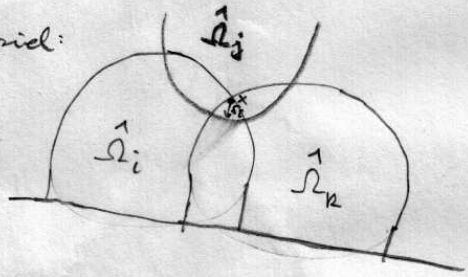
Sei $\{\hat{\Omega}_i\}_{i=1}^{p}$ eine Zerlegung von Ω in überlappende Teilgebiete.

Für $i = 1, \dots, p$ existiere je ein $\delta_i > 0$ so dass

$$\forall x \in \hat{\Omega}_i \quad \exists\, j(x) \in \{1, \dots, p\} \text{ so dass } x \in \hat{\Omega}_{j(x)} \wedge \operatorname{dist}(x, \partial\hat{\Omega}_{j(x)} \setminus \partial\Omega) \geq \delta_i .$$

Das heißt $x \in \hat{\Omega}_i$ muss mindestens in einem Teilgebiet $j(x)$ den Abstand δ_i vom Rand haben.

Beispiel:



Voraussetzung 02 (Endliche Überdeckung)

Es existiert eine Färbung von $\{\hat{\Omega}_i\}_{i=1}^{p}$ mit höchstens N^c Farben.

Formal: Es gibt eine Abb $c : \{1, \dots, p\} \to \{1, \dots, N^c\}$ so dass

$$c(i) = c(j) \Rightarrow \hat{\Omega}_i \cap \hat{\Omega}_j = \emptyset .$$

Daraus folgt:

Jedes $x \in \Omega$ ist Element von höchstens N^c Teilgebieten.

Beweis: durch Widerspruch. Setze $J_x = \{ j \in 1, \dots, p \mid x \in \hat{\Omega}_j \}$. Beh. $|J_x| \leq N^c$

Für $i, j \in J_x$, $i \neq j$ gilt $\hat{\Omega}_i \cap \hat{\Omega}_j$ (offensichtlich) und somit $c(i) \neq c(j)$ wg 02.

also muss $N^c \geq |J_x|$ sein. ∎

Andererseits folgt alleine aus 02 keine Aussage über die Zahl der „Nachbarn" von $\hat{\Omega}_i$, d.h. $J_i = \{ j \in \{1, \dots, p\} \mid \hat{\Omega}_j \cap \hat{\Omega}_i \neq \emptyset \}$.

10. V. 09
18

Lemma 7.9 (Partition der Eins)

Sei $\{\hat\Omega_i\}_{i=1}^p$ eine Zerlegung von Ω in überlappende Teilgebiete, die die Voraussetzungen O1 und O2 erfüllt. Dann gibt es Funktionen $\{\tilde\theta_i\}_{i=1}^p$ aus $W^{1,\infty}(\Omega)$ so dass

(a) $0 \le \tilde\theta_i(x) \le 1 \quad x \in \overline{\Omega}$,

(b) $\operatorname{supp}(\tilde\theta_i) = \{x \in \Omega \mid \tilde\theta_i(x) \neq 0\} \subset \overline{\hat\Omega_i}$,

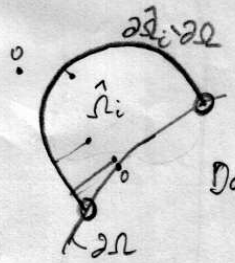
(c) $\sum_{i=1}^p \tilde\theta_i(x) = 1$,

(d) $\|\nabla \tilde\theta_i\|_\infty \le \frac{C}{\delta_i} \quad 1 \le i \le p.$ alternativ: $\|\nabla \tilde\theta_i(x)\|_\infty \le \frac{C}{\delta_i} \quad \forall x \in \hat\Omega_i.$

$W^{1,\infty}(\Omega)$: Funktionswert und Ableitung ist „fast überall" beschränkt, d.h. höchstens auf einer Menge vom Maß $\emptyset$ unbeschränkt.

Beweis: Konstruktiv! Definiere

$1 \le i \le p$: $d_i(x) = \begin{cases} \operatorname{dist}(x, \partial\hat\Omega_i \setminus \partial\Omega) & x \in \overline{\hat\Omega_i} \cup (\partial\hat\Omega_i \cap \partial\Omega) \\ 0 & \text{sonst} \end{cases}$



offensichtlich ist $d_i(x) = 0$ für $x \in \Omega \setminus \overline{\hat\Omega_i}$

Dann setze

$$\tilde\theta_i(x) = \frac{d_i(x)}{\sum_{k=1}^p d_k(x)}.$$

(a) $\tilde\theta_i \ge 0$ folgt aus $d_i \ge 0$ (Abstandsfunktion)

(c) $\sum_{i=1}^p \tilde\theta_i(x) = \sum_{i=1}^p \frac{d_i(x)}{\sum_k d_k(x)} = 1 \quad \Rightarrow \tilde\theta_i(x) = 1 - \underbrace{\sum_{\substack{k=1 \\ k \neq i}}^p \tilde\theta_i(x)}_{\ge 0} \le 1$

(b) $\tilde\theta_i(x) = 0$ auf $\partial\hat\Omega_i$ und ausserhalb.

Darüberhinaus gilt $\tilde\theta_i \in C^0(\overline{\Omega})$

10.V.09
19

(d) ist das eigentlich schwierige.

(i) Um den Gradienten zu beschränken zeigen wir

$$|\tilde\theta_i(x) - \tilde\theta_i(y)| \le \frac{c}{\delta_i}|x-y|. \quad \text{für alle } y \text{ genügend nahe an } x.$$

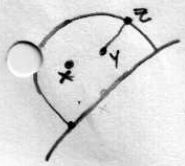
Denn mit so einem Resultat gilt

$$|\partial_j \tilde\theta_i(x)| = \lim_{h\to 0}\left|\frac{\tilde\theta_i(x+he_j)-\tilde\theta_i(x)}{h}\right| \le \lim_{h\to 0}\frac{c}{\delta_i}\frac{h}{h} = \frac{c}{\delta_i}.$$

(ii) Setze $\delta_k(x,y) := d_k(x) - d_k(y)$. Wir zeigen

$$|\delta_k(x,y)| \le |x-y|$$

Fall I: $x \in \hat\Omega_k$: wähle $y \in \hat\Omega_k$ und dazu $z \in \partial\Omega_k \setminus \partial\Omega$ sodass $d_k(y) = \operatorname{dist}(y, \partial\Omega_k \setminus \partial\Omega) = |y-z|$



somit gilt

$$d_k(x) = \operatorname{dist}(x, \partial\Omega_k \setminus \partial\Omega) \le |x-z| = |x-y+y-z| \le |x-y| + \underbrace{|y-z|}_{=d_k(y)}$$

($z \in \partial\Omega_k \setminus \partial\Omega$; Dreiecks)

$$\iff d_k(x) - d_k(y) \le |x-y|$$

Nun wähle $z' \in \partial\hat\Omega_k \setminus \partial\Omega$ so dass $d_k(x) = \operatorname{dist}(x, \partial\hat\Omega_k \setminus \partial\Omega) = |x-z'|$. Dann gilt

$$d_k(y) \le |y-z'| = |y-x+x-z'| \le |y-x| + d_k(x) \iff d_k(y) - d_k(x) \le |y-x|$$

aus beidem folgt somit $|d_k(x) - d_k(y)| \le |x-y|$

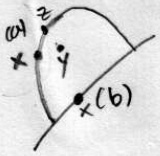
Fall II $x \in \hat\Omega_i \setminus \overline{\hat\Omega}_k$ (also außerhalb $\overline{\hat\Omega}_k$) für ein $i \neq k$

dann wähle $y \in \hat\Omega_i \setminus \overline{\hat\Omega}_k$ und es gilt

$$|d_k(x) - d_k(y)| = |0-0| = 0 \le |x-y|$$

Fall III $x \in \partial\hat\Omega_k$

(a) $x \in \partial\hat\Omega_k \setminus \partial\Omega$. Wähle $y \in \hat\Omega_k$



1.) Sei $z \in \partial\hat\Omega_k \setminus \partial\Omega$ so dass $\operatorname{dist}(y, \partial\hat\Omega_k \setminus \partial\Omega) = |y-z|$

also $d_k(x) = 0 \le |x-z| = |x-y+y-z| \le |x-y| + d_k(y)$

$\iff d_k(x) - d_k(y) \le |x-y|$

2. $d_k(y) = |y-z| \le |y-x|$ da z der nächste Randpunkt auf $\partial\hat\Omega_k \setminus \partial\Omega$ ist.

(b) $x \in \partial\hat\Omega_k \cap \partial\Omega$. Wähle $y \in \hat\Omega_k$ und z wie oben.

1.) $d_k(x) \le |x-z| = |x-y+y-z| \le |x-y| + d_k(y) \iff d_k(x) - d_k(y) \le |x-y|$ wie immer.

2.) Wähle $z' \in \partial\hat\Omega_k \setminus \partial\Omega$ so dass $d_k(x) = \operatorname{dist}(x, \partial\hat\Omega_k \setminus \partial\Omega) = |x-z'|$

$d_k(y) \le |y-z'| = |y-x+x-z'| \le |y-x| + d_k(x) \iff d_k(y) - d_k(x) \le |y-x|$

10.V.09
20

(iii) $\sum_{k=1}^{p} d_k(x) \geq d_{j(x)}(x) \geq \delta_i$ mit $j(x)$ aus Voraussetzung 01. (ein Abstand ist größer als δ_i)

(iv) $|\tilde\theta_i(x) - \tilde\theta_i(y)| \overset{\text{Def.}}{=} \left| \frac{d_i(x)}{\sum_{k=1}^{p} d_k(x)} - \frac{d_i(y)}{\sum_{k=1}^{p} d_k(y)} \right|$

Hauptnenner $= \left| \frac{d_i(x)\sum_{k=1}^{p} d_k(y) - d_i(y)\sum_{k=1}^{p} d_k(x)}{\left(\sum_{k=1}^{p} d_k(x)\right)\left(\sum_{k=1}^{p} d_k(y)\right)} \right|$

$k=i$ heftet sich weg $= \left| \frac{d_i(x)\sum_{\substack{k=1\\k\neq i}}^{p} d_k(y) - d_i(y)\sum_{\substack{k=1\\k\neq i}}^{p} d_k(x)}{\left(\sum_{k=1}^{p} d_k(x)\right)\left(\sum_{k=1}^{p} d_k(y)\right)} \right|$

Erweitern $= \left| \frac{d_i(x)\sum_{\substack{k=1\\k\neq i}}^{p} d_k(y) - d_i(x)\sum_{\substack{k=1\\k\neq i}}^{p} d_k(x) + d_i(x)\sum_{\substack{k=1\\k\neq i}}^{p} d_k(x) - d_i(y)\sum_{\substack{k=1\\k\neq i}}^{p} d_k(x)}{\left(\sum_{k=1}^{p} d_k(x)\right)\left(\sum_{k=1}^{p} d_k(y)\right)} \right|$

$= \frac{1}{\sum_{k=1}^{p} d_k(y)} \left| \underbrace{\frac{d_i(x)}{\sum_{k=1}^{p} d_k(x)}}_{\tilde\theta_i(x)} \sum_{\substack{k=1\\k\neq i}}^{p} \underbrace{(d_k(y) - d_k(x))}_{\delta_k(y,x)} + \underbrace{\frac{\sum_{\substack{k=1\\k\neq i}}^{p} d_k(x)}{\sum_{k=1}^{p} d_k(x)}}_{1-\tilde\theta_i(x)} \underbrace{(d_i(x) - d_i(y))}_{\delta_i(x,y)} \right|$

(ch positiv)

Betrag reinziehen $\leq \underbrace{\frac{1}{\sum_{k=1}^{p} d_k(y)}}_{\leq \frac{1}{\delta_i} \text{ laut (iii)}} \left[\underbrace{\tilde\theta_i(x)}_{\leq 1} \underbrace{\sum_{\substack{k=1\\k\neq i}}^{p} |\delta_k(y,x)|}_{\leq N^c |x-y|} + \underbrace{(1-\tilde\theta_i(x))}_{\leq 1} \underbrace{|\delta_i(x,y)|}_{\leq |x-y|} \right]$

da x in höchstens N^c Teilgebieten enthalten

$\leq \frac{N^c+1}{\delta_i} |x-y|$

10.V.09
21

7.2 Anwendung der abstrakten Theorie für das überlappende Schwarz-Verfahren

Lemma 6.6 und
Bemerkung 6.7 liefern direkt die Abschätzung nach oben:

$$\left\langle \sum_{i=0}^{p} P_i x_1, x \right\rangle_A \leq (N+1) \langle x_1, x \rangle_A$$

wobei N die maximale

$$N = \max_i N_i \qquad N_i = \left\{ j \in 1, \dots, p \mid \hat{\Omega}_i \cap \hat{\Omega}_j \neq \emptyset \right\}.$$

Ein „direktes Färbungsargument" liefert eine etwas bessere Konstante:

Lemma 7.10

Es sei N^c die Anzahl der Farben aus Voraussetzung 01.
(Naja, etwas erweitert: $\operatorname{dist}(\Omega_i, \Omega_j) \geq k$ (Gitterweite))
Dann gilt

$$\left\langle \sum_{i=0} P_i x_1, x \right\rangle_A \leq (N^c+1) \langle x_1, x \rangle_A.$$

Beweis: Setze $J_k = \{ i \in 1, \dots, p \mid c(i) = k \}$ für $k = 1, \dots, N^c$.

$$\left\langle \sum_{i=0}^{p} P_i x_1, x \right\rangle_A = \left\langle \sum_{k=1}^{N^c} \sum_{i \in J_k} P_i x_1, x \right\rangle_A$$

$$= \sum_{k=1}^{N^c} \sum_{i \in J_k} \langle P_i x_1, x \rangle_A = \sum_{k=1}^{N^c} \sum_{i \in J_k} \left\langle P_i x_1, \underbrace{P_i x}_{\text{wg. 6.5(iii)}} + \underbrace{\sum_{\substack{j \in J_k \\ j \neq i}} P_j x}_{\text{weil } \langle P_i x_1, P_j x \rangle_A = 0} \right\rangle_A$$

$$= \sum_{k=1}^{N^c} \left\langle \sum_{i \in J_k} P_i x_1, \sum_{i \in J_k} P_i x \right\rangle_A$$

Summe von orth. Projektionen ist wieder eine orthogonale Projektion. Das ist nicht gezeigt. Ü Ü?

$$\leq \sum_{k=1}^{N^c} \langle x_1, x \rangle_A = N^c \langle x_1, x \rangle_A \qquad \square$$

10. V. 09
22

Lemma 7.11. (Voraussetzung A1 (stabile Zerlegung))

Für das additive Schwarz-Verfahren gilt

$$\left[\left(1+\frac{H}{\delta}\right)\right]^{-1}\langle x,x\rangle_A \leq \langle \sum_{i=0}^{p} P_i x, x\rangle_A .$$

Beweis. Nach dem Zerlegungslemma 6.8 folgt dies aus der Stabilität der Zerlegung (Voraussetzung A1), die somit nachzuweisen ist.

(i) Konstruktion der Zerlegung. Gegeben sei $u_h \in V^h$ (P_1-Finite Elemente!).

Setze $u_0 := \tilde{I}^H u_h \in V^H$

dann

$$w = u_h - \underbrace{I^h u_0}_{=u_0 \text{ für } V^H \subset V^h} \in V^h.$$

wir lassen aber den allgemeinen Fall zu

Mittels der Partition der Eins setzen wir

$$\theta_i = I^h \tilde{\theta}_i \in V^h \qquad \{\theta_i\}_{i=1}^p \text{ ist immer noch eine Partition oder 1 in } V^h:$$

- Knotenwerte addieren sich zu 1
- $\Rightarrow$ auch im Elementinneren 1

Und schließlich

$$u_i = R_i'' I^h(\theta_i w) \in V_{h,i} \qquad 1 \leq i \leq p.$$

R_i'': Hat sind Null außerhalb $\tilde{\Omega}_i$. $\Rightarrow$ Wegfallen!

$I^h(\theta_i w) \in V^h$ $\subset V^h$ stückw. quadratisch $\in V_h$

Die Fkt $I^h u_0, u_i \mid 1 \leq i \leq p$ sind eine Zerlegung von $u_h \in V^h$:

$$I^h u_0 + \sum_{i=1}^{p} u_i = I^h \tilde{I}^H u_h + \sum_{i=1}^{p} I^h(\theta_i w)$$

I^h linear

$$= I^h \tilde{I}^H u_h + I^h\Big(\underbrace{\sum_{i=1}^{p} \theta_i w}_{=w \text{ da } \{\theta_i\} \text{ Partition der Eins}}\Big)$$

$$= I^h \tilde{I}^H u_h + I^h(u_h - I^h \tilde{I}^H u_h)$$

Linearität $[I^h]^2 = I^h$ $= I^h u_h = u_h$.

Nodale Interpolation I^h

Entsprechend für die zugehörigen Koeffizienten: $x = R_0^T x_0 + \sum_{i=1}^{p} R_i^T x_i$

(Koeffiz. von u_0; Koeffiz. von u_i; Teilgebiet fast oder)

10. V. 09
23

Nun ist zu zeigen (A1)

$$\sum_{i=0}^{p} \langle R_i^T x_i, R_i^T x_i \rangle_A = \underbrace{a(I^h u_0, I^h u_0)}_{(ii)} + \underbrace{\sum_{i=1}^{p} a(u_i, u_i)}_{(iii)}$$

(kontinuierlich)

(ii) Es gilt $\underline{c} |v|^2_{H^1(\Omega)} \le a(v,v) \le \bar{c} |v|^2_{H^1(\Omega)}$ (gleichmäßige Elliptiz und Beschränktheit der Koeffizienten K)

(iii)

$$a(I^h u_0, I^h u_0) \le C |I^h u_0|^2_{H^1(\Omega)}$$

(quasi Stetigkeit der BLF i.e. Beschränktheit der Koeff.)

~~$C |I^h u_0|^2_{H^1(\Omega)}$~~

$= : u_H$ (← zwei Bez. für das selbe $u_H = u_0$ d.i. klöd 24.5.)

Lemma 7.6 anwenden (ausführlich)

$$|I^h u_0|^2_{H^1(\Omega)} = |u_0 - u_0 + I^h u_0|^2_{H^1(\Omega)}$$

$$\le 2 \left(|u_0|^2_{H^1(\Omega)} + |u_0 - I^h u_0|^2_{H^1(\Omega)} \right)$$

$$= 2 \left(|u_0|^2_{H^1(\Omega)} + \sum_{k \in J_h} |u_0 - I^h u_0|^2_{H^1(k)} \right)$$

(Lemma 7.6)

Finite covering argument einmal ausführlich

Lemma 7.6 →

$$\le 2 \left(|u_0|^2_{H^1(\Omega)} + C \sum_{k \in J_h} \underbrace{|u_0|^2_{H^1(\omega_k)}}_{\sum_{k' \subset \omega_k} |u_0|^2_{H^1(k')}} \right)$$

Finite covering $k' \subset \omega_k$ für endlich viele

$$\le C |u_0|^2_{H^1(\Omega)}$$

$$\le C |\tilde{I}^H u_H|^2_{H^1(\Omega)}$$

Lemma 7.5

Finite covering Argument auf $\mathcal{T}_H$ →

$$\le C |u_h|^2_{H^1(\Omega)}$$

Elliptizität →

$$\le C\, a(u_h, u_h) = C \langle x, x \rangle_A .$$

10. V. 09
24

(iv) Teilgebiete $1 \le i \le p$

$$a(u_i,u_i) \le c\,|u_i|^2_{H^1(\hat\Omega_i)} \overset{\text{def}}{=} c\,|I^h(\theta_i w)|^2_{H^1(\hat\Omega_i)}$$

($\theta_i w$: quadr. auf Kette)

Lemma 7.7 $\rightarrow \le c\,|\theta_i w|^2_{H^1(\hat\Omega_i)}$

$$= c\int_{\hat\Omega_i} \nabla(\theta_i w)\cdot\nabla(\theta_i w)\,dx$$

Produktregel

$$\le c\int_{\hat\Omega_i} \|w\,\nabla\theta_i\|_2^2 + \|\theta_i\nabla w\|_2^2\,dx$$

① ②

Produktregel

$$\nabla(uv)\cdot\nabla(uv) = \sum_{j=1}^d (\partial_j(uv))^2$$
$$= \sum_{j=1}^d (\partial_j u)v + u\partial_j v)^2$$
$$\le \sum_{j=1}^d (2(\partial_j u)^2 v^2 + 2u^2(\partial_j v)^2$$
$$\le c\left(\|v\nabla u\|_2^2 + \|u\nabla v\|_2^2\right)$$

(v) Zwischenresultat. Für $s=0,1$ gilt:

$$|w|^2_{H^s(\hat\Omega_i)} \overset{\text{Def. } w}{=} |u_h - I^h u_0|^2_{H^s(\hat\Omega_i)}$$

$$= |u_h - u_0 + u_0 - I^h u_0|^2_{H^s(\hat\Omega_i)}$$

$u_0 = \tilde I^H u_h \rightarrow$

$$\le 2\,|u_h - \tilde I^H u_h|^2_{H^s(\hat\Omega_i)} + 2\,|u_0 - I^h u_0|^2_{H^s(\hat\Omega_i)}$$

(a) (b)

$h_K = \frac{h_K}{H_K} H_K$

(b) $= 0$ für $V_H \subset V_h$!

$$= \sum_{\substack{K\in\mathcal{T}^h \\ K\subset\hat\Omega_i}} |u_0 - I^h u_0|^2_{H^s(K)}$$

Lemma 7.6

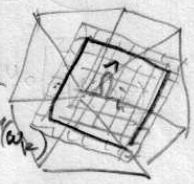
$$\le c\sum_{\substack{K\in\mathcal{T}^h \\ K\subset\hat\Omega_i}} h_K^{2(1-s)}\,|u_0|^2_{H^1(\omega_K)}$$

(a)

$$\le \sum_{K\in\mathcal{T}^H,\ K\cap\hat\Omega_i\neq 0} |u_h - \tilde I^H u_h|^2_{H^s(\hat\Omega_i)}$$

Lemma 7.5 ($s=0$: Dreiecksungl.)

$$\le c\sum_{\substack{K\in\mathcal{T}^H \\ K\cap\hat\Omega_i\neq\emptyset}} H_K^{2(1-s)}\,|u_h|^2_{H^1(\omega_K)}$$



$$\le c\sum_{K\in\mathcal{T}^H,\ K\cap\hat\Omega_i\neq 0} \sum_{\substack{k\in\mathcal{T}^h \\ k\subset\hat\Omega_i}} h_k^{2(1-s)}\,|u_0|^2_{H^1(\omega_k)}$$

(evtl. mehrfach gezählt)

$$\le c\sum_{K\in\mathcal{T}^H,\ K\cap\hat\Omega_i\neq 0} \sum_{\substack{k\in\mathcal{T}^h \\ k\cap K\neq\emptyset}} h_K^{2(1-s)}\,|u_0|^2_{H^1(\omega_K)}$$

$h_K \le H_K$, Finite Überdeckung

$$\le c\sum_{K\in\mathcal{T}^H,\ K\cap\hat\Omega_i\neq 0} H_K^{2(1-s)}\,|\tilde I^H u_h|^2_{H^1(K)}$$

Lemma 7.5(2)

$$\le c\sum_{K\in\mathcal{T}^H,\ K\cap\hat\Omega_i\neq 0} H_K^{2(1-s)}\,|u_h|^2_{H^1(\omega_K)}$$

also

$$|w|^2_{H^s(\hat\Omega_i)} \le c\sum_{\substack{K\in\mathcal{T}^H \\ K\cap\hat\Omega_i\neq 0}} H_K^{2(1-s)}\,|u_h|^2_{H^1(\omega_K)}$$

10. V. 09
25

(vi) nun weiter mit (iv)

(2) $\int_{\hat\Omega_i} \|\theta_i \nabla w\|_2^2 \, dx \le \int_{\hat\Omega_i} \|\nabla w\|_2^2 \, dx = |w|^2_{H^1(\hat\Omega_i)}$ ($\theta_i \le 1$)

$\overset{(v)}{s=1} \le C \sum_{\substack{K \in \mathcal{T}_H \\ K \cap \hat\Omega_i \neq 0}} |u_h|^2_{H^1(\omega_K)}$

(1) $\int_{\hat\Omega_i} \|w \nabla\theta_i\|_2^2 \, dx = \int_{\hat\Omega_{i,\delta_i}} |w|^2 \underbrace{\|\nabla\theta_i\|_2^2}_{\le (\frac{C}{\delta_i})^2} dx \le \left(\frac{C}{\delta_i}\right)^2 \|w\|^2_{L^2(\hat\Omega_{i,\delta_i})}$

($\nabla\theta_i = 0$, da $\theta_i = 1$ im Inneren; Skalar ranziehen)

Lemma 7.8 Spursatz $\le \frac{C}{\delta_i^2} \delta_i^2 \left((1+\frac{H}{\delta_i}) |w|^2_{H^1(\hat\Omega_i)} + \frac{1}{H_i \delta_i} \|w\|^2_{L^2(\hat\Omega_i)} \right)$

$\overset{(v)}{s=1}$ bzw. $\overset{(v)}{s=0}$

$\le C \left(C (1+\frac{H_i}{\delta_i}) \sum_{\substack{K \in \mathcal{T}_H \\ K \cap \hat\Omega_i \neq 0}} |u_h|^2_{H^1(\omega_K)} + \frac{1}{H_i \delta_i} C'' \sum_{\substack{K \in \mathcal{T}_H \\ K \cap \hat\Omega_i \neq \emptyset}} H_K^2 |u_h|^2_{H^1(\omega_K)} \right)$

local quasi-uniform: $\le H_i \le H_K \le \bar C H_i$

$\le C \left((1+\frac{H_i}{\delta_i}) \sum_{\dots} |u_h|^2_{H^1(\omega_K)} + \frac{H_i}{\delta_i} \sum_{\dots} |u_h|^2_{H^1(\omega_K)} \right)$

finite overlap

$\le C (1+\frac{H_i}{\delta_i}) \sum_{\substack{K \in \mathcal{T}_H \\ K \cap \hat\Omega_i \neq 0}} |u_h|^2_{H^1(\omega_K)}$

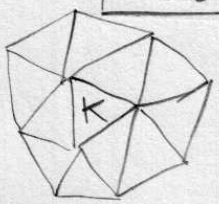
(1) + (2) liefert nun in Fortsetzung von (iv):

$$a(u_i, u_i) \le C (1+\frac{H_i}{\delta_i}) \sum_{\substack{K \in \mathcal{T}_H \\ K \cap \hat\Omega_i \neq 0}} |u_h|^2_{H^1(\omega_K)}$$

10.V.09
26

(vii) Summe über Subdomains:

$$\sum_{i=1}^{p} a(u_i,u_i) \le \sum_{i=1}^{p} C\left(1+\frac{H_i}{\delta_i}\right) \sum_{\substack{K\in\mathcal{T}^H \\ K\cap\tilde{\Omega}_i\neq 0}} |u_{2h}|^2_{H^1(\omega_K)}$$



$\frac{H_i}{\delta_i} \sim \frac{H}{\delta}$

$$\le C\left(1+\frac{H}{\delta}\right) \sum_{K\in\mathcal{T}^H} |u_{2h}|^2_{H^1(\omega_K)}$$

$K\in\mathcal{T}^H$ kommt nur N mal vor unabh. von p (Finite Covering)

$$\le C\left(1+\frac{H}{\delta}\right) \sum_{K\in\mathcal{T}^H} |u_{2h}|^2_{H^1(K)} = C\left(1+\frac{H}{\delta}\right) |u_{2h}|^2_{H^1(\Omega)}$$

shape regularity von $\mathcal{T}^H$, $K\in\mathcal{T}^H$ kommt nur in konstant vielen ω_K vor

$|u_{2h}|^2 = a(u_{2h},u_{2h})$

$$\le C\left(1+\frac{H}{\delta}\right) a(u_{2h},u_{2h}) = C\left(1+\frac{H}{\delta}\right) \langle x,x\rangle_A .$$

Zusammen mit (iii) gilt also für $u_{2h} = I^h u_0 + \sum_{i=1}^{p} u_i$ entsprechend $x = \sum_{i=0}^{p} R_i^T x_i$

$$\sum_{i=0}^{p} \langle R_i^T x_i, R_i^T x_i\rangle_A = a(I^h u_0, I^h u_0) + \sum_{i=1}^{p} a(u_i,u_i)$$

(iii)

$$\le C\langle x,x\rangle_A + C'\left(1+\frac{H}{\delta}\right)\langle x,x\rangle_A$$

$H \le \delta$ ausführlich

$$\le C''\langle x,x\rangle_A + C'\frac{H}{\delta}\langle x,x\rangle_A$$

$$\le \max\{C',C''\}\left(1+\frac{H}{\delta}\right)\langle x,x\rangle_A$$

10. V. 09
27

Und damit finally

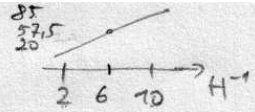
Satz 7.12

Für das additive Schwarz-Verfahren mit Grobgitterkorrektur gilt

$$\kappa\left(\sum_{i=0}^{p} P_i\right) \leq C\left(1+\frac{H}{\delta}\right)$$

10^{-8} Reduktion

Skalierbarkeit in H.



1) variiere H bei H/h und H/δ fest

H^{-1}	1	2	3	4	5	6	7	8	9	10
ohne GK	1	20	30	38	46	55	62	70	77	85
mit GK	1	18	23	(24)	25	25	25	25	26	25

$h = H/2^6$ overlap $\delta = 4h$

2) Variiere overlap $\frac{H=1/4}{(P=16)}$, $h = H/2^6$

overlap	1h	2h	4h	8h	16h
ohne GK	74	53	38	29	20
mit GK	43	32	(24)	19	16

← geringerer Anstieg mit GK.

3) Skalierbarkeit in h: H/δ fest, H=1/4

H/h	8	16	32	64	128
ovlp	1	2	4	8	16
IT ohne GK	28	29	29	29	28
mit GK	20	20	20	19	19

4) Minimum overlap, H=1/4

ovlp $\delta = h$, d.h. $\frac{H}{\delta} = \frac{H}{H/2^e} = 2^e$

H/h	8	16	32	64	128
ohne GK	28	38	[53]	(74)	102
mit GK	20	25	[33]	(43)	56

ovlp $\delta = 2h$, d.h. $\frac{H}{\delta} = \frac{H}{H/2^{e-1}} = 2^{e-1}$

H/h	8	16	32	64	128
ohne GK	20	29	38	[53]	(73)
mit GK	17	20	25	[32]	(43)

8 Mehrgitterverfahren

8.1 Gitterhierarchie, geschachtelte FE-Räume

Das feine Gitter $\mathcal{T}^l$, $l > 0$ entstehe durch regelmäßige Verfeinerung eines groben Gitters $\mathcal{T}^0$, wie in Abbildung 8.1 dargestellt.

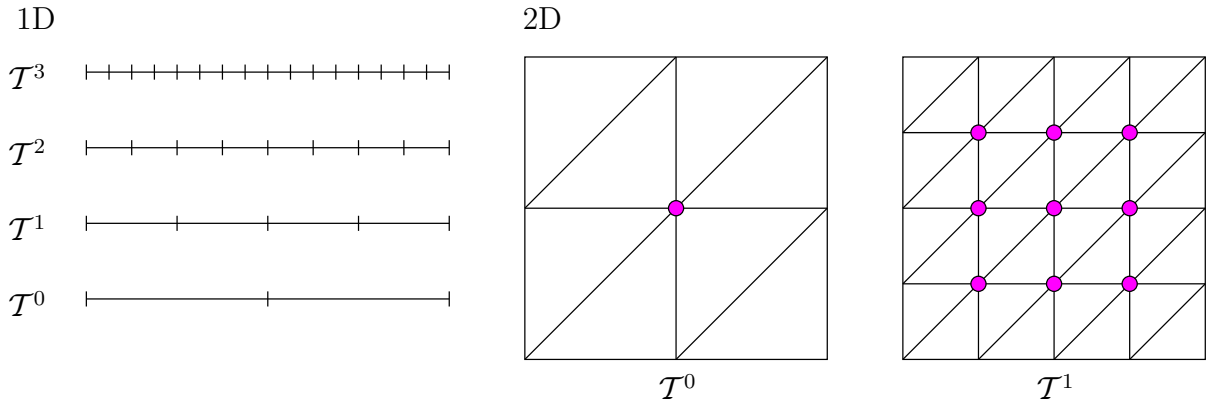


Abbildung 8.1: Regelmäßige Verfeinerung von Gittern

Die Stufen bezeichnen wir mit

$$l = 0, 1, \ldots, L,$$

die Gitter mit $\mathcal{T}^0, \ldots, \mathcal{T}^L$. Der Bezug zur bisherigen Notation ist gegeben durch (der Level-Index steht oben, um Verwechslung mit dem Teilgebiet zu vermeiden)

$$\mathcal{T}^H = \mathcal{T}^0, \;\; \mathcal{T}^h = \mathcal{T}^L,$$

nur jetzt betrachten wir auch explizit alle Zwischenstufen.

Zu jedem $\mathcal{T}^l$, mit $l \in \{0, \ldots, L\}$ gehört eine Indexmenge:

$$I^l, \quad \text{Indexmenge der Knoten der Stufe } l.$$

Die Nummerierung verschiedener Stufen ist konsistent, d.h. Gitterpunkt (x_i, y_j), $i \in \mathcal{T}^l$ hat auch Nummer i in allen $\mathcal{T}^k$ mit $k > l$.

Die Standard-Knotenbasis auf Stufe l ist

$$\Phi^l = \{\varphi_k^l \mid k \in I^l\}$$

mit dem entsprechenden FE-Raum

$$V^l = \operatorname{span} \Phi^l.$$

Es gilt

$$V_H = V^0 \subseteq V^1 \subseteq \cdots \subseteq V^{L-1} \subseteq V^L = V_h \subset H_0^1(\Omega)$$

Jede Basisfunktion der Stufe l kann durch eine Linearkombination von Basisfunktionen der Stufe $m > l$ ausgedrückt werden. Für $m = l + 1$ und $m = L$ führen wir folgende Schreibweise ein:

$$\varphi_k^l = \sum_{j \in I^L} \Theta_{k,j}^l \varphi_j^L \qquad \text{Darstellung auf Stufe } L \tag{8.1}$$

$$\varphi_k^l = \sum_{j \in I^{l+1}} \theta_{k,j}^l \varphi_j^{l+1} \qquad \text{Darstellung auf Stufe } l+1 \tag{8.2}$$

Siehe dazu Abbildung 8.2.

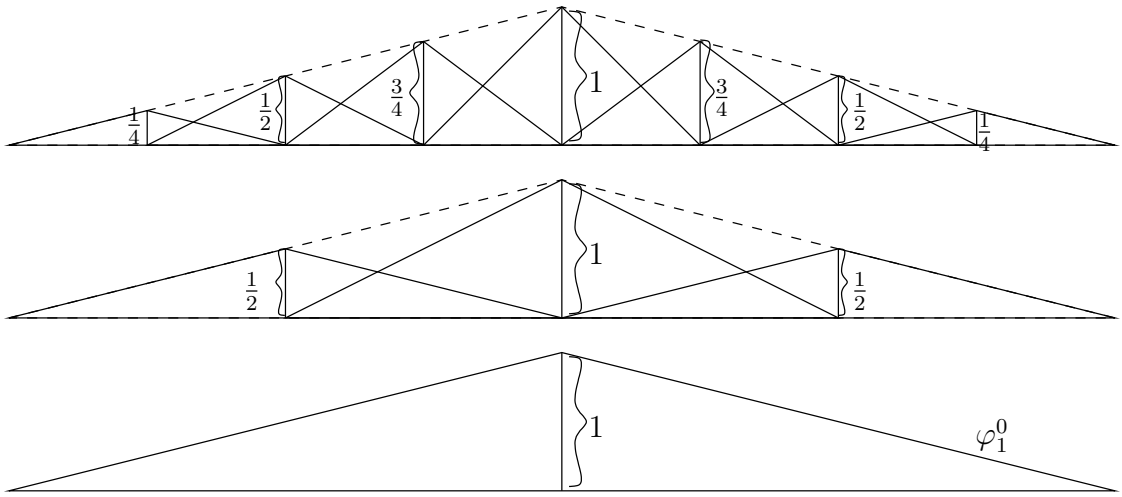


Abbildung 8.2: Darstellung höherer Stufen in 1D

Bemerkung 8.1 Es gilt

$$\Theta_{k,j}^l \neq 0 \quad \Longleftrightarrow \quad \mathcal{T}^L \ni (x_j, y_j) \in \operatorname{supp} \varphi_k^l$$

d.h. $O(2^{(L-l)d})$ Koeffizienten sind von Null verschieden. *Aber*

$$\theta_{k,j}^l \neq 0 \quad \Longleftrightarrow \quad \mathcal{T}^{l+1} \ni (x_j, y_j) \in \operatorname{supp} \varphi_k^l$$

sind nur $O(3^d)$ viele.

8.2 Abstrakte Formulierung von Teilraumkorrekturverfahren

Wir formulieren nun lineare Iterationsverfahren in Finite-Elemente-Räumen statt im $\mathbb{R}^I$.

Die Funktion u_h sei Lösung der Variationsgleichung in $V_h = V^L$

$$a(u_h, w) = (f, w)_{L^2(\Omega)} \quad \forall w \in V_h. \tag{8.3}$$

Mittels Bilinearformen kann man auf folgende Weise Operatoren definieren: Definiere $\mathcal{A}_h \colon V_h \to V_h$ mittels

$$(\mathcal{A}_h v, w)_{L^2(\Omega)} = a(v, w) \quad \forall v, w \in V_h \tag{8.4}$$

und die L^2-Projektion $\mathcal{Q}_h \colon L^2(\Omega) \to V_h$ mittels

$$(\mathcal{Q}_h v, w)_{L^2(\Omega)} = (v, w)_{L^2(\Omega)} \quad \forall v \in L^2(\Omega), w \in V_h. \tag{8.5}$$

Mittels der Abkürzung $f_h := \mathcal{Q} f$ ist

$$\mathcal{A}_h u_h = f_h \tag{8.6}$$

die äquivalente *Operatorform* zu (8.3), denn mit (8.4) und (8.5) gilt:

$$\begin{aligned} \mathcal{A}_h u_h = f_h \quad &\underset{\text{endlich dimensional}}{\Leftrightarrow} \quad (\mathcal{A}_h u_h, w)_{L^2(\Omega)} = (f_h, w)_{L^2(\Omega)} \quad \forall w \in V_h \\ &\underset{(8.4),(8.5)}{\Leftrightarrow} \quad a(u_h, w) = (f, w)_{L^2(\Omega)} \quad \forall w \in V_h \end{aligned}$$

Bemerkung 8.2 Da $\mathcal{A}_h \colon V_h \to V_h$ müssen wir mittels $\mathcal{Q}_h$ das $f \in L^2(\Omega)$ nach V_h projezieren. $\mathcal{Q}_h$ spielt dabei die Rolle einer Restriktion.

Wir formulieren nun, was eine *Teilraumkorrektur* ist. Sei dazu $u_h^{alt} \in V_h$ eine gegebene Näherung von u_h, die korrigiert werden soll und $V_i \subseteq V^l \subseteq V^L$ ein Teilraum von V^l. Beachte, dass $0 \leq l \leq L$ erlaubt ist und dass V_i nur ein Teil von V_l sein darf!

Es ist nun eine Korrektur $v_i \in V_i$ mittels der Gleichung

$$\boxed{a(u_h^{alt} + v_i, w) = (f, w)_{L^2(\Omega)} \quad \forall w \in V_i} \tag{8.7}$$

zu bestimmen. Falls $V_i = V_h$, so würde $u_h^{alt} + v_i = u_h$ gelten (siehe (8.3)). Ist $V_i \subset V_h$, so ist $u_h^{alt} + v_i$ eine verbesserte Näherung.

Analog zu oben wollen wir (8.7) in Operatorform schreiben. Dazu definieren wir $\mathcal{A}_i \colon V_i \to V_i$ mittels

$$(\mathcal{A}_i v, w)_{L^2(\Omega)} = a(v, w) \quad \forall v, w \in V_i \tag{8.8}$$

sowie $\mathcal{Q}_i \colon L^2(\Omega) \to V_i$ mittels

$$(\mathcal{Q}_i v, w)_{L^2(\Omega)} = (v, w)_{L^2(\Omega)} \quad \forall v \in L^2(\Omega), w \in V_i. \tag{8.9}$$

Somit gilt

$$\begin{aligned} & a(v_i, w) = (f, w)_{L^2(\Omega)} - a(u_h^{alt}, w) \quad \forall w \in V_i \quad (8.7) \\ \iff \quad & (\mathcal{A}_i v_i, w)_{L^2(\Omega)} = (\mathcal{Q}_i f_h, w)_{L^2(\Omega)} - \left(\mathcal{Q}_i \mathcal{A}_h u_h^{alt}, w\right)_{L^2(\Omega)} \\ & \qquad = \left(\mathcal{Q}_i (f_h - \mathcal{A}_h u_h^{alt}), w\right)_{L^2(\Omega)} \quad \forall w \in V_i \\ \iff \quad & \mathcal{A}_i v_i = \mathcal{Q}_i (f_h - \mathcal{A}_h u_h^{alt}) \end{aligned}$$

also insgesamt:

$$\boxed{u_h^{neu} = u_h^{alt} + \mathcal{A}_i^{-1} \mathcal{Q}_i (f_h - \mathcal{A}_h u_h^{alt})} \tag{8.10}$$

Bemerkung 8.3 $f_h - \mathcal{A}_h u_h^{alt}$ ist der Defekt als FE-Funktion!

Mehrere Teilraumkorrekturen können nun zu einem Gesamtiterationsverfahren kombiniert werden. Dazu wählen wir M Teilräume:

$$V_i \subseteq V_h, i \in \{1, \ldots, M\}.$$

Als notwendige Bedingung für Konvergenz des Verfahrens fordert man

$$V_1 + V_2 + \cdots + V_m = V_h,$$

ansonsten kann die Zerlegung völlig beliebig sein.

Die *Kombination* der Teilraumkorrekturen kann sequentiell oder parallel erfolgen, entsprechend erhält man das multiplikative oder additive Teilraumkorrekturverfahren:

Algorithmus 8.4 (Multiplikative Teilraumkorrektur) Gegeben Zerlegung $V_1, \ldots, V_M$ und Startwert u_h^0

for $(k = 0, 1, \ldots)$
  for $(i = 1, \ldots, M)$
    $u_h^{k+\frac{i}{M}} = u_h^{k+\frac{i-1}{M}} + \mathcal{A}_i^{-1} \mathcal{Q}_i (f_h - \mathcal{A}_h u_h^{k+\frac{i-1}{M}})$

Algorithmus 8.5 (Additive Teilraumkorrektur) Gegeben Zerlegung $V_1, \ldots, V_M$, und Startwert u_h^0

for $(k = 0, 1, \ldots)$
  $u_h^{k+1} = u_h^k + \sum_{i=1}^M \mathcal{A}_i^{-1} \mathcal{Q} (f_h - \mathcal{A}_h u_h^k)$

Die abstrakte Formulierung der Teilraumkorrekturverfahren ist sehr mächtig. Wir können alle bisher behandelten Verfahren (Jacobi, Gauß-Seidel, Blockvarianten, überlappende Schwarz-Verfahren) als solche formulieren.

Dies zeigen wir am Beispiel des Jacobi-Verfahrens: Wähle

$$V_i = \operatorname{span}\{\varphi_i^L\}, \ i \in I^L$$

und betrachte die Korrektur v_i für ein $i \in I^L$:

$$\begin{aligned} & \mathcal{A}_i v_i = \mathcal{Q}_i(f_h - \mathcal{A}_h u_h^k) \\ \iff \quad & a(v_i, \varphi_i^L) = \left(f_h, \varphi_i^L\right)_{L^2(\Omega)} - a(u_h^k, \varphi_i^L) \\ \iff \quad & (y)_i \underbrace{a(\varphi_i^L, \varphi_i^L)}_{\substack{\text{Diagonal-}\\ \text{element}\\ \text{von } A}} = \underbrace{(b)_i - \sum_{j \in I^L} (x^k)_j a(\varphi_j^L, \varphi_j^L)}_{\substack{i\text{-te}\\ \text{Komponente}\\ \text{von}\\ d = b - Ax^k!}} \end{aligned}$$

wenn man $v_i = (y)_i \varphi_i^L$, und $u_h^k = \sum_{j \in I^L} (x^k)_j \varphi_j^L$ schreibt. Somit ist $v = \sum_{i \in I^L} v_i = \sum_{i \in I^L} (y)_i \varphi_i^L$ und $u = D^{-1}(b - Ax^k)$ mit $D = \operatorname{diag}(A)$. Also entspricht dem Jacobi-Verfahren auf $Ax - b$ nichts anderes als die Wahl

- $V_i = \operatorname{span}\{\varphi_i^L\}$
- additive Kombination der $v_i \in V_i$

Das Gauß-Seidel-Verfahren ergibt sich durch multiplikative Kombination.

Bemerkung 8.6 Eine Darstellun von $\mathcal{A}_i$ ergibt sich durch einsetzen einer Basis von V_i. Das abstrakte Verfahren (8.10) ist „basisfrei“

8.3 Beispiele für Teilraumkorrekturverfahren

- *Jacobi und Gauß-Seidel*

$$V_i = \operatorname{span}\{\varphi_i^L\},\ i \in I^L$$

  additiv → Jacobi
  multiplikativ → Gauß-Seidel

- *Block-Jacobi, Block-Gauß-Seidel*

$$I^L = I = \bigcup_{i=1}^{M} I_i, I_i \cap I_j = \emptyset \text{ falls } i \neq j \text{ (nicht überlappende Zerlegung)}$$

$$V_i = \operatorname{span}\{\varphi_j^L \mid j \in I_i\}$$

  additiv → Block-Jacobi
  multiplikativ → Block-Gauß-Seidel

- *Eingitter-Schwarz*

$$V_i = \operatorname{span}\{\varphi_j^L \mid j \in \hat{I}_i\}, i = 1, \ldots, p$$

  additiv → additiver Schwarz
  multiplikativ → multiplikativer Schwarz

- *Zweigitter-Schwarz*

$$\begin{aligned} V_i & \quad \text{für } i = 1, \ldots, p \text{ wie oben,} \\ V_0 & = \operatorname{span}\{\varphi_j^0 \mid j \in I^0\} \end{aligned}$$

Unter Zuhilfenahme der $V^1, \ldots, V^{L-1}$ erhalten wir folgende Mehrgittervarianten:

- *Hierarchische Basis*
  Wir nutzen folgende *nicht überlappende* Zerlegung:

$$V^L = V^0 + \sum_{l=1}^{L} \sum_{i \in I^l \setminus I^{l-1}} \overbrace{\operatorname{span}\{\varphi_i^l\}}^{V_{l,i}}$$

  additive Kombination → hierarchische Basis Verfahren HB
  multiplikative Kombination → hierarchische Basis Mehrgitter HBM
  Die Konvergenzrate dieser Verfahren hat die Form $\rho = 1 - \frac{1}{O(L)}$ in 2D, aber leider $\rho = 1 - O(h)$ in 3D.

- *Additives Mehrgitter*
  Überlappende Zerlegung:

$$V^L = V^0 + \sum_{l=1}^{L} \sum_{i \in I^l} \operatorname{span}\{\varphi_i^l\}$$

  In Operatorform ergibt sich die Darstellung:

$$u_h^{k+1} = u_h^k + \mathcal{A}_0^{-1} \mathcal{Q}_0 (f_h - \mathcal{A}_h u_h^k) + \sum_{l=1}^{L} \sum_{i \in I^l} \frac{\left(f_h - \mathcal{A}_h u_h^k, \varphi_i^l\right)_{L^2(\Omega)}}{a(\varphi_i^l, \varphi_i^l)} \cdot \varphi_i^l, \qquad (8.11)$$

  daher heisst das Verfahren auch Multilevel Diagonal Scaling.

  Die Näherung $a(\varphi_i^l, \varphi_i^l) \approx h_l^{d-2}$ (für $-\Delta u = f$ auf quasiuniformem Gitter) ergibt das *BPX*-Verfahren.

  Die Konvergenzrate ist unabhängig von h.

- *klassisches (multiplikatives) Mehrgitterverfahren*
  Mit $n^l = \left|I^l\right|$ setze

$$\begin{array}{ccccccccccc} V^L = & \operatorname{span}\{\varphi_1^L\} & + \ldots + & \operatorname{span}\{\varphi_{n^L}^L\} & + \operatorname{span}\{\varphi_1^{L-1}\} & + \ldots + & \operatorname{span}\{\varphi_{n^1}^1\} & + & V_0 \\ & \| & & & & & \| & & \| \\ & V_1 & & & & & V_{M-1} & & V_M \end{array}$$

  und kombiniere multiplikativ.

  Dies ist ein klassisches Mehrgitterverfahren mit einem Schritt Gauß-Seidel zur Vorglättung.

Die so definierte Iteration ist nicht symmetrisch. Dies benötigt man aber, um die Iteration innerhalb eines Gradienten- oder CG-Verfahrens einzusetzen. Hier bietet sich die *symmetrische* Variante an:

$$V^L = \underbrace{V_1 + V_2 + \cdots + V_{M-1}}_{\substack{\text{wie oben}\\ \text{definiert}}} + \underbrace{V_M}_{\substack{= V^0,\\ \text{Grob-}\\ \text{gitter-}\\ \text{problem}}} + V_{M-1} + \cdots + V_1$$

Die Konvergenzrate ist unabhängig von h.

8.4 Klassische Formulierung von Mehrgitterverfahren

Wir gehen nun den umgedrehten Weg und leiten aus der Operatorform eine Matrixformulierung her.

Dies geschieht natürlich durch Einsetzen einer Basis.

Wir betrachen zunächst das MDS-Verfahren aus (8.11). Die auf Stufe $1 \leq l \leq L$ berechnete Korrektur lautet

$$v^l = \sum_{i \in I^l} (y^l)_i \varphi_i^l, \quad (y^l)_i = \frac{\left(f_h - \mathcal{A}_h u_h^k, \varphi_i^l\right)_{L^2(\Omega)}}{a(\varphi_i^l, \varphi_i^l)}. \tag{8.12}$$

Offensichtlich ist

$$y^l = \left(D^l\right)^{-1} d^l, \quad D^l = \operatorname{diag}(A^l),\ (d^l)_i = \left(f_h - \mathcal{A}_h u_h^k, \varphi_i^l\right)_{L^2(\Omega)} \tag{8.13}$$

mit A^l der Steifigkeitsmatrix auf Stufe l.

Betrachten wir den Vektor d^l näher. Offensichtlich ist $d_h = f_h - \mathcal{A}_h u_h^k \in V^L$ eine Feingitterfunktion. φ_i^l ist hingegen eine Funktion der Stufe $1 \leq l \leq L$. Setzen wir für φ_i^l die Darstellung (8.1) ein, so ergibt sich

$$(d^l)_i = \left(f_h - \mathcal{A}_h u_h^k, \varphi_i^l\right)_{L^2(\Omega)} = \sum_{j \in I^L} \Theta_{i,j}^l \left(f_h - \mathcal{A}_h u_h^k, \varphi_j^L\right)_{L^2(\Omega)} \tag{8.14}$$

Mit der Darstellung $u_h^k = \sum_{m \in I^L} (x^k)_m \varphi_m^L$ erhalten wir

$$(d^l)_i = \cdots = \sum_{j \in I^L} \Theta_{i,j}^l \Big[\underbrace{\left(f_h, \varphi_j^l\right)_{L^2(\Omega)}}_{(b)_j} - \underbrace{\sum_{m \in I^L} (x^k)_m a(\varphi_m^l, \varphi_j^L)}_{(A \cdot x^k)_j} \Big] \tag{8.15}$$

Die $\Theta_{i,j}^l$ bilden die Koeffizienten einer rechteckigen Restriktionsmatrix $\mathbf{r}^l \colon \mathbb{R}^{I^L} \to \mathbb{R}^{I^l}$ und es ist

$$y^l = D^{-1} \mathbf{r}^l (b - A x^k) \tag{8.16}$$

mit $Ax = b$ dem Gleichungssystem auf Stufe L. Für die gröbste Stufe erhält man die Formel

$$y^0 = (A^0)^{-1} \mathbf{r}^0 (b - A x^k) \tag{8.17}$$

mit $(A^0)_{i,j} = a(\varphi_i^0, \varphi_j^0)$ der Diskretisierung auf Stufe 0.

Schreiben wir v^l in der Basis von Stufe L ergibt sich

$$v^l = \sum_{i \in I^l} (y^l)_i \varphi_i^l \quad \underset{(8.1)}{=} \quad \sum_{i \in I^l} \sum_{j \in I^l} (y^l)_i \Theta_{i,j}^l \varphi_j^l \tag{8.18}$$

In der Basis Φ^L hat v^l die Koeffizienten $(\mathbf{r}^l)^T y^l$! Die MDS-Iteration in Matrixdarstellung lautet:

$$x^{k+1} = x^k + (\mathbf{r}^0)^T (A^0)^{-1} \mathbf{r}^0 (b - Ax^k) + \sum_{l=1}^{L} (\mathbf{r}^l)^T (D^l)^{-1} \mathbf{r}^l (b - Ax^k) \tag{8.19}$$

Bemerkung 8.7 Der zweite Term in (8.19) entspricht exakt der Grobgitterkorrektur im Schwarz-Verfahren, d.h.

$$\begin{aligned} \mathbf{r}^0 &= R_H, \quad A^0 = A_H \text{ !} \\ A^0 &= \mathbf{r}^0 A (\mathbf{r}^0)^T \text{ sieht man durch Einsetzen von (8.1) in } a(\cdot,\cdot). \end{aligned}$$

Allerdings ist (8.19) wegen Bemerkung 8.1 nicht effizient. Die Berechnung von $\mathbf{r}^l(b - Ax^k)$ erfordert $O(N^L)$ Operationen *unabhänging* von l. Die Berechnung *aller* $\mathbf{r}^l(b - Ax^k)$, $l = 0, \ldots, L$ braucht somit $O(n^L \log n^L)$ Operationen. Dies geht besser:

Zunächst bemerken wir, dass

$$\left(f_h - \mathcal{A}_h u_h^k, \varphi_j^l\right)_{L^2(\Omega)} = (b - Ax^k)_j.$$

Die Berechnung aller

$$\left(f_h - \mathcal{A}_h u_h^k, \varphi_i^l\right)_{L^2(\Omega)} = \sum_{j \in I^{l+1}} \theta_{i,j}^l \left(f_h - \mathcal{A}_h u_h^k, \varphi_j^{l+1}\right)_{L^2(\Omega)}$$

für $i \in I^l$, $l < L$, braucht mittels (8.7) genau $O(n^{l+1})$ Operationen.

Bei Stufenweiser Berechnung der restringierten Defekte ist die Gesamtkomplexität

$$n^L + n^{L-1} + \cdots + n^0 = n^L (1 + \frac{1}{\eta} + \frac{1}{\eta^2} + \cdots + \frac{1}{\eta^L}) \leq n^L \frac{\eta}{\eta - 1},$$

wobei wir geometrisches Wachstum $\frac{n^{l+1}}{n^l} = \eta$ vorausgesetzt haben. In 2D gilt $\eta = 4$ in 3D $\eta = 8$.

Wir können die $\theta_{i,j}^l$ auch als Restriktion $R^l \colon \mathbb{R}^{I^{l+1}} \to \mathbb{R}^{I^l}$ deuten. Offensichtlich ist $\mathbf{r}^l = R^l R^{l+1} \ldots R^{L-1}$.

Damit kommen wir zu

Algorithmus 8.8 (additives MG)

amgc $(x^k \in \mathbb{R}^{I^L})$
{
$d^L = b - Ax^k \in I^L$

```
    for (l = L − 1; l ≥ 0; l = l − 1)   // sequentiell
        d^l = R^l d^{l+1};               //  über Level
    ∀l ∈ {1, ..., L}   v^l = ( underbrace(D^l, diag(A^l)) )^{-1} d^l;  // alle Level
    v^0 = (A^0)^{-1} d^0;                //  gleichzeitig
    for (l = 1; l ≤ L; l = l + 1)        // sequentiell
        v^l = v^l + (R^{l-1})^T v^{l-1}; //  über Level
    return x^{k+1} = x^k + v^L;
}
```

Das multiplikative Mehrgitterverfahren in algorithmischer Form ergibt sich durch Berechnung neuester Defekte auf jeder Stufe:

Algorithmus 8.9 (multiplikatives MG)

```
mmgc (x^k ∈ R^{I^L})
{
    d^L = b^L − A^L x^k ∈ I^L;
    for (l = L; l > 0; l = l − 1)
    {
        v^l = 0;                         // wichtig, Startwert 0
        for (m = 0; m < ν_1; m = m + 1)  Vorglättungsiteration
        {
            y^l = (W^l)^{-1} d^l;        // z.B. W^l = L(A^l)
            v^l = v^l + y^l;
            d^l = d^l − A^l y^l;
        }
        d^{l-1} = R^{l-1} d^l;           // Restriktion
    }
    v^0 = (A^0)^{-1} d^0;                // Grobgitter exakt lösen
    for (l = 1; l ≤ L; l = l + 1)
    {
        y^l = (R^{l-1})^T v^{l-1};       // Prolongation
        v^l = v^l + y^l;
        d^l = d^l − A^l y^l;
        for (m = 0; m < ν_2; m = m + 1)  Nachglättung
            {
            y^l = (W^l)^{-1} d^l;
            v^l = v^l + y^l;
            d^l = d^l − A^l y^l;
        }
    }
    return x^{k+1} = x^k + v^L;
}
```

Bemerkung 8.10

- Oben wurde nur $W^l = L(A^l)$, $\nu_1 = \nu_2 = 1$ behandelt.
- Dies ist ein sogenannter V-Zyklus.

Satz 8.11 Die Konvergenzrate von additivem und multiplikativem Mehrgitterverfahren ist unabhängig von h, H (aber abhängig von den Koeffizienten der DGL)

BEWEIS: siehe (SMITH, BJØRSTAD und GROPP 1996), (HACKBUSCH 1991). Wir wollen das hier nicht machen, da es nicht spezifisch parallel ist. □

8.5 Parallele Implementierung von MG-Verfahren

Das Mehrgitter-Verfahren ist kein spezifisch paralleles Verfahren, es lässt sich jedoch durch eine entsprechende Datenverteilung gut parallelisieren. Wir zeigen mehrere Möglichkeiten.

Interessant ist vor allem die Parallelisierung der Gittertransformation. Wir beginnen mit einer allgemeineren Darstellung und wenden diese auf verschiedene Situationen an.

Es sei I^l die Indexmenge der Stufe l, $l = 0, \ldots, L$. Wir betrachten folgende Datenaufteilung

(i). Für jedes l sei $\{\tilde{I}^l\}$ eine überlappende Zerlegung

$$I^l = \bigcup_{i=1}^{p} \tilde{I}_i^l$$

in p Teilmengen.

(ii). Die $\tilde{I}_i^l$, $\tilde{I}_i^{l+1}$ erfüllen für alle $l = 0, \ldots, L-1$ (siehe Abbildung 8.3):

$$j \in \tilde{I}_i^{l+1} \wedge \theta_{k,j}^l \neq 0 \;\Rightarrow\; k \in \tilde{I}_i^l$$

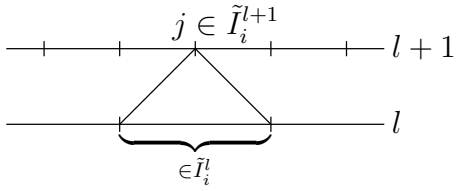


Abbildung 8.3: Überlappende Zerlegung für das MG-Verfahren

Wir benötigen nun verschiedene Restriktionen. Mit $\tilde{R}_i^l \colon \mathbb{R}^{I^l} \to \mathbb{R}^{\tilde{I}_i^l}$ bezeichnen wir die Restriktion auf das „Teilgebiet“ i, d.h. für $x^l \in \mathbb{R}^{I^l}$ ist

$$(\tilde{R}_i^l x^l)_j = (x^l)_j \quad \forall j \in \tilde{I}_i^l$$

wie bei den Schwarz-Verfahren.

Unter $R^l\colon \mathbb{R}^{I^{l+1}} \to \mathbb{R}^{I^l}$ verstehen wir die Mehrgitterrestriktion wie oben eingeführt, d.h. für $x^{l+1} \in \mathbb{R}^{I^{l+1}}$ ist

$$(R^l x^{l+1})_k = \sum_{j \in I^{l+1}} \theta^l_{k,j} (x^{l+1})_j \tag{8.20}$$

mit dem $\theta^l_{k,j}$ aus (8.2).

Die Beschränkung von R^l auf Teilgebiet i ist $R^l_i\colon \mathbb{R}^{\tilde{I}^{l+1}_i} \to \mathbb{R}^{\tilde{I}^l_i}$, und ist gegeben für $x^{l+1}_i \in \mathbb{R}^{\tilde{I}^{l+1}_i}$ durch

$$(R^l_i x^{l+1}_i)_k = \sum_{j \in \tilde{I}^{l+1}_i} \theta^l_{k,j} (x^{l+1}_i)_j. \tag{8.21}$$

Beachte: R^{l+1}_i lässt sich wegen Bedingung (ii) lokal in jedem Prozessor durchführen.

Die drei Restriktionen erfüllen für jedes $x^{l+1}_i \in \mathbb{R}^{\tilde{I}^{l+1}_i}$ folgende Beziehung

$$R^l (\tilde{R}^{l+1}_i)^T x^{l+1}_i = (\tilde{R}^l_i)^T \underbrace{R^l_i x^{l+1}_i}_{\text{lokal}} \tag{8.22}$$

oder anders

$$\begin{array}{ccc} \mathbb{R}^{\tilde{I}^{l+1}_i} & \xrightarrow[\text{lokale MG-Restr.}]{R^l_i} & \mathbb{R}^{\tilde{I}^l_i} \\ {\scriptstyle (\tilde{R}^{l+1}_i)^T}\downarrow{\scriptstyle \text{Teil-gebiets-restriktion}} & & \downarrow{\scriptstyle (\tilde{R}^l_i)^T} \\ \mathbb{R}^{I^{l+1}} & \xrightarrow[\text{globale MG-Restr.}]{R^l} & \mathbb{R}^{I^l} \end{array}$$

Beweis: Setze $x^{l+1} = (\tilde{R}^{l+1}_i)^T x^{l+1}_i$ und $x^l = R^l x^{l+1}$. Es leistet $(x^{l+1})_j$ einen Beitrag zu $(x^l)_k$ genau dann, wenn $\theta^l_{k,j} \neq 0$ (rechte Seite in (8.20)). Dies kann wegen Bedingung (ii) aber auch lokal durchgeführt werden! □

Beziehung (8.22) nutzen wir nun, um einen globalen Vektor zu restringieren. Sei $x^{l+1} \in \mathbb{R}^{I^{l+1}}$ *additiv zerlegt:*

$$x^{l+1} = \sum_{i=1}^{p} (\tilde{R}^{l+1}_i)^T x^{l+1}_i, \quad x^{l+1}_i \in \mathbb{R}^{\tilde{I}^{l+1}_i}, \tag{8.23}$$

dann gilt

$$\begin{aligned} R^l x^{l+1} &= R^l \sum_{i=1}^{p} (\tilde{R}^{l+1}_i)^T x^{l+1}_i \underset{R^l \text{ linear}}{=} \sum_{i=1}^{p} R^l (\tilde{R}^{l+1}_i)^T x^{l+1}_i = \\ &\underset{(8.22)}{=} \sum_{i=1}^{p} (\tilde{R}^l_i)^T R^l_i x^{l+1}_i. \end{aligned}$$

D.h. Ist der Defekt additiv zerlegt, so können wir lokal restringieren und erhalten, ohne Kommunikation, eine additive Zerlegung des restringierten Defekts.

Dies lässt sich natürlich über mehrere Stufen rekursiv fortsetzen, und wir erhalten für $0 \leq m \leq l$

$$R^m R^{m+1} \dots R^l x^{l+1} = \sum_{i=1}^{p} (\tilde{R}^m_i)^T R^m_i R^{m+1}_i \dots R^l_i x^{l+1}_i \tag{8.24}$$

Für die Prolongation erhält man die folgende Beziehung

$$\tilde{R}_i^{l+1}(R^l)^T x^l = (R_i^l)^T \tilde{R}_i^l x^l \tag{8.25}$$

oder als Diagramm

$$\begin{array}{ccc} \mathbb{R}^{I^l} & \xrightarrow[\text{globale Prolongation}]{(R^l)^T} & \mathbb{R}^{I^{l+1}} \\ {\scriptstyle \tilde{R}_i^l}\downarrow & & \downarrow{\scriptstyle \tilde{R}_i^{l+1}} \\ \mathbb{R}^{\tilde{I}_i^l} & \xrightarrow[\text{lokale Prolongation}]{(R_i^l)^T} & \mathbb{R}^{\tilde{I}_i^{l+1}} \end{array}$$

(8.25) folgt wieder aus der Definition der Prolongation und Bedingung (ii).

(8.25) sagt, dass der lokale Anteil der prolongierten Korrekturen aus dem lokalen Anteil der Korrekturen auf dem groben Gitter in jedem Prozessor separat berechnet werden kann.

Rekursive Anwendung liefert für $l \leq m < L$

$$\tilde{R}_i^{m+1}(R^m)^T \dots (R^{l+1})^T (R^l)^T x^l = (R_i^m)^T \dots (R_i^{l+1})^T (R_i^l)^T \tilde{R}_i^l x^l \tag{8.26}$$

Grobgitterkorrektur im Schwarz-Verfahren

Es sei gegeben

- eine Gitterhierarchie $\mathcal{T}_H = \mathcal{T}^0, \mathcal{T}^1, \dots, \mathcal{T}^L = \mathcal{T}_h$
- eine überlappende Zerlegung der Indexmengen jeder Stufe

$$I^l = \bigcup_{i=1}^{p} \tilde{I}_i^l, \quad I^l = \bigcup_{i=1}^{p} \hat{I}_i^l$$

- Forderung (ii) an die $\tilde{I}_i^l$ sei erfüllt
- zusätzlich eine nicht überlappende Zerlegung

$$I^l = \bigcup_{i=1}^{p} I_i^l \quad I_i^l \cap I_j^l = \emptyset \; \forall i \neq j \quad I_i^l \subseteq \hat{I}_i^l \subseteq \tilde{I}_i^l$$

Definiere $U_i^l \colon \mathbb{R}^{I^l} \to \mathbb{R}^{I_i^l}$ mittels

$$(U_i^l x^l)_j = (x^l)_j \quad \forall j \in I_i^l$$

Dann wird die Grobgitterkorrektur wie folgt realisiert

- Jeder Prozessor i speichert alle Indizes $\tilde{I}_i^l$
- Berechne $d_i^L = U_i^L(b - Ax^k)$ in jedem Prozessor; dies geht lokal ohne Kommunikation, da $I_i^L \subseteq \hat{I}_i^L$

- Somit erfüllen die d_i^L die Bedingung (8.23)

$$d^l = b - Ax^k = \sum_{i=1}^{p} (\tilde{R}_i^L)^T d_i^L$$

- Dann lässt sich d^0 (die rechte Seite des Grobgitterproblems) berechnen als

$$d^0 = R^0 R^1 \dots R^{L-1} d^L = \sum_{i=1}^{p} (\tilde{R}_i^0)^T R_i^0 R_i^1 \dots R_i^{L-1} d_i^L$$

- Um $\tilde{R}_i^0 d^0$ im Prozessor i zu kennen, ist eine Kommunikation über alle gemeinsamen Knoten nötig:

$$(d^0)_j = \sum_{i \in \{k | j \in \tilde{I}_k^0\}} (R_i^0 R_i^1 \dots R_i^{L-1} d_i^L)_j$$

- Löse Grobgitterproblem; die Prolongation der Korrektur geschieht nach (8.26) lokal

Mehrgitter mit minimaler Überlappung

Wir betrachten die Datenaufteilung von Abbildung 8.4. Wir setzen

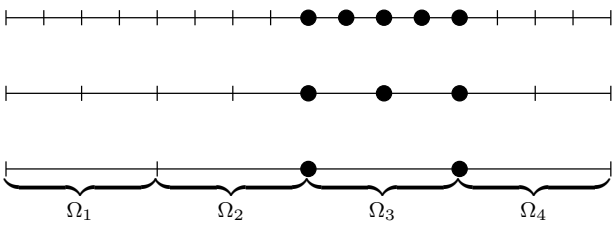


Abbildung 8.4: Datenaufteilung für Mehrgitter (1D)

$$\tilde{I}_i^l = \left\{ j \in I^l \;\middle|\; (x_j, y_j) \in \hat{\Omega}_i \right\}$$

Diese Datenverteilung erfüllt (ii) von oben.

Prozessor i berechnet und speichert die Korrektur für alle Indizes $\tilde{I}_i^l$. Die lokalen Anteile am Gleichungssystem im Prozessor i sind:

$$(A_i^l)_{\alpha,\beta} = \int_{\Omega_i} (K \nabla \varphi_\alpha^l) \cdot \nabla \varphi_\beta^l \, dx \tag{8.27a}$$

$$(b_i^l)_\alpha = \int_{\Omega_i} f \varphi_\alpha^l \, dx. \tag{8.27b}$$

Somit gilt ($A^l x^l = b^l$ ist das lineare Gleichungssystem im sequentiellen Fall):

$$b^l = \sum_{i=1}^{p} (\tilde{R}_i^l)^T b_i^l \tag{8.28a}$$

$$A^l = \sum_{i=1}^{p} (\tilde{R}_i^l)^T A_i^l \tilde{R}_i^l. \tag{8.28b}$$

Damit gilt für den Defekt (einer beliebigen Stufe):

$$\begin{aligned} d^l &= b^l - A^l x^{l,k} = \sum_{i=1}^{p} (\tilde{R}_i^l)^T b_i^l - \sum_{i=1}^{p} (\tilde{R}_i^l)^T A_i^l \underbrace{\tilde{R}_i^l x^{l,k}}_{x_i^l} \\ &= \sum_{i=1}^{p} (\tilde{R}_i^l)^T \underbrace{[b_i^l - A_i^l x_i^{l,k}]}_{d_i^{l,k}} \quad \text{jeder muss } x^l \text{ in } \tilde{I}_i^l \text{ kennen} \\ &= \sum_{i=1}^{p} (\tilde{R}_i^l)^T d_i^{l,k} \end{aligned}$$

also:

- $d_i^{l,k}$ lassen sich lokal ohne Kommunikation berechnen
- $\{d_i^{l,k}\}$ bilden eine additive Zerlegung.

Mit der *Richardson-Iteration* als Glätter erhalten wir folgende parallele Version von Algorithmus 8.9:

Algorithmus 8.12 (paralleles, multiplikatives MG)

- Datenverteilung wie oben definiert.
- Jeder Vektor mit Subskript i und Superskript l ist in $\mathbb{R}^{\tilde{I}_i^l}$.

pmmgc $(x_i^k \ in \mathbb{R}^{\tilde{I}_i^l})$
{
  $d_i^L = b_i^l - A_i^L x_i^k$;
  for $(l = L;\ l > 0;\ l = l - 1)$
  {
    $v_i^l = 0$;
    for $(m = 0;\ m < \nu_1;\ m = m + 1)$
    {
      $y_i^l = \omega d_i^l$; // Richardson!
      $y_i^l = \tilde{R}_i^l \sum_{j=1}^{p} (\tilde{R}_j^l)^T y_j^l$; // Kommunikation

```
        v_i^l = v_i^l + y_i^l;
        d_i^l = d_i^l − A_i^l y_i^l;
      }
      d_i^{l−1} = R_i^{l−1} d_i^l;                    // lokal
    }
    v_i^0 = R̃_i^0 (A^0)^{−1} d^0;                     // erfordert Kommunikation
    for (l = 1; l ≤ L; l = l + 1)
    {
      y_i^l = (R_i^{l−1})^T v_i^{l−1};                // lokal
      v_i^l = v_i^l + y_i^l;
      d_i^l = d_i^l − A_i^l y_i^l;                    // Defekt ist additiv!
      for (m = 0; m < ν_2; m = m + 1)
      {
        y_i^l = ω d_i^l;
        y_i^l = R̃_i^l Σ_{j=1}^p (R̃_j^l)^T y_j^l;
        v_i^l = v_i^l + y_i^l;
        d_i^l = d_i^l − A_i^l y_i^l;
      }
    }
    return x_i^{k+1} = x_i^k + v_i^L;
}
```

- Braucht eine Kommunikation pro Glättungsschritt und
- Kommunikation zur Lösung des Stufe-0-Problems (wie Schwarz-Verfahren).

Das additive MG-Verfahren (Algorithmus 8.8) lässt sich analog parallelisieren. Dort kann die Kommunikation aller Glättungsschritte der Stufen $1, \ldots, L$ in *einem* (entsprechend größeren) Kommunikationsschritt abgewickelt werden.

Eingitter-Schwarz als „Glätter“ im MG

Hier zeigen wir, wie man überlappende Schwarz-Verfahren als Glätter im MG-Verfahren einsetzen kann. Hier wird man die Teilgebietsprobleme sicher *nicht* exakt lösen wollen.

Wir beschränken uns auf die additive Variante, d.h. additiver überlappender Schwarz als Glätter im additiven Mehrgitterverfahren.

Wir arbeiten mit der selben Datenaufteilung wie im überlappenden Schwarz-Verfahren, nur dass nun alle Stufen $0, \ldots L$ zum Einsatz kommen und auf jeder Stufe eine Zerlegung in überlappende Teilgebiete benötigt wird. Hier die Konstruktion:

(a). Ω Gebiet

(b). $\bar{\Omega} = \bigcup_{i=1}^p \bar{\Omega}_i, \quad \Omega_i \cap \Omega_j = \emptyset$ für $i \neq j$: nicht überlappende Zerlegung

(c). $\hat{\Omega}_i$: Ω_i vergrößert um $m_i \cdot h_i$ mit h_i der Gitterweite auf Stufe i

(d). Wir verlangen $m_i \leq 2m_{i-1}$; damit folgt

$$\Omega_i \subseteq \Omega_{i-1}$$

(e). $\hat{I}_i^l = \left\{ j \in I^l \;\middle|\; (x_j, y_j) \in \hat{\Omega}_i \right\}$
$\tilde{I}_i^l = \left\{ j \in I^l \;\middle|\; (x_j, y_j) \in \overline{\hat{\Omega}_i} \right\}$
Bedingung (d) sichert Bedingung (ii) an die $\tilde{I}_i^l$.
$I^l = \bigcup_{i=1}^p I_i^l, \quad I_i^l \cap I_j^l = \emptyset$ für $i \neq j$ sei eine eindeutige Zerlegung von I^l.

Algorithmus 8.13 (Additiver Schwarz + additives MG) Prozessor i besitzt

$$(A_i^l)_{\alpha,\beta} = a(\varphi_\alpha^l, \varphi_\beta^l) \quad \alpha, \beta \in \hat{I}_i^l$$

und jeder Vektor mit Subskript i und Superskript l ist in $\mathbb{R}^{\tilde{I}_i^l}$.

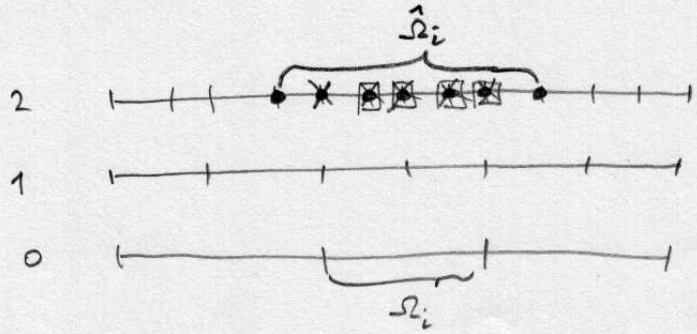
samgc $(x_i^k \in \mathbb{R}^{I^L})$
{
$(d^L)_j = \begin{cases} (b_i^l - A_i^l x_i^k)_j & j \in I_i^l \\ 0 & \text{sonst} \end{cases}$; // additive Zerlegung des Defektes
for $(l = L-1;\ l \geq 0;\ l = l-1)$
$d_i^l = R_i^l d_i^{l+1}$; // lokale Restriktion
$\forall l \in \{0, \ldots, L\}$
$(d_i^l)_j = \sum_{m \in \{n | j \in \tilde{I}_n^l\}} (d_m^l)_j \quad j \in \hat{I}_i^l$;
$v_i^0 = \tilde{R}_i^0 (A^0)^{-1} d^0$;
$\forall l \in \{1, \ldots, L\}$ // Schwarz nur inexakt lösen!
$v_i^l = \begin{cases} (A_i^l)^{-1} d_i^l & \text{auf } \hat{I}_i^l \\ 0 \text{ sonst} \end{cases}$;
$\forall l \in \{1, \ldots, L\}$
$(v_i^l)_j = \sum_{m \in \{n | j \in \tilde{I}_n^l\}} (v_m^l)_j \quad j \in \tilde{I}_i^l$;
for $(l = 1;\ l \leq L;\ l = l+1)$
$v_i^l = (R_i^{l-1})^T v_i^{l-1}$; // lokale Prolongation
return $x_i^{k+1} = x_i^k + v_i^L$;
}

Bemerkung 8.14

- Die Kommunikation aller Stufen kann in einer Message zusammengefasst werden.
- Im multiplikativen Mehrgitter schalte nach dem Schwarz-Glätter eine Restriktion auf die Knoten I_i^l ein.

Parallelisierung

Zweigitter überlappender Schwarz



• $\tilde{\mathcal{I}}_i^L$ × $\hat{\mathcal{I}}_i^L$ ○ $\bar{\mathcal{I}}_i^L$

$\mathcal{I}^H = \mathcal{I}^0, \dots, \mathcal{I}^L = \mathcal{I}^h$ Gitterhierarchie

Indexmengen auf Stufe L:

$$\tilde{\mathcal{I}}_i^L = \{ k \mid x_k \in \tilde{\Omega}_i \}$$

$$\hat{\mathcal{I}}_i^L = \{ k \mid x_k \in \hat{\Omega}_i \}$$

$$\mathcal{I}^L = \bigcup_{i=1}^{P} \tilde{\mathcal{I}}_i^L = \bigcup_{i=1}^{P} \hat{\mathcal{I}}_i^L$$ überlappende Zerlegung

$$\mathcal{I}^L = \bigcup_{i=1}^{P} \bar{\mathcal{I}}_i^L$$ sei eine nicht überlappende Zerlegung sodass

$$\bar{\mathcal{I}}_i^L \subseteq \hat{\mathcal{I}}_i^L \subseteq \tilde{\mathcal{I}}_i^L$$

Indexmengen auf Stufe $l < L$

$$k \in \tilde{\mathcal{I}}_i^l \iff \exists j \in \tilde{\mathcal{I}}^{l+1} : \theta_{k,j}^l \neq 0 \quad (\text{dort neu } \tilde{\mathcal{I}}_i^l)$$

parallele Implementierung:

- x_i^L (Koeffizienten zu u_h) kennt Prozessor i auf $\tilde{\mathcal{I}}^L$
- A^L : Prozessor i kennt Zeilen zu $\hat{\mathcal{I}}_i^L$
- b^L : " i " Einträge zu $\hat{\mathcal{I}}_i^L$
- $d_i^L = b_i^L - A_i^L x_i^L$ kann Prozessor i auf $\hat{\mathcal{I}}_i^L$ berechnen
  $\Rightarrow$ Löse subdomain Problem, Kommunikation für v_i^L auf $\tilde{\mathcal{I}}_i^L$
- $d_i^L = \underset{\text{zu } I_i^L}{\bar{R}_i} \underset{\text{zu } \hat{\mathcal{I}}_i^L}{(\hat{R}_i^L)^T} d_i^L \rightsquigarrow d_i = \sum (\hat{R}^L)^T d_i^L$ (additive Darstellung)

  $\Rightarrow d_i^0 = R^0 \dots R^{L-1} d_i^L$ mit $d^0 = \sum_{i=1}^{P} \tilde{R}_0^T d_i^0$
- Alle an alle um $A^0 y^0 = d^0$ zu lösen – umgekehrter Weg für v.

BPX

h-robustness

L	h	H=1/2 #it (10^{-8} red)	H=1/16 (10^{-3} red) #it (10^{-8})	H=1/16 (10^{-6} red) #it (10^{-8})
1	4^{-1}	6		
2	8^{-1}	12		
3	16^{-1}	15		
4	32^{-1}	16	17	17
5	64^{-1}	17	21	21
6	128^{-1}	18	22	22
7	256^{-1}	18	24	24
8	512^{-1}	18	24	24
9	1024^{-1}	18	25	25

strong scaling $h = 1/2048$

P	1	2	4	8	16	32	64	128	256
Time (25 its)	27.34	14.82	7.64	3.94	1.99	0.98	0.5	0.46	0.22
Speedup	–	1,8	3,6	6,9	13,7	27,9	54,7	59,4	124,3

weak scaling 1 Mio DOF / procesn

P	1	4	16	64	256
iT	25	25	25	25	25
time	5.83	5.84 / 7.6 (4 nodes / 1 node)	5.95 / 8 (16 nod / 4 nodes)	8.33	9.14

Konvergenzbeweis für Multilevel Diagonal Scaling

Es viele verschiedene Varianten!
Wir folgen hier Smith, Bjørstad, Gropp.

Funktionenräume.

- geschachtelte Gitterhierarchie
- $\Phi^\ell = \{ \phi_i^\ell \mid i \in \mathcal{I}^\ell \}$ Lagrangebasis für P_1 oder Q_1 auf Stufe ℓ
- $V_i^\ell = \operatorname{span}\{\phi_i^\ell\}$ $\ell = 1, \dots, L$, $i \in \mathcal{I}^\ell$
- $V^\ell = \operatorname{span} \Phi^\ell = \operatorname{span}\{ \phi_i^\ell \mid i \in \mathcal{I}^\ell \}$ FE-Raum auf Stufe $\ell = 0, \dots, L$

a-Projektion.

$\mathcal{P}_\ell : H^1(\Omega) \to V^\ell$

$$a(\mathcal{P}_\ell u, v) = a(u, v) \quad \forall v \in V^\ell.$$

Praktisch bedeutet dies das Lösen eines Systems auf Stufe ℓ:

$u^\ell = \mathcal{P}_\ell u$: $a(u^\ell, v) = a(u, v) \ \forall v \in V^\ell$.
↳ Vorsicht! das ist nicht dasselbe $\mathcal{P}_\ell$ von unten!

Splitting.

Gegeben ein $u \in V^h = V^L$. Wir zerlegen u zunächst in die einzelnen Stufen rekursiv:

$$u^0 = \mathcal{P}_0 u$$
$$u^\ell = \underbrace{(\mathcal{P}_\ell - \mathcal{P}_{\ell-1})}_{\tilde{\mathcal{P}}_\ell \text{ neue Projektion}} u \quad \text{für } \ell > 0$$

Wir zeigen nun einige Eigenschaften dieser Projektionen

(i)
$$\sum_{\ell=0}^{L} u^\ell = \mathcal{P}_0 u + (\mathcal{P}_1 u - \mathcal{P}_0 u) + (\mathcal{P}_2 u - \mathcal{P}_1 u) + \dots + (\mathcal{P}_L u - \mathcal{P}_{L-1} u)$$
$$= \mathcal{P}_L u = u \qquad \text{(Teleskopsumme)}$$

(ii) $$a(u^\ell, u^0) = \begin{cases} 0 & \ell > 0 \\ a(u, u^0) & \ell = 0 \end{cases}$$

$\ell > 0$: $a(u^\ell, u^0) = a(P_\ell u - P_{\ell-1} u, P_0 u)$

$$= a(P_\ell u, P_0 u) - a(P_{\ell-1} u, P_0 u)$$

($P_0 u \in V^0 \subset V^\ell$, $P_0 u \in V^0 \subset V^{\ell-1}$)

Def. von $P_\ell, P_{\ell-1}$ ↘

$$= a(u, P_0 u) - a(u, P_0 u) = 0$$

$\ell = 0$: $a(u^0, u^0) = a(u, u^0)$ da $u^0 \in V^0$ nach Def. von P_0.

(iii) $$a(u^\ell, u^k) = \begin{cases} 0 & \ell > k > 0 \\ a(u, u^\ell) & \ell = k > 0 \end{cases}$$

$$a(u^\ell, u^k) = a(P_\ell u - P_{\ell-1} u, P_k u - P_{k-1} u)$$

($\ell \geq k > 0$)

$$= a(P_\ell u, P_k u) - \underbrace{a(P_\ell u, P_{k-1} u)}_{a(u, P_{k-1} u)} - a(P_{\ell-1} u, P_k u) + \underbrace{a(P_{\ell-1} u, P_{k-1} u)}_{= a(u, P_{k-1} u)}$$

da $P_{k-1} u \in V^{k-1} \subseteq V^\ell$; da $P_{k-1} u \in V^{k-1} \subseteq V^{\ell-1}$

$$= a(P_\ell u, P_k u) - a(P_{\ell-1} u, P_k u)$$

$$= \begin{cases} a(u, \underset{\in V^\ell}{P_k u}) - a(u, \underset{\in V^{\ell-1}}{P_k u}) = 0 & \text{falls } \ell > k \\ \quad \text{da } k < \ell \Leftrightarrow k \leq \ell - 1 \\ a(P_\ell u, u) - a(P_{\ell-1} u, u) = a(u^\ell, u) & \ell = k \end{cases}$$

Wo braucht man das?

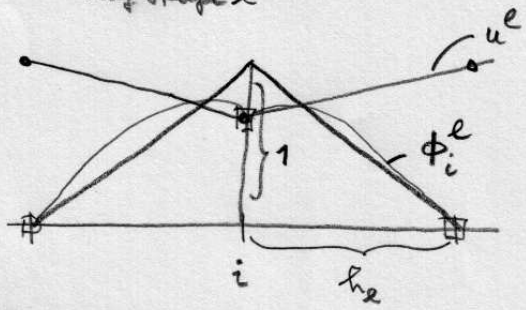
Weiter in der Zerlegung.

Die levelweisen Funktionen werden nun weiter zerlegt in die Teilräume $V_i^\ell \subset V^\ell$ für $\ell > 0$

$$V_i^\ell \ni u_i^\ell = \underbrace{u^\ell(x_i)}_{\text{Auswertung am Knoten } i}\, \phi_i^\ell = I^{h_\ell}\left(\phi_i^\ell u^\ell\right)$$

I^{h_ℓ}: Knoten-interpolation auf Stufe ℓ

$\{\phi_i^\ell\}$ agieren quasi als Partition der Eins!



Es gilt
$$\left(\phi_i^\ell u^\ell\right)(x_j^\ell) = \begin{cases} u^\ell(x_i)\cdot 1 & j = i \\ u^\ell(x_j)\cdot 0 = 0 & j \neq i \end{cases}$$

Damit ist die Zerlegung komplett und es gilt

$$u^0 + \sum_{\ell=1}^{J} \sum_{i \in \mathcal{I}^\ell} u_i^\ell = u^0 + \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}^\ell} I^{h_\ell}\left(\phi_i^\ell u^\ell\right) = u^0 + \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}^\ell} u^\ell(x_i)\,\phi_i^\ell$$

$$= \sum_{\ell=0}^{L} u^\ell = u.$$

05.06.09
4

Lemma 1 (Stabilität der Zerlegung)

Das Problem $a(u,v) = \ell(v)$ sei H^2-regulär. Dann gilt für obige Zerlegung einer Funktion $u \in V^h = V^L$

$$a(u^0,u^0) + \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}^\ell} a(u_i^\ell, u_i^\ell) \leq C\, a(u,u).$$

mit C unabhängig von h.

Beweis:

sei $\ell > 0$.

(i) $$\sum_{i \in \mathcal{I}^\ell} a(u_i^\ell, u_i^\ell) \leq \sum_{i \in \mathcal{I}^\ell} C\, |u_i^\ell|^2_{H^1(\Omega)}$$

(Bach. Koeffiziate)

$$= \sum_{i \in \mathcal{I}^\ell} C\, |u_i^\ell|^2_{H^1(\Omega_i^\ell)} \qquad \text{mit } \Omega_i^\ell := \operatorname{supp} \phi_i^\ell$$

$$= C \sum_{i \in \mathcal{I}^\ell} |I^{h_\ell}(\phi_i^\ell u^\ell)|^2_{H^1(\Omega_i^\ell)}$$

(Lemma 7.7, $\phi_i^\ell u^\ell$ stückw. quadrat.)

$$\leq C \sum_{i \in \mathcal{I}^\ell} |\phi_i^\ell u^\ell|^2_{H^1(\Omega_i^\ell)}$$

(wie im Beweis von Schwarz; Produktregel, Triangle, Young)

$$= C \sum_{i \in \mathcal{I}^\ell} \int_{\Omega_i^\ell} \|\nabla(\phi_i^\ell u^\ell)\|_2^2 \, dx$$

$$\leq C \sum_{i \in \mathcal{I}^\ell} \int_{\Omega_i^\ell} \underbrace{\|\phi_i^\ell \nabla u^\ell\|_2^2}_{\phi_i^\ell \leq 1} + \underbrace{\|u^\ell \nabla \phi_i^\ell\|_2^2}_{|u^\ell|^2 \|\nabla \phi_i^\ell\|_2^2 \leq \left(\frac{1}{h_\ell}\right)^2} dx$$

(quasi-uniformity (alle h's gleich))

$$\leq C \sum_{i \in \mathcal{I}^\ell} \left\{ |u^\ell|^2_{H^1(\Omega_i^\ell)} + \left(\frac{1}{h_\ell}\right)^2 \|u^\ell\|^2_{L^2(\Omega_i^\ell)} \right\}$$

$$\leq C \left\{ |u^\ell|^2_{H^1(\Omega)} + \left(\frac{1}{h_\ell}\right)^2 \|u^\ell\|^2_{L^2(\Omega)} \right\}$$

(ii) Das kontinuierliche Problem $u \in H_0^1(\Omega)$: $a(u,v) = \ell(v)$ $\forall v \in H_0^1(\Omega)$ sei H^2-regulär. Mit Hilfe des Aubin-Nitsche-Lemmas (z.B. Braess, Lemma 7.6) gilt für die Lösung u_h in $V^h \subset H_0^1(\Omega)$

$$\| u - u_h \|_{L^2(\Omega)} \leq C h \| u - u_h \|_{H^1(\Omega)} .$$

- auf $H_0^1(\Omega)$ sind $\|.\|_{H^1(\Omega)}$ und $|u|_{H^1(\Omega)}$ äquivalent
- Anwendg auf die Projektionen P_ℓ: „beliebiges" $W \in V^h \subset H_0^1(\Omega)$ ist gegeben

$$a(W,v) = a(W,v) \quad \forall v \in V^{\ell} \text{“ hat Lösung } w$$

$$a(w^\ell, v) = \underbrace{a(W,v)}_{\text{„}\ell(v) := a(W,v)\text{“ für ein festes, gegebenes } W.} \quad \forall v \in V^\ell \text{ mit } w^\ell = P_\ell w$$

somit gilt

$$\| W - P_\ell W \|_{L^2(\Omega)} \leq C h_\ell |W|_{H^1(\Omega)}$$

(iii)

$$a(P_{\ell-1}(P_\ell u), v) = a(P_\ell u, v) \quad \forall v \in V^{\ell-1}$$

Projektion von $P_\ell u$

$$= a(u, v) \; \forall v \in V^{\ell-1}; \text{ aus Def von } P_\ell \text{ und } V^{\ell-1} \subset V^\ell$$

$$= a(P_{\ell-1} u, v) \; \forall v \in V^{\ell-1};$$

$$\Leftrightarrow P_{\ell-1} P_\ell u = P_{\ell-1} u$$

und damit für $\ell \geq 1$: Def. von u^ℓ

$$(I - P_{\ell-1}) u^\ell = (I - P_{\ell-1})(P_\ell - P_{\ell-1}) u$$

$$= (P_\ell - P_{\ell-1} - P_{\ell-1} P_\ell + P_{\ell-1}^2) u \qquad P_{\ell-1}^2 = P_{\ell-1}, \; P_{\ell-1} P_\ell = P_{\ell-1}$$

$$= (P_\ell - P_{\ell-1}) u = u^\ell \qquad \text{liegt a-orthogonal.}$$

d.h. $u^\ell = u^\ell - P_{\ell-1} u^\ell$

mit (ii) $\| u^\ell \|_{L^2(\Omega)} = \| u^\ell - P_{\ell-1} u^\ell \|_{L^2(\Omega)} \leq C h_{\ell-1} |u^\ell|_{H^1(\Omega)}$

(iv) damit weiter in (i):

$$\sum_{i \in \mathcal{I}^\ell} a(u_i^\ell, u_i^\ell) \leq C \left\{ |u^\ell|^2_{H^1(\Omega)} + \left(\frac{1}{h_\ell}\right)^2 \|u^\ell\|^2_{L^2(\Omega)} \right\}$$

$$\overset{\text{(iii)}}{=} \|u^\ell - \mathcal{P}_{\ell-1} u^\ell\|^2_{L^2(\Omega)}$$

$$\overset{\text{(ii)}}{\leq} C h^2_{\ell-1} |u^\ell|^2_{H^1(\Omega)}$$

$$\leq C \left\{ |u^\ell|^2_{H^1(\Omega)} + \underbrace{\left(\frac{h_{\ell-1}}{h_\ell}\right)^2}_{\substack{=2 \text{ bei} \\ \text{uniformer} \\ \text{Verfeinerung}}} |u^\ell|^2_{H^1(\Omega)} \right\}$$

$$\leq C \, |u^\ell|^2_{H^1(\Omega)}.$$

(v) Summe über alle Level

$$a(u^0, u^0) + \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}^\ell} a(u_i^\ell, u_i^\ell)$$

$$\leq a(u^0, u^0) + \sum_{\ell=1}^{L} C \, |u^\ell|^2_{H^1(\Omega)}$$

Äquivalenz

$$\leq C \sum_{\ell=0}^{L} a(u^\ell, u^\ell) = C \sum_{\ell=0}^{L} \left(a(u^\ell, u^\ell) + \sum_{\ell' \neq \ell} \underbrace{a(u^{\ell'}, u)}_{=0} \right)$$

wg. (i)', (ii) bei Eigenschaften

$$= C \sum_{\ell=0}^{L} a(u, u^\ell)$$

$$= C \, a\Big(u, \underbrace{\sum_{\ell=0}^{L} u^\ell}_{=u}\Big)$$

$$= C \, a(u, u)$$

Für die Abschätzung nach oben benützt man wieder ein Färbungsargument

Lemma 2

Es sei R^0 die Restriktion von $R^{\mathcal{I}^L} \to R^{\mathcal{I}^0}$ wie im Zweigitter Schwarz-Verfahren, also $(R^0)^T$ die Darstellung von Grobgitterfunktionen als Feingitterfunktionen.
$(R_i^\ell)^T$ sei analog die Prolongation von V_i^ℓ nach V^ℓ in Koeffizienten.
Entsprechend ist wie immer

$$A^0 = R^0 A (R^0)^T, \qquad P^0 = (R^0)^T (A^0)^{-1} R^0 A,$$
$$A_i^\ell = R_i^\ell A (R_i^\ell)^T, \qquad P_i^\ell = (R_i^\ell)^T (A_i^\ell)^{-1} R_i^\ell A;$$

Dann gilt

$$\left\langle \left(P^0 + \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}^\ell} P_i^\ell \right) x, x \right\rangle_A \leq C(\#L) \langle x, x \rangle_A .$$

Beweis. Auf jeder Stufe gibt es eine Färbung der Knoten in N^C Gruppen

$$\mathcal{I}^\ell = \bigcup_{c=1}^{N^C} \mathcal{I}^{\ell,c}$$

so dass $\langle P_i^\ell x, P_j^\ell x \rangle_A = 0$ für $c(i) \neq c(j)$

mit N^C unabhängig von h. ($c=9$ für Q_1 auf strukt. Gitter)

Mit dem Lemma 7.10 gilt dann für $\ell \geq 1$:

$$\left\langle \sum_{i \in \mathcal{I}^\ell} P_i^\ell x, x \right\rangle_A \leq N^C \langle x, x \rangle_A$$

und damit

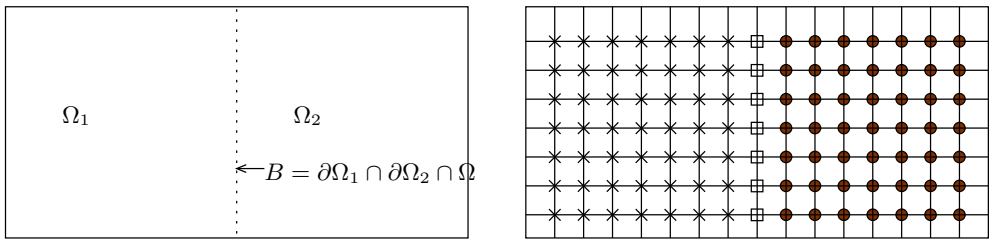
$$\left\langle \left(P^0 + \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}^\ell} P_i^\ell \right) x, x \right\rangle_A \leq \underbrace{\langle x, x \rangle_A}_{\ell=0} + \sum_{\ell=1}^{L} N^C \langle x, x \rangle_A$$

$$\leq N^C (1+L) \langle x, x \rangle_A \qquad \square$$

9 Nichtüberlappende Gebietszerlegungsverfahren

9.1 Einführung und Motivation

Betrachte die Zerlegung von $\Omega = (0,2) \times (0,1)$ in zwei nichtüberlappende Teilgebiete Ω_1 und Ω_2, wie sie in Abbildung 9.1a) dargestellt ist.



(a) (b)

Abbildung 9.1: Zerlegung (a) und Diskretisierung (b) bei zwei Teilgebieten

Nach einer passenden Diskretisierung auf den Teilgebieten (Substrukturen) wie in Abbildung 9.1b) nummerieren wir zuerst die Unbekannten in Teilgebiet 1 dann die in Teilgebiet 2 und zuletzt die auf dem inneren Rand $B = \partial\Omega_1 \cap \partial\Omega_2 \cap \Omega$. Das entstehende Gleichungssystem $Ax = b$ hat eine 3×3 Blockgestalt entsprechend dieser Zerlegung:

$$\begin{pmatrix} A_{II}^{(1)} & 0 & A_{IB}^{(1)} \\ 0 & A_{II}^{(2)} & A_{IB}^{(2)} \\ A_{BI}^{(1)} & A_{BI}^{(2)} & A_{BB} \end{pmatrix} \begin{pmatrix} x_I^{(1)} \\ x_I^{(2)} \\ x_B \end{pmatrix} = \begin{pmatrix} b_I^{(1)} \\ b_I^{(2)} \\ b_B \end{pmatrix} \tag{9.1}$$

Dabei entspricht $x_I^{(1)}$ den Unbekannten in Ω_1, $x_I^{(2)}$ denen in Ω_2 und x_B denen auf B.

Block-Gauß-Elimination der Blöcke $A_{BI}^{(1)}$ und $A_{BI}^{(2)}$ ergibt

$$\begin{pmatrix} A_{II}^{(1)} & 0 & A_{IB}^{(1)} \\ 0 & A_{II}^{(2)} & A_{IB}^{(2)} \\ 0 & 0 & S \end{pmatrix} \begin{pmatrix} x_I^{(1)} \\ x_I^{(2)} \\ x_B \end{pmatrix} = \begin{pmatrix} b_I^{(1)} \\ b_I^{(2)} \\ g \end{pmatrix}$$

mit

$$S = A_{BB} - A_{BI}^{(1)} {A_{II}^{(1)}}^{-1} A_{IB}^{(1)} - A_{BI}^{(2)} {A_{II}^{(2)}}^{-1} A_{IB}^{(2)}$$
$$g = b_B - A_{BI}^{(1)} {A_{II}^{(1)}}^{-1} b_I^{(1)} - A_{BI}^{(2)} {A_{II}^{(2)}}^{-1} b_I^{(2)}.$$

Der generelle Algorithmus geht wie folgt vor:

(i). berechne rechte Seite g (paralleles Lösen $A_{II}^{(i)^{-1}}$)

(ii). löse $Sx_B = g$ – Ziel: iterativ, ohne Aufstellen von S!

(iii). löse $A_{II}^{(i)} x_I^{(i)} = b_I^{(i)} - A_{IB}^{(i)} x_B$ (parallel)

Alternativ kann man folgende LDU-Zerlegung von A nutzen

$$A = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ A_{BI}^{(1)} A_{II}^{(1)^{-1}} & A_{BI}^{(2)} A_{II}^{(2)^{-1}} & I \end{pmatrix} \begin{pmatrix} A_{II}^{(1)} & 0 & 0 \\ 0 & A_{II}^{(2)} & 0 \\ 0 & 0 & S \end{pmatrix} \underbrace{\begin{pmatrix} I & 0 & A_{II}^{(1)^{-1}} A_{IB}^{(1)} \\ 0 & I & A_{II}^{(2)^{-1}} A_{IB}^{(2)} \\ 0 & 0 & I \end{pmatrix}}_{\substack{\text{Trafo in „diskret} \\ \text{harmonische Basis“}}},$$

S sowie $A_{II}^{(i)^{-1}}$ durch Näherungen ersetzen und daraus ein Iterationsverfahren für das Gesamtsystem gewinnen.

Wir werden im Folgenden den ersten Ansatz weiter verfolgen. Der alternative Ansatz hat Vorteile bei der Anwendung inexakter Teilgebietslöser, dazu aber später.

Das *„Schurkomplement“* S ist positiv definit und voll besetzt (bei $p > 2$ Teilgebieten ist es blockweise voll besetzt), ist jedoch besser konditioniert als A selbst: $\kappa(S) = O(h^{-1})$ (A ist Diskretisierung eines elliptischen Operators zweiter Ordnung).

Die *iterative* Lösung des Schurkomplementsystems $Sx_B = g$ mit dem CG-Verfahren erfordert nur Matrix-Vektor Multiplikationen der Form $Sy = (A_{BB} - \sum_{i=1}^{2} A_{BI}^{(i)} A_{II}^{(i)^{-1}} A_{IB}^{(i)}) y$, die sich durch *paralleles* Lösen in den Teilgebieten realisieren lassen.

Ziel aller im Folgenden beschriebenen Verfahren ist die Konstruktion geeigneter Vorkonditionierer (approximativer Inversen) für das Schurkomplementsystem. Hierbei betrachten wir zunächst ausführlich den Fall von zwei Teilgebieten (2D) und dann den Fall vieler Teilgebiete (2D und 3D).

9.2 Vorkonditionierer bei zwei Teilgebieten

9.2.1 J-Operator

In speziellen Fällen kennt man eine Transformation von S auf Diagonalgestalt. Es sei Gebiet $\Omega = \Omega_1 \cup \Omega_2$ und Gitter wie in Abbildung 9.2.

Ist A eine Diskretisierung von $-\Delta$ mit dem 5-Punkte-Stern, so lässt sich S schreiben als

$$S = F \Lambda F$$

mit der symmetrischen und unitären Matrix F –

$$(F)_{ij} = \sqrt{\tfrac{2}{n+1}} \sin\left(\tfrac{ij\pi}{n+1}\right) \quad i, j \in \{1, \ldots, n\}$$

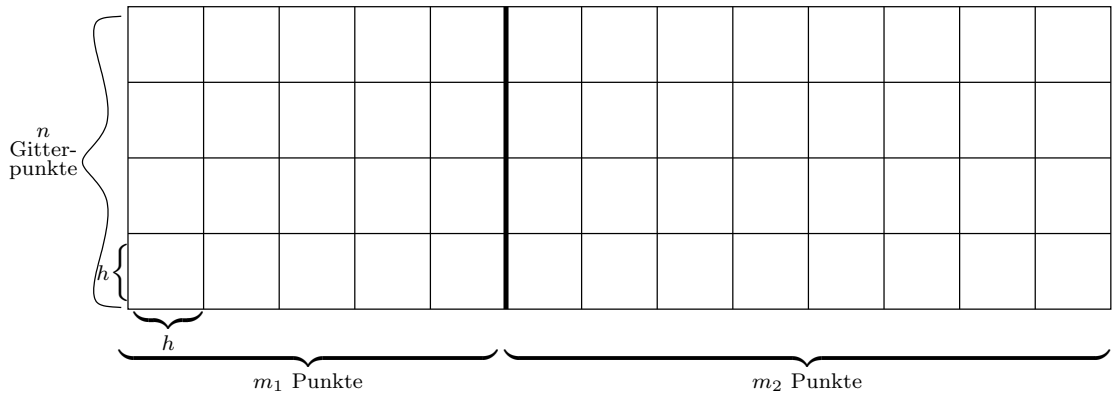


Abbildung 9.2: Gitter

– und der Diagonalmatrix

$$(\Lambda)_{i,i} - \left(\frac{1+\gamma_i^{m_1+1}}{1-\gamma_i^{m_1+1}} + \frac{1+\gamma_i^{m_2+1}}{1-\gamma_i^{m_2+1}}\right)\sqrt{\sigma_i + \sigma_i^2/4}$$

wobei

$$\sigma_i = 4\sin^2\left(\tfrac{i\pi}{2(n+1)}\right), \quad \gamma_i = 1 + \sigma_i/2 - \sqrt{\sigma_i + \sigma_i^2/4} \in (\tfrac{1}{2}, 1)$$

siehe (Chan 1994) und Referenzen dort. Für m_1, m_2 genügend groß kann man z.B.

$$S \approx J = F\Sigma^{1/2}F, \quad \Sigma = \mathrm{diag}(\sigma_i)$$

als spektraläquivalente Näherung verwenden. Die Lösung des Systems

$$Jv = r$$

kann in $O(n \log n)$ Schritten mittels schneller Fourier-Sinustransformation durchgeführt werden. Der Vorkonditionierer lautet

$$\boxed{B_J = F\Sigma^{1/2}F.}$$

Ist das Gitter nicht mehr regelmäßig oder die Koeffizienten des elliptischen Operators variabel, so kann J noch als Vorkonditionierer verwendet werden, die Konvergenzrate wird jedoch (viel) schlechter.

9.2.2 Neumann-Dirichlet Vorkonditionierer

Bei einer FE-Disktretisierung zerfällt der A_{BB}-Block in Gleichung (9.1) in natürlicher Weise in zwei Teile $A_{BB} = A_{BB}^{(1)} + A_{BB}^{(2)}$, die durch Integration der Bilinearform in den Teilgebieten Ω_i, $i = 1, 2$ entstehen. Das Schurkomplement S schreibt sich dann als

$$S = \underbrace{A_{BB}^{(1)} - A_{BI}^{(1)}{A_{II}^{(1)}}^{-1}A_{IB}^{(1)}} + \underbrace{A_{BB}^{(2)} - A_{BI}^{(1)}{A_{II}^{(2)}}^{-1}A_{IB}^{(2)}} = S^{(1)} + S^{(2)},$$

wird also aus Beiträgen von jedem Teilgebiet zusammengesetzt. $S^{(i)}$ kann auch als Schurkomplement der Matrix

$$A^{(i)} = \begin{pmatrix} A_{II}^{(i)} & A_{IB}^{(i)} \\ A_{BI}^{(i)} & A_{BB}^{(i)} \end{pmatrix} \tag{9.2}$$

verstanden werden, wobei links bzw. oben die Unbekannten in Ω_i, rechts bzw. unten die auf B stehen.

Der Neumann-Dirichlet Vorkonditionierer basiert auf der Idee, dass

$$S^{(1)} = S^{(2)}, \quad \text{also } S = 2S^{(1)},$$

falls Gebiet und Gitter spiegelsymmetrisch zum internen Rand sind (A ist Diskretisierung von $-\Delta$, sowieso). In diesem Fall ist $S^{(1)}$ ein idealer Vorkonditionierer, im allgemeinen Fall wird gehofft, dass $S^{(1)}$ immer noch akzeptabel ist.

Im Sinne eines Rechtsvorkonditionierers

$$\boxed{B_{ND} = {S^{(1)}}^{-1}}$$

wird die CG-Iteration angewandt auf das transformierte Gleichungssystem

$$S{S^{(1)}}^{-1} w_B = g, \quad x_B = {S^{(1)}}^{-1} w_B,$$
$$S{S^{(1)}}^{-1} = (S^{(1)} + S^{(2)}){S^{(1)}}^{-1} = I + \underbrace{S^{(2)}{S^{(1)}}^{-1}}_{\substack{= I \text{ im} \\ \text{Idealfall}}}.$$

Eine Iteration benötigt somit

(i). eine Multiplikation mi ${S^{(1)}}^{-1}$

(ii). eine Multiplikation mit $S^{(2)}$

Zum Schritt (i) Für die Matrix $A^{(i)}$ aus (9.2) gilt die Block-LDU-Zerlegung

$$A^{(i)} = \begin{pmatrix} I & 0 \\ A_{BI}^{(1)} {A_{II}^{(1)}}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{II}^{(1)} & 0 \\ 0 & S^{(1)} \end{pmatrix} \begin{pmatrix} I & {A_{II}^{(1)}}^{-1} A_{IB}^{(1)} \\ 0 & I \end{pmatrix} =: LDU$$

Für die Inverse ${A^{(i)}}^{-1}$ rechnet man nach:

$$\begin{aligned} {A^{(i)}}^{-1} &= \begin{pmatrix} I & -{A_{II}^{(1)}}^{-1} A_{IB}^{(1)} \\ 0 & I \end{pmatrix} \begin{pmatrix} {A_{II}^{(1)}}^{-1} & 0 \\ 0 & {S^{(1)}}^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{BI}^{(1)} {A_{II}^{(1)}}^{-1} & I \end{pmatrix} \\ &= \begin{pmatrix} {A_{II}^{(1)}}^{-1} + {A_{II}^{(1)}}^{-1} A_{IB}^{(1)} {S^{(1)}}^{-1} A_{BI}^{(1)} {A_{II}^{(1)}}^{-1} & -{A_{II}^{(1)}}^{-1} A_{IB}^{(1)} {S^{(1)}}^{-1} \\ -{S^{(1)}}^{-1} A_{BI}^{(1)} {A_{II}^{(1)}}^{-1} & {S^{(1)}}^{-1} \end{pmatrix} \end{aligned}$$

Somit kann ${S^{(1)}}^{-1}$ geschrieben werden als

$$\underbrace{{S^{(1)}}^{-1} = \begin{pmatrix} 0 & I_B \end{pmatrix} {A^{(1)}}^{-1} = \begin{pmatrix} * & * \\ * & {S^{(1)}}^{-1} \end{pmatrix}}_{\text{}} \begin{pmatrix} 0 \\ I_B \end{pmatrix},$$

und der rechte Vektor hat oben die Unbekannten in Ω_1 und unten die auf B.

Für ${S^{(1)}}^{-1} r$ ist also ein System der Form $A^{(1)} y = \binom{0}{r}$ zu lösen. $A^{(1)}$ ist eine Diskretisierung von $-\Delta$ mit *Neumann-Randbedingungen auf B.* Als Randdaten dient der Vektor r. Die Multiplikation $\begin{pmatrix} 0 & I_B \end{pmatrix} y$ extrahiert dann das Ergebnis auf den Knoten auf B.

Zum Schritt (ii) Die Multiplikation $S^{(2)} y = (A_{BB}^{(2)} - A_{BI}^{(2)} {A_{II}^{(2)}}^{-1} A_{IB}^{(2)}) y$ erfordert eine Lösung von $A_{II}^{(2)} z_I = A_{IB}^{(2)} y$, d.h. *Dirichlet Randdaten auf B.*

Pro Iterationsschritt ist somit ein Problem in Ω_1 mit Neumann-Randdaten auf B und eines in Ω_2 mit Dirichlet Randdaten auf B zu lösen.

Bemerkungen 9.1

1. Bei zwei Teilgebieten ist das Verfahren sequentiell auf Teilgebietsebene.
2. Die Konvergenz ist unabhängig von h, hängt jedoch von den Koeffizienten und der Form der Teilgebiete ab.
3. Linksvorkonditionierung führt zum Dirichlet-Neumann Algorithmus.

9.2.3 Neumann-Neumann Vorkonditionierer

S ist die Summe aus Anteilen beider Teilgebiete

$$S = \sum_{i=1}^{2} S^{(i)}.$$

Idee: Setze

$$S^{-1} = \left(\sum_{i=1}^{2} S^{(i)} \right)^{-1} \approx \frac{1}{4} \sum_{i=1}^{2} {S^{(i)}}^{-1}.$$

Warum ist das sinnvoll?

$$\left(\sum_{i=1}^{2} {S^{(i)}}^{-1} \right) \left(\sum_{i=1}^{2} S^{(i)} \right) = \underbrace{{S^{(1)}}^{-1} S^{(1)}}_{=I} + \underbrace{{S^{(1)}}^{-1} S^{(2)}}_{\substack{I \text{ wg.} \\ S^{(1)} = S^{(2)} \\ \text{im Spiegel-} \\ \text{sym.} \\ \text{Fall}}} + \underbrace{{S^{(2)}}^{-1} S^{(1)}}_{I} + \underbrace{{S^{(2)}}^{-1} S^{(2)}}_{=I}$$

$$\approx 4I$$

Der Neumann-Neumann Vorkonditionierer lautet

$$\boxed{B_{NN} = \frac{1}{4} \left({S^{(1)}}^{-1} + {S^{(2)}}^{-1} \right)}$$

und erfordert pro Iterationsschritt die Lösung

- zweier Neumann-Probleme in Ω_1 und Ω_2,
- sowie zweier Dirichlet-Probleme in Ω_1 und Ω_2.

9.3 Der Fall vieler Teilgebiete

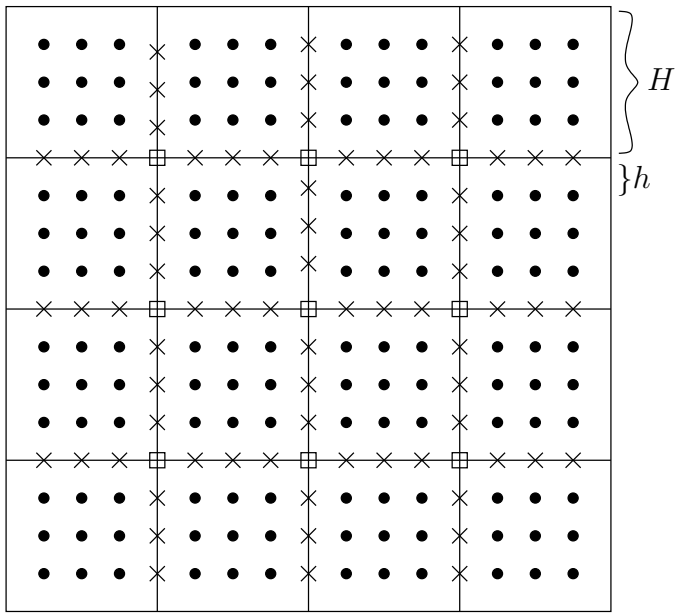


Abbildung 9.3: Konstruktion der nichtüberlappenden Zerlegung bei mehreren Teilgebieten

Das Gebiet Ω sei unterteilt in p nichtüberlappende, polygonale Teilgebiete $\Omega_1, \ldots, \Omega_p$ mit Durchmesser $O(H)$ (siehe Abbildung 9.3). Der Rand B,

$$B = \bigcup_{i=1}^{p} \partial\Omega_i \cap \Omega = \bigcup_{i,j} E_{ij} \cup V,$$

besteht aus Kanten $E_{ij} = \partial\Omega_i \cap \partial\Omega_j \setminus \{\text{Endpunkte}\}$ sowie den Knoten $V = \{V_1, \ldots, V_k\}$, an denen sich mehr als zwei Teilgebiete treffen. $V_i \in V$ heißt Koppelknoten oder cross point. Partitioniert man den Vektor der Unbekannten $x = (x_I, x_B)^T$ wieder entsprechend den inneren bzw. Randknoten, so hat das zu lösende Gleichungssystem die 2×2 Blockgestalt

$$\begin{pmatrix} A_{II} & A_{IB} \\ A_{BI} & A_{BB} \end{pmatrix} \begin{pmatrix} x_I \\ x_B \end{pmatrix} = \begin{pmatrix} b_I \\ b_B \end{pmatrix}$$

wie oben, wobei

$$A_{II} = \begin{pmatrix} A_{II}^{(1)} & & & \\ & A_{II}^{(2)} & & 0 \\ & & \ddots & \\ & 0 & & \ddots \\ & & & & A_{II}^{(p)} \end{pmatrix}$$

nun eine Blockdiagonalmatrix mit p Blöcken ist.

Das Schurkomplementsystem

$$Sx_B = g$$

ergibt sich in analoger Weise mit

$$S = A_{BB} - \sum_{i=1}^{p} A_{BI}^{(i)} {A_{II}^{(i)}}^{-1} A_{IB}^{(i)},$$

$$g = b_B - \sum_{i=1}^{p} A_{BI}^{(i)} {A_{II}^{(i)}}^{-1} b_I^{(i)}$$

(dabei wurde die offensichtliche Blockzerlegung von A_{BI} und A_{IB} verwendet).

Die Matrix S ist nicht voll besetzt. Ein Knoten $v \in B$ ist mit einem anderen Knoten $w \in B$ gekoppelt, falls v *und* w auf dem Rand eines Teilgebietes $\partial\Omega_i$ sind. Dies *verhindert* im allgemeinen das explizite Aufstellen der Matrix S.

Weitere Zerlegungen der Indexmenge aller Knoten sowie zugehörige Blockzerlegungen der Matrizen und Vektoren werden nach Bedarf eingeführt.

9.4 Hierarchische Basis für das Schurkomplementsystem

Dieses Verfahren ist sowohl für zwei als auch für viele Teilgebiete geeignet.

9.4.1 Das Verfahren der hierarchischen Basis

Betrachten wir zunächst das Verfahren der hierarchischen Basis (Yserentant 1986). Dazu benötigt man wieder eine Mehrgitterhierarchie $\mathcal{T}^H = \mathcal{T}^0, \mathcal{T}^1, \mathcal{T}^2, \ldots, \mathcal{T}^L = \mathcal{T}^h$ wie in Kapitel 8. $\Phi^l = \{\varphi_1^l, \ldots, \varphi_{n^l}^l\}$ sei die Standardknotenbasis auf der Stufe l. Die hierarchische Basis Ψ^h für V^h (entsprechend $\mathcal{T}^h$) ist definiert als

$$\Psi^h = \Phi^0 \cup \bigcup_{l=1}^{L} \bigcup_{i \in I^l \setminus I^{l-1}} \{\varphi_i^l\}.$$

Ψ^h und $\Phi^h = \Phi^L$ sind zwei verschiedene Basen des selben Raumes V^h (Finite-Elemente-Funktionen). Es sei

$$\psi_i = \sum_{j=1}^{n^l} \omega_{ij} \varphi_j^L$$

die Darstellung der hierarchischen Basis in der Knotenbasis. Führt man das Verfahren der Finiten Elemente mit der Basis Ψ^h statt Φ^h durch, so erhält man ein Gleichungssystem

$$\hat{A}\hat{x} = \hat{b}$$

gleicher Dimension. $\hat{A}$ ist sehr viel dichter besetzt als die Matrix A zur Basis Φ^h, jedoch hat $\hat{A}$ die wesentlich bessere Kondition ($\kappa(\hat{A}) = O(H^{-2}(1 + \log\left(\frac{H}{h}\right)^2))$; $\log 1 = 0$!). Als approximative Inverse genügt eine einfach gebaute Matrix $\hat{D}^{-1}$ mit

$$(\hat{D})_{ij} = \begin{cases} (\hat{A})_{ij} & i = j \vee (i \in I^0 \wedge j \in I^0) \\ 0 & \text{sonst} \end{cases}.$$

Dies entspricht der im Abschnitt 8.3 gegebenen Definition. Man zeigt dann, dass $\kappa(\hat{D}^{-1}\hat{A}) = O(1 + \log^2(\frac{H}{h})) = O(L^2)$.

Der Zusammenhang zwischen dem Gleichungssystem $\hat{A}\hat{x} = \hat{b}$ in der hierarchischen Basis und dem Gleichungssystem $Ax = b$ in der Knotenbasis ist gegeben durch

$$H^T A H \hat{x} = H^T b;$$

dabei transformiert die Matrix $H\colon \mathbb{R}^{n^L} \to \mathbb{R}^{n^L}$ die Koeffizienten bezüglich der hierarchischen Basis in die bezüglich der Knotenbasis. Für eine beliebige Funktion u_h gilt:

$$u_h = \sum_i (\hat{x})_i \psi_i = \sum_i (\hat{x})_i \sum_j \omega_{ij}\varphi_j = \sum_j \left(\sum_i \omega_{ij} (\hat{x})_i \right) \varphi_j = \sum_j (H\hat{x})_j \varphi_j.$$

Somit ist $(H)_{ij} = \omega_{ji}$.

Wendet man die Transformation in jedem Iterationsschritt an, so schreibt sich eine Iteration in der Standardknotenbasis als

$$x^{neu} = x^{alt} + H\hat{D}^{-1}H^T(b - Ax^{alt})$$

(mit Prolongation H von hierarchischer auf Knotenbasis, und Restriktion H^T).

9.4.2 Anwendung auf das Schurkomplementproblem

Es sei $x = (x_I, x_B)^T$, $\hat{x} = (\hat{x}_I, \hat{x}_B)^T$ die Blockzerlegung bezüglich Basisfunktionen zu inneren (I) bzw. Randknoten (B). Dann gilt wegen $\hat{A} = H^T A H$:

$$\begin{pmatrix} \hat{A}_{II} & \hat{A}_{IB} \\ \hat{A}_{BI} & \hat{A}_{BB} \end{pmatrix} = \begin{pmatrix} H_{II}^T & 0 \\ H_{IB}^T & H_{BB}^T \end{pmatrix} \begin{pmatrix} A_{II} & A_{IB} \\ A_{BI} & A_{BB} \end{pmatrix} \begin{pmatrix} H_{II} & H_{BI} \\ 0 & H_{BB} \end{pmatrix}.$$

Dies bedeutet, dass die Transformation von hierarchischer Basis in Knotenbasis auf dem Koppelrand B nur Werte auf B braucht, d.h. $H_{BI} = 0$.

Weiterhin ist

$$\hat{D} = \begin{pmatrix} \hat{D}_{II} & 0 \\ 0 & \hat{D}_{BB} \end{pmatrix}$$

eine Blockdiagonalmatrix ($\hat{D}_{II}$ ist selbst diagonal, $\hat{D}_{BB}$ enthält das Grobgittersystem + Diagonalen).

Für das weitere Vorgehen benötigen wir folgenden

Hilfssatz 9.2 Sei A eine Matrix mit 2×2 Blockgestalt bezüglich $I = I_I \cup I_B$. Mit $\mathrm{Schur}(A) = A_{BB} - A_{BI}A_{II}^{-1}A_{IB}$ bezeichnen wir das Schurkomplement des Blockes A_{II}. Weiter sei T eine *obere Blockdreiecksmatrix* (Basistransformation). Dann gilt

$$\begin{aligned} \mathrm{Schur}(T^T A T) &= \mathrm{Schur}\left(\begin{pmatrix} T_{II}^T & 0 \\ T_{IB}^T & T_{BB}^T \end{pmatrix} \begin{pmatrix} A_{II} & A_{IB} \\ A_{BI} & A_{BB} \end{pmatrix} \begin{pmatrix} T_{II} & T_{IB} \\ 0 & T_{BB} \end{pmatrix} \right) \\ &= T_{BB}^T \,\mathrm{Schur}(A) T_{BB} \end{aligned}$$

BEWEIS: Ausrechnen! □

Das hierarchische Basis-Verfahren löst iterativ das vorkonditionierte System

$$\hat{D}^{-1}H^TAH\hat{x} = \hat{D}^{-1}H^Tb.$$

Mit $\hat{x} = \hat{D}^{-\frac{1}{2}}\hat{y}$ können wir dies symmetrisieren zu

$$\hat{D}^{-\frac{1}{2}}H^TAH\hat{D}^{-\frac{1}{2}}\hat{y} = \hat{D}^{-\frac{1}{2}}H^Tb.$$

Nun strukturieren wir die Indexmenge I in $I = I_I \cup I_B$ und wenden zweimal den Hilfssatz an:

$$\text{Schur}(\hat{D}^{-\frac{1}{2}}H^TAH\hat{D}^{-\frac{1}{2}}) = \hat{D}_{BB}^{-\frac{1}{2}}H_{BB}^T\,\text{Schur}(A)H_{BB}\hat{D}_{BB}^{-\frac{1}{2}}.$$

Somit lautet das vorkonditionierte Schurkomplementsystem in der hierarchischen Basis

$$\hat{D}_{BB}^{-\frac{1}{2}}H_{BB}^TSH_{BB}\hat{D}_{BB}^{-\frac{1}{2}}\hat{y}_B = \hat{D}_{BB}^{-\frac{1}{2}}H_{BB}^Tg \qquad (\square)$$

(mit S und g wie zuvor). Rücktransformieren mit $\hat{x}_B = \hat{D}_{BB}^{-\frac{1}{2}}\hat{y}_B$ ergibt

$$\hat{D}_{BB}^{-1}H_{BB}^TSH_{BB}\hat{x}_B = \hat{D}_{BB}^{-1}H_{BB}^Tg.$$

Eine Iteration auf diesem System schreibt sich als:

$$x_B^{neu} = x_B^{alt} + \omega H_{BB}\hat{D}_{BB}^{-1}H_{BB}^T(g - Sx_B^{alt}),$$

d.h. man behält x_B in der Knotenbasis und transformiert in jeder Iteration. Der hierarchische Basis Vorkonditionierer lautet also:

$$\boxed{B_{HB} = H_{BB}D_{BB}^{-1}H_{BB}^T}$$

9.4.3 Zur Konvergenzgeschwindigkeit

Die Konvergenzabschätzung lässt sich auf die der ursprünglichen hierarchischen Basis Methode zurückführen. Dazu brauchen wir als

Hilfssatz 9.3 Sei A eine positiv definite Matrix mit 2×2 Blockstruktur wie oben. Dann gilt

$$\kappa(\text{Schur}(A)) \leq \kappa(A)$$

BEWEIS: Sei $S = \text{Schur}(A)$. Zerlege einen beliebigen Vektor x als

$$\underbrace{\begin{pmatrix} x_I \\ x_B \end{pmatrix}}_{=x} = \underbrace{\begin{pmatrix} Ex_B \\ x_B \end{pmatrix}}_{=x'} + \underbrace{\begin{pmatrix} x_I - Ex_B \\ 0 \end{pmatrix}}_{=x''}$$

mit $E = -A_{II}^{-1}A_{IB}$, so gilt

$$\begin{aligned}\langle x, Ax\rangle = \langle x' + x'', A(x' + x'')\rangle = \\ = \langle x', Ax'\rangle + 2\langle x'', Ax'\rangle + \langle x'', Ax''\rangle = \\ = \left\langle \begin{pmatrix} * \\ x_B \end{pmatrix}, \begin{pmatrix} 0 \\ Sx_B \end{pmatrix}\right\rangle + 2\left\langle \begin{pmatrix} * \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ Sx_B \end{pmatrix}\right\rangle + \left\langle \begin{pmatrix} x_I'' \\ 0 \end{pmatrix}, \begin{pmatrix} A_{II} \\ x_I'' \end{pmatrix}\right\rangle \\ = \langle x_B, Sx_B\rangle + \langle x_I - Ex_B, A_{II}(x_I - Ex_B)\rangle\,.\end{aligned}$$

Nun rechne für die extremen Eigenwerte von S:

$$\lambda_{\max}(S) = \max_{x_B \neq 0} \frac{\langle x_B, Sx_B\rangle}{\langle x_B, x_B\rangle} \leq \max_{\substack{x_B \neq 0 \\ x_I = 0}} \frac{\langle x_B, Sx_B\rangle + \overbrace{\langle x_I - Ex_B, A_{II}(x_I - Ex_B)\rangle}^{\geq 0, \text{ da } A_{II} \text{ pos. def.}}}{\langle x_B, x_B\rangle + \langle x_I, x_I\rangle} \leq$$

$$\leq \max_{x \neq 0} \frac{\langle x, Ax\rangle}{\langle x, x\rangle} = \lambda_{\max}(A)$$

$$\lambda_{\min}(S) = \min_{x_B \neq 0} \frac{\langle x_B, Sx_B\rangle}{\langle x_B, x_B\rangle} \geq \min_{\substack{x_B \neq 0 \\ x_I = Ex_B}} \frac{\langle x_B, Sx_B\rangle + \langle x_I - Ex_B, A_{II}(x_I - Ex_B)\rangle}{\langle x_B, x_B\rangle + \langle x_I, x_I\rangle} \geq$$

$$\geq \min_{x \neq 0} \frac{\langle x, Ax\rangle}{\langle x, x\rangle} = \lambda_{\min}(A)$$

□

Satz 9.4 Das vorkonditionierte Schurkomplementsystem hat die Konditionszahl

$$\kappa(\hat{D}_{BB}^{-1} H_{BB}^T S H_{BB}) \leq C\left(1 + \log^2(\tfrac{H}{h})\right).$$

Beweis:

$$\begin{aligned}
\kappa(\hat{D}_{BB}^{-1} H_{BB}^T S H_{BB}) &= \\
&= \kappa(\hat{D}_{BB}^{-\frac{1}{2}} H_{BB}^T S H_{BB} \hat{D}_{BB}^{-\frac{1}{2}}) && \text{wegen } (\square) \\
&= \kappa(\text{Schur}(\hat{D}^{-\frac{1}{2}} H^T A H \hat{D}^{-\frac{1}{2}})) && \text{obige Ableitung nur rückwärts} \\
&\leq \kappa(\hat{D}^{-\frac{1}{2}} H^T A H \hat{D}^{-\frac{1}{2}}) && \text{Hilfssatz 9.3} \\
&\leq C\left(1 + \log^2(\tfrac{H}{h})\right) && \text{Resultat von (Yserentant 1986)}
\end{aligned}$$

□

9.5 Bramble-Pasciak-Schatz-Verfahren (BPS)

9.5.1 Konstruktion

Beim BPS-Verfahren (auch „iterative substructuring“) wird die Indexmenge I_B (Knoten auf Koppelrand) in Freiheitsgrade auf Kanten und Koppelknoten strukturiert. Die Freiheitsgrade auf Kanten können bei Bedarf noch den einzelnen Kanten zugeordnet werden:

$$I_B = I_E \cup I_V = I_{E^1} \cup I_{E^2} \cup \cdots \cup I_{E^{n^E}} \cup I_V.$$

n^E bezeichnet hier die Anzahl der Kanten. Das Schurkomplement hat die entsprechende Blockstruktur:

$$S = \begin{pmatrix} S_{EE} & S_{EV} \\ S_{VE} & S_{VV} \end{pmatrix} = \begin{pmatrix} S_{E^1E^1} & \dots & \dots & S_{E^1E^{n^E}} & S_{E^1V} \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ S_{E^{n^E}E^1} & & & S_{E^{n^E}E^{n^E}} & S_{E^{n^E}V} \\ S_{VE^1} & \dots & \dots & S_{VE^{n^E}} & S_{VV} \end{pmatrix}$$

Nun wechseln wir in eine *partielle hierarchische Basis,* bei der nur auf den Koppelknoten die hierarchischen Basisfunktionen verwendet werden. Der Übergang von dieser Basis in die Standardbasis wird beschrieben durch ("uberstrichen bezeichnet die partiell hierarchische Basis)

$$\bar{H}_{BB} = \begin{pmatrix} I_{EE} & \bar{H}_{EV} \\ 0 & I_W \end{pmatrix}$$

(die obere Zeile gehört zu E, die untere zu V). Für das transformierte Schurkomplement $\bar{S} = \bar{H}^T S \bar{H}$ gilt

$$\bar{S} = \bar{H}_{BB}^T S \bar{H}_{BB} = \begin{pmatrix} S_{EE} & \bar{S}_{EV} \\ \bar{S}_{VE} & \bar{S}_{VV} \end{pmatrix},$$

d.h. der Block S_{EE} ist unverändert. Als Vorkonditionierer für $\bar{S}$ verwendet man

$$\bar{D}_{BB}^{-1} = \begin{pmatrix} D_{EE}^{-1} & 0 \\ 0 & \bar{S}_{VV}^{-1} \end{pmatrix}, \quad D_{EE}^{-1} = \begin{pmatrix} S_{E^1E^1}^{-1} & & 0 \\ & \ddots & \\ 0 & & S_{E^{n^E}E^{n^E}}^{-1} \end{pmatrix}.$$

Dabei bezeichnet $S_{E^iE^i}^{-1}$ die exakte Lösung eines Systems $S_{E^iE^i}v = r$. In der Praxis kann dies auch näherungsweise mit einem der Vorkonditionierer für *zwei* Teilgebiete aus Abschnitte 9.2 gemacht werden, z.B. mit dem J-Operator. Die Kopplungen zwischen den Kanten wurden weggelassen.

Bleibt noch zu zeigen, wie das System $\bar{S}_{VV}$ aufzustellen ist.

Sei $R_V \colon \mathbb{R}^{I_B} \to \mathbb{R}^{I_V}$ die punktweise Restriktion vom ganzen Koppelrand auf die Koppelknoten. Sei R_H die Mehrgitterrestriktion vom feinsten auf das grobe Gitter wie bei den überlappenden Gebietszerlegungsverfahren in Kapitel 5.

Mit diesen Bezeichnungen gilt für einen Vektor $x_V \in \mathbb{R}^{I_V}$:

$$R_H^T x_V = \begin{pmatrix} -A_{II}^{-1} A_{IB} \\ I \end{pmatrix} \bar{H}_{BB} R_V^T x_V.$$

Dabei wurde angenommen, dass A_{II} die Diskretisierung von $-\Delta u$ im inneren der Teilgebiete ist, die Teilgebiete Dreiecksgestalt haben und mit linearen finiten Elementen dis-

kretisiert sind. Damit erhalten wir

$$\begin{aligned} A_H = R_H A R_H^T &= R_V \bar{H}_{BB}^T \begin{pmatrix} -A_{BI}A_{II}^{-1} & I \end{pmatrix} \underbrace{A \begin{pmatrix} -A_{II}^{-1}A_{IB} \\ I \end{pmatrix}}_{\begin{pmatrix} 0 \\ S \end{pmatrix}} \bar{H}_{BB} R_V^T \\ &= R_V \underbrace{\bar{H}_{BB}^T S \bar{H}_{BB}}_{\bar{S}} R_V^T \\ &= R_V^T \bar{S} R_V^T = \bar{S}_{VV}. \end{aligned}$$

Somit ist $\bar{S}_{VV}$ nichts anderes als die Grobgittermatrix des Zweigitter-Schwarz-Verfahrens.

Wegen der Blockdiagonalgestalt von $\bar{D}_{BB}$ ergibt sich der BPS Vorkonditionierer als Summe von Korrekturen entsprechend der einzelnen Kanten und der Koppelknoten:

$$\boxed{B_{\text{PBS}} = \sum_{i=1}^{n^E} R_{E^i}^T \tilde{S}_{E^iE^i}^{-1} R_{E^i} + \bar{H}_{BB}^T R_V^T A_H^{-1} R_V \bar{H}_{BB}^T.}$$

Dabei ist $R_{E^i} \colon \mathbb{R}^{I_B} \to \mathbb{R}^{I_{E^i}}$ die punktweise Restriktion vom Koppelrand auf die Kante E^i. Für die Knoditionszahl gilt die Abschätzung

$$\kappa(B_{BPS}S) \leq C(1 + \log(\tfrac{H}{h}))$$

mit C unabhängig von h, H und den Diffusionskoeffizienten, sofern diese in einem Teilgebiet konstant sind.

Bemerkung zur Implementierung: Die Anzahl der Kanten ist ungefähr $\frac{3}{2} \cdot$ #Elemente bei dreickigen Teilgebieten und $2 \cdot$ #Elemente bei viereckigen Teilgebieten. Jede Kante erfordert dann z.B. Lösen zweier Teilgebietsprobleme beim Neumann-Dirichlet Vorkonditionierer.

9.5.2 Interpretation als Schwarz Verfahren

Wir gehen aus von der üblichen Zweigitterkonstruktion $\mathcal{T}^H, \mathcal{T}^h$ mit entsprechenden FE-Räuen V^H, V^h, die stückweise linear sein sollen (Dreieckselemente, 2 Raumdimension, das Verfahren ist *nicht* direkt auf 3D erweiterbar!).

Das Variationsproblem lautet: Finde $u_h \in V^h$, so dass

$$a(u_h, v) = (f, v)_{L^2(\Omega)} \quad \forall v \in V^h, \tag{FE}$$

mit

$$a(u, v) = \sum_{i=1}^{p} \int_{\Omega_i} \rho_i \nabla u \cdot \nabla v \, dx = \sum_{i=1}^{p} \rho_i \, a_i(u, v).$$

Die Koeffizienten $\rho_i > 0$ können in *jedem* Teilgebiet beliebig gewählt werden. Das zu lösende Gleichungssystem für die Koeffizienten ist wie oben

$$\begin{pmatrix} A_{II} & A_{IB} \\ A_{BI} & A_{BB} \end{pmatrix} \begin{pmatrix} x_I \\ x_B \end{pmatrix} = \begin{pmatrix} b_I, b_B \end{pmatrix}$$

mit $I = I_I \cup I_B$ der Zerlegung der Indexmenge in innere und Koppelrandknoten.

Es sei $\mathcal{P}\colon \mathbb{R}^I \to V^h$ der FE-Isomorphismus, der jedem Koeffizientenvektor x seine FE-Funktion in der Standard-Knotenbasis zuordnet:

$$\mathcal{P}x = \sum_{i\in I} (x)_i \varphi_i^h.$$

Nun definiere folgende Teilräume von V^h:

$$\tilde{V}^h = \left\{ u \in V^h \,\middle|\, u = \mathcal{P}x \ \wedge\ x = \begin{pmatrix} -A_{II}^{-1}A_{IB}x_B \\ x_B \end{pmatrix} \right\}$$
$$\hat{V}^h = \left\{ u \in V^h \,\middle|\, u = \mathcal{P}x \ \wedge\ x = \begin{pmatrix} x_i & 0 \end{pmatrix}^T \right\}$$

$\tilde{V}^h$ ist der Teilraum der „diskret harmonischen Funktionen“. (Man überzeuge sich, dass $\tilde{V}^h$ wirklich ein Teilraum ist, d.h. abgeschlossen gegen Addition und skalare Multiplikation.)

Die beiden Teilräume bilde eine *direkte Zerlegung* von V^h:

$$V^h = \tilde{V}^h \oplus \hat{V}^h. \tag{$\oplus$}$$

Mit $u = \mathcal{P}x$, $v = \mathcal{P}y$ gilt

$$a(u,v) = x_B^T S y_B \qquad \text{falls } u, v \in \tilde{V}^h \tag{9.3a}$$
$$a(u,v) = 0 \qquad \text{falls } u \in \hat{V}^h, v \in \tilde{V}^h \tag{9.3b}$$
$$a(u,v) = x_I^T A_{II} y_I \qquad \text{falls } u, v \in \hat{V}^h \tag{9.3c}$$

Beweis:

zu (9.3a) $a(u,v) = \left\langle \begin{pmatrix} -A_{II}A_{IB}x_B \\ x_B \end{pmatrix}, \begin{pmatrix} 0 \\ Sy_B \end{pmatrix} \right\rangle = x_B^T S y_B$

zu (9.3b) $a(u,v) = \langle x, Ay \rangle = \left\langle \begin{pmatrix} x_I \\ 0 \end{pmatrix}, A \begin{pmatrix} -A_{II}^{-1}A_{IB}y_B \\ y_B \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} x_I \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ Sy_B \end{pmatrix} \right\rangle = 0$

zu (9.3c) $a(u,v) = \left\langle \begin{pmatrix} x_I \\ 0 \end{pmatrix}, A \begin{pmatrix} y_I \\ 0 \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} x_I \\ 0 \end{pmatrix}, \begin{pmatrix} A_{II}y_I \\ A_{IB}y_I \end{pmatrix} \right\rangle = x_I^T A_{II} y_I.$

□

Für das (FE-) Problem gilt dann wegen ($\oplus$): Jedes $u \in V^h$ kann eindeutig in $u_h = \tilde{u}_h + \hat{u}_h$ mit $\tilde{u}_h \in \tilde{V}^h, \hat{u}_h \in \hat{V}^h$ zerlegt werden, also ist

$$\begin{aligned}
& a(u_h, v) = (f,v)_{L^2(\Omega)} \quad \forall v \in V^h \\
\iff & a(\tilde{u}_h + \hat{u}_h, \tilde{v}^h + \hat{v}^h) = \left(f, \tilde{v}^h + \hat{v}^h\right)_{L^2(\Omega)} \quad \forall \tilde{v}^h \in \tilde{V}^h, \hat{v}^h \in \hat{V}^h \\
\iff & a(\tilde{u}_h, \tilde{v}^h) + a(\hat{u}_h, \hat{v}^h) = \left(f, \tilde{v}^h\right)_{L^2(\Omega)} + \left(f, \hat{v}^h\right)_{L^2(\Omega)} \quad \forall \tilde{v}^h \in \tilde{V}^h, \hat{v}^h \in \hat{V}^h
\end{aligned}$$

da man für $\tilde{v}^h = 0$ bzw. $\hat{v}^h = 0$ wählen kann:

$$\begin{aligned}
\iff & \begin{cases} a(\tilde{u}_h, \tilde{v}^h) = \left(f, \tilde{v}^h\right)_{L^2(\Omega)} & \forall \tilde{v}^h \in \tilde{V}^h \\ a(\hat{u}_h, \hat{v}^h) = \left(f, \hat{v}^h\right)_{L^2(\Omega)} & \forall \hat{v}^h \in \hat{V}^h \end{cases} \\
\iff & \begin{cases} Sx_B = g \qquad g = b_B - A_{BI}A_{II}^{-1}b_I \\ A_{II}x_I = b_I \end{cases}.
\end{aligned}$$

Nun ist

$$u_h = \tilde{u}_h + \hat{u}_h \qquad = \mathcal{P}\left[\begin{pmatrix} -A_{II}^{-1}A_{IB}x_B \\ x_B \end{pmatrix} + \begin{pmatrix} A_{II}^{-1}b_I \\ 0 \end{pmatrix}\right] =$$
$$= \mathcal{P}\begin{pmatrix} A_{II}^{-1}(b_I - A_{IB}x_B) \\ x_B \end{pmatrix}.$$

Die Zerlegung $V^h = \tilde{V}^h \oplus \hat{V}^h$ zerlegt das Problem (FE) in zwei völlig entkoppelte Teilprobleme. Das Variationsproblem in $\tilde{V}^h$ entspricht der Lösung des Schurkomplementproblems. Das Variationsproblem in $\hat{V}^h$ zusammen mit der harmonischen Fortsetzung in das Innere ergibt das rückwärts Einsetzen.

Zur Beschreibung des *BPS*-Verfahrens auf dem Schurkomplement bemerken wir zunächst, dass

$$V^H \subset \tilde{V}^h,$$

da die Basisfunktionen auf $\mathcal{T}^h$ diskret harmonisch sind (lineare Finite Elemente, $const \cdot \Delta$ im Teilgebiet).

Sei $R_{E^k} \colon \mathbb{R}^{I_B} \to \mathbb{R}^{I_{E^k}}$ die punktweise Restriktion vom Interface B auf die Kante $E^k \subset B$. Dann definiere

$$\tilde{V}^h_{E^k} = \left\{ u \in \tilde{V}^h \mid u = \mathcal{P}x \ \wedge \ (\forall i \in I_B \setminus I_{E^k} : (x)_i = 0) \right\}$$

Damit gilt

$$\tilde{V}^h = V^H \oplus \bigoplus_{k=1}^{n^E} \tilde{V}^h_{E^k}, \tag{D}$$

also eine direkte Zerlegung. Der BPS-Vorkonditionierer entspricht nun einer additiven Schwarz-Iteration bezüglich der Zerlegung (D) von $\tilde{V}^h$.

Denn: Sei $u^{alt} = \mathcal{P}x^{alt} \in \tilde{V}^h$, $u^{neu} = \mathcal{P}x^{neu} \in \tilde{V}^h$.

Grobgitterkorrektur: Finde $v^H = \mathcal{P}\begin{pmatrix} -A_{II}^{-1}A_{IB}R_C^T y_V \\ R_C^T y_V \end{pmatrix}$, so dass $a(u^{alt}+v^H, w) = (f,w)_{L^2(\Omega)}\ \forall w \in V^H$; dabei ist $R_C \colon \mathbb{R}^{I_B} \to \mathbb{R}^{I_V}$ die Mehrgitterrestriktion auf B und somit $R_C^T = \bar{H}_{BB}R_V^T$ die lineare Interpolation von Koppelknoten auf den ganzen Koppelrand. Es gilt also

$$\underbrace{A_H}_{=R_C S R_C^T}\ y_V = R_C(g - Sx_B^{alt}).$$

Korrektur auf einer Kante: Finde $\tilde{V}^H_{E^k} \ni v^{E^k} = \mathcal{P}\begin{pmatrix} -A_{II}^{-1}A_{IB}R_{E^k}^T y_{E^k} \\ R_{E^k}^T y_{E^k} \end{pmatrix}$ so dass $a(u^{alt} + v^{E^k}, w) = (f,w)_{L^2(\Omega)}\ \forall w \in \tilde{V}^k_{E^k}$. Man hat

$$(\underbrace{R_{E^k}SR_{E^k}^T}_{\substack{S_{E^kE^k} \\ \text{Hauptun-} \\ \text{termatrix} \\ \text{zu } E^k}})y_{E^k} = R_{E^k}(g - Sx_B^{alt}),$$

d.h. $B_{BPS} = \sum_{k=1}^{n^E} R_{E^k}^T S_{E^kE^k}^{-1} R_{E^k} + R_C^T A_H^{-1} R_C$ (man vergleiche mit oben).

9.5.3 Konvergenzabschätzung

Zu zeigen ist wie beim Schwarz-Verfahren

$$\boxed{\gamma \langle x_B, x_B \rangle_S \leq \langle B_{BPS} S x_B, x_B \rangle_S \leq \Gamma \langle x_B, x_B \rangle_S};$$

dann gilt $\kappa(B_{BPS}S) \leq \frac{\Gamma}{\gamma}$. Beachte, dass S positiv definit ist!

Abschätzung nach oben (Γ)

Analog wie beim Schwarz-Verfahren sind die $k+1$ Terme S-orthogonale Projektionen.

Mittels der Färbung der *Kanten,* wie sie in Abbildung 9.4 dargestellt ist, kann man wieder alle Projektionen der selben Farbe zu einer S-orthogonalen Projektion zusammenfassen und erhält wieder $\Gamma = 4 + 1 = 5$.

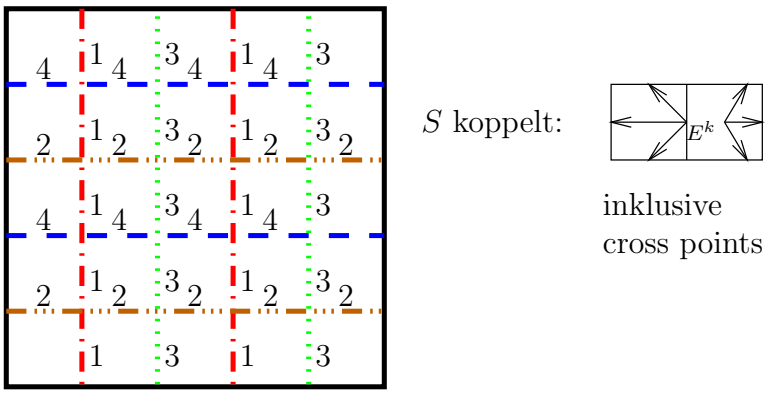


Abbildung 9.4: Zusammenfassung der Projektionen beim BPS-Verfahren

Abschätzung nach unten (γ)

Hier nutzt man wieder das Partition-Lemma:

$$\frac{1}{C_0} \langle x_B, x_B \rangle_S \leq \langle B_{BPS} S x_B, x_B \rangle_S$$

$$\iff \forall x_B \in \mathbb{R}^{I_B} \text{ gibt es eine Zerlegung } x_B = \sum_{k=1}^{n^E} R_{E^k}^T x_{E^k} + R_C^T x_V$$

$$\text{so dass } \sum_{k=1}^{n^E} \left\langle R_{E^k}^T x_{E^k}, R_{E^k}^T x_{E^k} \right\rangle_S + \left\langle R_C^T x_V, R_C^T x_V \right\rangle S \leq C_0 \langle x_B, x_B \rangle_S$$

Beachte: die Zerlegung ist hier eindeutig!

$$\Longleftrightarrow \forall u \in \tilde{V}^h \text{ gibt es eine Zerlegung } u = \sum_{k=1}^{n^E} u_{E^k} + u_V \text{ (mit } u_{E^k} \in \tilde{V}^h_{E^k}, u_V \in V^H)$$

$$\text{so dass } \sum_{k=1}^{n^E} a(u_{E^k}, u_{E^k}) + a(u_V, u_V) \leq C_0 a(u,u)$$

Der Nachweis dieser Abschätzung erfordert wieder funktionalanalytische Hilfsmittel.

Wir sind daran interessiert zu zeigen, dass C_0 (und somit γ) *unabhängig* von den Koeffizienten ρ_i in der Bilinearform ist.

Kann man für $u = \sum_k u_{E^k} + u_V$, mit $u_{E^k} \in \tilde{V}^h_{E^k}$ und $u_V \in V^H$, *lokal für jedes Teilgebiet* $i = 1, \ldots, p$ zeigen, dass

$$\sum_{k=1}^{n^E} a_i(u^i_{E^k}, u^i_{E^k}) + a_i(u^i_V, u^i_V) \leq C^i_0 a_i(u^i, u^i) \quad \forall u^i \in \tilde{V}^h \cap H^1(\Omega_i) \qquad (*)$$

mit C^i_0 *unabhängig von* ρ_i, so folgt daraus wegen $a(u,v) = \sum_{i=1}^p \rho_i \, a_i(u,v)$ die globale Abschätzung mit

$$C_0 = \max C^i_0.$$

In $(*)$ ist zu beachten, dass für u^i auf Ω_i im allgemeinen *keine* Nullrandbedingungen mehr vorliegen. Insbesondere gilt für Teilgebiete i mit $\partial\Omega_i \cap \partial\Omega = \emptyset$, dass $a_i(u^i, u^i) = 0$ für alle Funktionen $u^i \in \tilde{V}^h \cap H^1(\Omega_i)$ mit $u^i \equiv const.$ In diesem Fall kann $(*)$ nur gelten, wenn auch die linke Seite 0 ergibt. Dies ist jedoch der Fall, da dann gilt $u^i_V = u^i = 1$ und $u^i_{E^k} = 0$ (eindeutige Zerlegung!). Dies ist äquivalent dazu, dass BPS die „Null space property“ besitzt.

Wir können uns somit im Folgenden auf *ein* Teilgebiet zurückziehen und $(*)$ nachweisen.

Behandlung eines Teilgebietes Ω_i

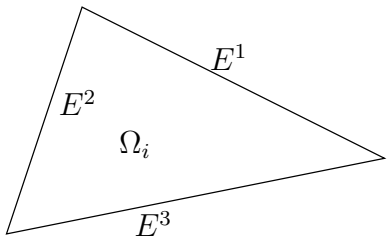


Abbildung 9.5: Lokale Kantennummern in einem Teilgebiet

Wir verwenden lokale Kantennummern (Abbildung 9.5): E^1, E^2, E^3.

$$\text{in diesem Abschnitt setze} \quad \boldsymbol{a(u,v) := \int_{\Omega_i} \nabla u \nabla v dx}$$

Zu zeigen ist: zu jedem $u \in \tilde{V}^h \cap H^1(\Omega_i)$ gibt es eine Zerlegung $u = \sum_{k=0}^{3} u_k$, mit $u_0 \in V^H \cap H^1(\Omega_i)$ und $u_k \in \tilde{V}^h_{E^k} \cap H^1(\Omega_i), k = 1, 2, 3$ mit

$$\sum_{k=0}^{3} a(u_k, u_k) \leq \underbrace{C_0}_{=C_0^i} a(u, u).$$

Da die Zerlegung des Funktionenraumes nicht überlappend ist, sind die u_k eindeutig bestimmt:

$$\boxed{u_0} = I_H u.$$

Setze $(u - u_0) = \mathcal{P}\begin{pmatrix} -A_{II}^{-1} A_{IB} x_B \\ x_B \end{pmatrix}$ (denn $u - u_0$ ist diekret harmonisch) dann setze für $k = 1, 2, 3$:

$$\boxed{u_k} = \mathcal{P}\begin{pmatrix} -A_{II}^{-1} A_{IB} R_k x_B \\ R_k x_B \end{pmatrix} \in \tilde{V}^h_{E^k} \cap H^1(\Omega_i);$$

wobei $R_k \colon \mathbb{R}^{I_{\partial\Omega_i}} \to \mathbb{R}^{I_{\partial\Omega_i}}$ die punktweise Restriktion auf Kante E^k ist.

Die entscheidende Abschätzung in den folgenden Beweisen ist:

> Sei u eine stückweise lineare, stetige FE-Funktion in zwei Raumdimensionen, so gilt
> $$\max_{x,y \in \Omega_i} |u(x) - u(y)|^2 \leq C(1 + \log(\tfrac{H}{h}))\, |u|^2_{H^1(\Omega_i)}\,.$$
> Siehe (Smith, Bjørstad und Gropp 1996) und die Referenzen dort.

Die linke Seite kann ersetzt werden durch

$$\max_{x,y \in \Omega_i} |u(x) - u(y)|^2 = \big(\underbrace{\max_{x \in \Omega_i} u(x)}_{=:u_{\max}} - \underbrace{\min_{y \in \Omega_i} u(y)}_{=:u_{\min}}\big)^2 = (u_{\max} - u_{\min})^2.$$

Abschätzung von u_0 (Dies ist eine lineare Funktion auf *einem* Dreieck)

$$\begin{aligned} a(u_0, u_0) = |u_0|^2_{H^1(\Omega_i)} &\leq C \underbrace{\frac{(u_{\max} - u_{\min})^2}{H^2}}_{\nabla u} \cdot \underbrace{H^2}_{\text{Integrieren}} \leq \\ &\leq C(u_{\max} - u_{\min})^2 \leq C(1 + \log(\tfrac{H}{h}))\, |u|_{H^1(\Omega_i)} = \\ &= C(1 + \log(\tfrac{H}{h})) a(u, u) \end{aligned} \tag{9.4}$$

Abschätzung der u_k**,** $k = 1, 2, 3$ Die u_k sind schwer zugänglich. Unter allen FE-Funktionen mit *gleichen Randdaten* auf $\partial\Omega_i$ ist die diskret harmonische Fortsetzung die mit minimaler Energie, denn:
sei $u \in \tilde{V}^h \cap H^1(\Omega_i)$, $w \in V^h \cap H^1(\Omega_i)$ und $u\,|_{\partial\Omega_i} = w\,|_{\partial\Omega_i}$. w kann zerlegt werden in $w = \tilde{w} + \hat{w}$ mit $\tilde{w} \in \tilde{V}^h \cap H^1(\Omega_i)$ und $\hat{w} \in \hat{V}^h \cap H^1(\Omega_i)$. Wegen gleicher Randdaten muß $\tilde{w} = u$ gelten, d.h. $w = u + \hat{w}$, also

$$a(w, w) = a(u + \hat{w}, u + \hat{w}) \underset{a(u, \hat{w}) = 0}{=} a(u, u) + a(\hat{w}, \hat{w}) \geq a(u, u),$$

also $a(u,u) \leq a(w,w)$.

Um $a(u_k,u_k)$ abzuschätzen, konstruieren wir eine Funktion w_k mit gleichen Randdaten, die sich leichter abschätzen läßt.

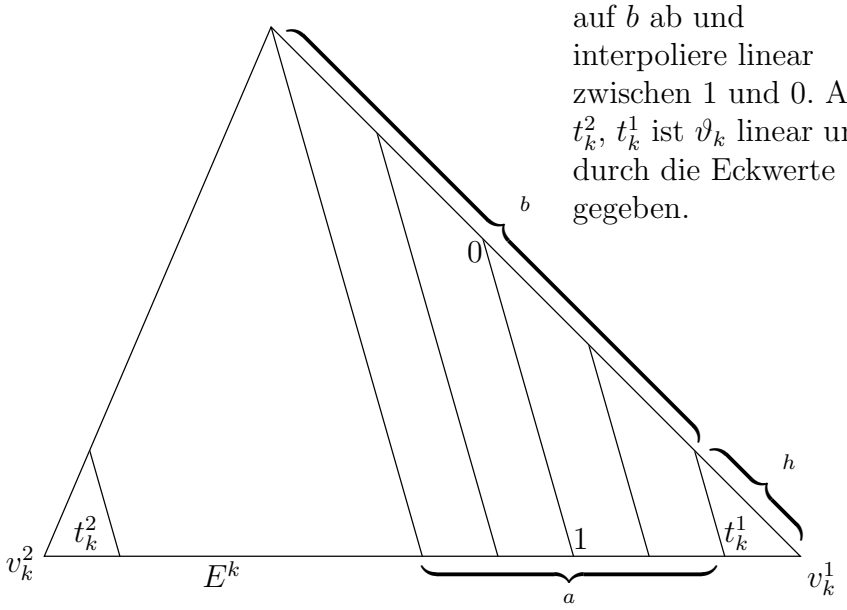
Dazu sei ϑ_k eine stetige Funktion auf Ω_i mit

$$\vartheta_k(x) = \begin{cases} 1 & x \in E^k \\ 0 & x \in \partial\Omega_i \setminus E^k \quad \text{d.h. auf Eckpunkten und anderen Kanten} \\ (\text{siehe Abb.9.6}) & \text{in } \Omega_i \end{cases} .$$

Dann setze

$$w_k = I_h(\vartheta_k(u - u_0)).$$

Wegen $w_k\,|_{\partial\Omega_i} = u_k\,|_{\partial\Omega_i}$ und obigem gilt $a(u_k,u_k) \leq a(w_k,w_k)$.



Bilde a isometrisch auf b ab und interpoliere linear zwischen 1 und 0. Auf t_k^2, t_k^1 ist ϑ_k linear und durch die Eckwerte gegeben.

Abbildung 9.6: Konstruktion von ϑ_k

Für den Gradienten von ϑ_k gilt:

$$\text{auf } t_k^1, t_k^2 : \|\nabla\vartheta_k\|_\infty \leq \frac{C}{h}$$

mit von der Form von t_k^1, t_k^2 abhängigem C; und

$$\text{auf } \Omega_i \setminus (t_k^1 \cup t_k^2) : \|\nabla\vartheta_k\|_\infty \leq \frac{C}{r(x)}$$

mit von der Form von Ω_i abhängigem C,

denn $\|\nabla\vartheta_k(x)\|_\infty = \frac{1}{l(x)}$ mit $l(x) \geq C^{-1}r(x)$ und $r(x) = \min(\mathrm{dist}(x, v_k^1), \mathrm{dist}(x, v_k^2))$. Wer es genau will:

$$\nabla\vartheta = \begin{cases} \nabla\vartheta(x) \cdot \frac{\vec{l}(x)}{\|\vec{l}(x)\|} = \frac{\partial\vartheta}{\partial r} = \frac{1}{\|\vec{l}(x)\|} \\ \nabla\vartheta(x) \cdot \frac{\vec{r}(x)}{\|\vec{r}(x)\|} = \frac{\partial\vartheta}{\partial r} = 0, \end{cases}$$

da $\vec{r}$ Höhenlinie von ϑ.

Abschätzen der Energie von $w_k = I_h(\vartheta_k(u - u_0))$ ergibt

$$a(w_k, w_k) = |w_k|^2_{H^2(\Omega_1)} = \sum_{j=1}^{2} |w_k|^2_{H^1(t_k^j)} + \sum_{t \neq t_k^1, t_k^2} |w_k|^2_{H^1(t)} \,.$$

Für die beiden speziellen Dreiecke t_k^1, t_k^2 gilt:

$$|w_k|^2_{H^1(t_k^j)} \quad \leq \quad \underbrace{C\frac{(u_{\max} - u_{\min})^2}{h^2}}_{\nabla w_k} \cdot \underbrace{h^2}_{\text{Integration}} \quad \leq \quad C(1 \;+\; \log(\tfrac{H}{h}))\, |u|^2_{H^1(\Omega_i)} \,, \quad \text{(I)}$$

mit von der Form von t_k^j abhängigem C.

Für ein $t \neq t_k^1, t_k^2$ gilt: (siehe den Beweis der additiven Schwarz-Iteration)

$$\begin{aligned} |w_k|_{H^1(t)} = |I_h(\vartheta_k(u - u_0))|^2_{H^1(t)} &\leq \left|I_h(\bar{\vartheta}_k^t(u - u_0)) + I_h((\vartheta_k - \bar{\vartheta}_k^t)(u - u_0))\right|^2_{H^1(t)} \leq \\ &\leq 2\bar{\vartheta}_k^t \, |u - u_0|^2_{H^1(t)} + 2\left|I_h((\vartheta_k - \vartheta_k^t)(u - u_0))\right|^2_{H^1(t)} ; \end{aligned}$$

dabei sei $\bar{\vartheta}_k^t$ der Mittelwert von ϑ_k auf t. Für die letzte Ungleichung benutzt man wieder $(a + b)^2 \leq 2a^2 + 2b^2$.

Für den ersten Term erhalten wir in der Summe: (man beachte $\bar{\vartheta}_k^t \leq 1$)

$$\begin{aligned} 2 \sum_{t \neq t_k^1, t_k^2} \bar{\vartheta}_k^t \, |u - u_0|^2_{H^1(t)} &\leq 2 \sum_{t \neq t_k^1, t_k^2} |u - u_0|^2_{H^1(t)} \\ &\leq 2\,|u - u_0|^2_{H^1(\Omega_i)} \leq 4\,|u|^2_{H^1(\Omega_i)} + 4\,|u_0|^2_{H^1(\Omega_i)} \leq \\ &\underset{(9.4)}{\leq} C(1 + \log(\tfrac{H}{h}))\, |u|^2_{H^1(\Omega_i)} \,. \end{aligned} \quad \text{(II)}$$

Für den zweiten Term kriegen wir:

$$2 \sum_{t \neq t_k^1, t_k^2} \left| I_h((\vartheta_k - \bar{\vartheta}_k^t)(u - u_0)) \right|^2_{H^1(t)} \leq$$

$$\underset{\substack{\text{inverse} \\ \text{Ugl.}}}{\leq} \quad 2 \sum_{t \neq t_k^1, t_k^2} Ch^{-2} \left\| I_h((\vartheta_k - \bar{\vartheta}_k^t)(u - u_0)) \right\|^2_{\mathbf{L^2(t)}} \leq$$

$$\leq 2 \sum_{t \neq t_k^1, t_k^2} Ch^{-2} \left\| I_h \left(\left(\frac{C' \cdot h}{r(x)} \right) (u - u_0) \right) \right\|^2_{L^2(t)} \leq C \sum_{t \neq t_k^1, t_k^2} \left\| I_h \left(\tfrac{1}{r} (u - u_0) \right) \right\|^2_{L^2(t)} \leq$$

$$\leq C \int\limits_{\Omega_i \setminus (t_k^1 \cup t_k^1)} \left[I_h \left(\tfrac{1}{r} (u - u_0) \right) \right]^2 dx \leq C (u_{\max} - u_{\min})^2 \cdot \int\limits_{\Omega_i \setminus (t_k^1 \cup t_k^1)} \left(I_h \left(\tfrac{1}{r} \right) \right)^2 dx$$

$$\leq C_1 (1 + \log(\tfrac{H}{h})) \, |u|^2_{H^1(\Omega_i)} \; + \; C_2 \int\limits_{\Omega_i \setminus (t_k^1 \cup t_k^1)} \tfrac{1}{r^2} dx$$

und mit $\int\limits_{\Omega_i \setminus (t_k^1 \cup t_k^1)} \frac{1}{r^2} dx = \int\limits_{r=h}^{H} \int\limits_{\theta} r^{-2} \cdot r \, d\theta dr \leq (1 + \log(\frac{H}{h}))$ ist das

$$\leq C \left(1 + \log(\tfrac{H}{h}) \right)^{\mathbf{2}} |u|^2_{H^1(\Omega_i)} \quad \text{(III)}$$

Nimmt man (I), (II) und (III) zusammen, so ergibt das

$$|w_k|^2_{H^1(\Omega_i)} \leq C \left(1 + \log(\tfrac{H}{h}) \right)^2 |u|^2_{H^1(\Omega_i)} .$$

Somit gilt für das Teilgebiet Ω_i

$$\sum_{k=0}^{3} a(u_k, u_k) \leq |u_0|^2_{H^1(\Omega_i)} + \sum_{i=1}^{3} |\mathbf{w_k}|^2_{H^1(\Omega_i)} \leq$$

$$\leq C \left(1 + \log(\tfrac{H}{h}) \right)^2 \underbrace{|u|^2_{H^1(\Omega_i)}}_{= a(u,u)};$$

somit ist $C_0^i = C \left(1 + \log(\frac{H}{h}) \right)^2$, mit nur von der Form des Ω_i und t_k^j, aber nicht von H,h ρ_i abhängigem C.

Für Teilgebiete am Dirichlet-Rand sind die entsprechenden Kanten einfach wegzulassen. Wir haben damit gezeigt, dass

$$\kappa(B_{BPS} \leq C \left(1 + \log(\tfrac{H}{h}) \right)^2$$
$$C \text{ unabhängig von } H, h, \rho_i$$

□

10 Algebraische Mehrgitterverfahren

18.6.05 1

10. Algebraisches Mehrgitterverfahren (AMG)

Geometrisches Mehrgitter
- vorgegebene Gitterhierarchie $T_H = T_0, T_1, \ldots, T_L = T$
- Gitterrestriktionen $R^e : \mathbb{R}^{I_{e+1}} \longrightarrow \mathbb{R}^{I_e}$
- Gitterprolongationen $P^e : \mathbb{R}^{I_e} \longrightarrow \mathbb{R}^{I_{e+1}}$
- Operatoren $A^0, \ldots, A^L$
- Glätter $W_1, \ldots, W_L$

(Algorithmus 8.9 ist multiplikatives Mehrgitter)

Nachteile:
- Bei Benutzung moderner CAD-basierter Gittergeneratoren wird ein sehr feines Gitter generiert (taugt nicht als Grobgitter)
  → Vergröberung liefert Grobgitter
- Abhängig vom Problem sind einfache Vergröberung mit Verdopplung der Gitterweite und Standard Interpolatoren nicht optimal

Verbesserungen:
- Problemabhängige Vergröberung (z.B. nur in eine Richtung, semi-coarsening)
- Operator-abhängige Interpolatoren
- Entwicklung von effizienten und robusten Glättern (Linienglätter, ILU-basierte Glätter)
  → Kompliziert für komplexe 3D-Gitter (ILU funktioniert nicht, Flächenglätter!)

Bei Mehrgitter müssen Glätter + Grobgitter Korrektur effizient zusammen spielen

AMG - Entwicklung ab den frühen 80'ern.

Idee:
1. Wähle fixen Glätter
2. Passe die Vergröberung daran an.

10.1 Algebraische Glattheit

Wir definieren folgende Skalarprodukte

Definition 10.1

Sei $A \in \mathbb{R}^{I \times I}$ s.p.d, $D := \mathrm{diag}(A)$ die Diagonalmatrix mit der gleichen Hauptdiagonale wie A. Dann definieren wir mittels des Euklidischen Skalarprodukts $\langle \cdot, \cdot \rangle$:

$$\begin{aligned} \langle u, v \rangle_0 &:= \langle Du, v \rangle \\ \langle u, v \rangle_1 &:= \langle Au, v \rangle \\ \langle u, v \rangle_2 &:= \langle D^{-1}Au, v \rangle \end{aligned} \qquad (10.0)$$

zusammen mit den assoziierten Normen $\|\cdot\|_i$, $i = 0, 1, 2$.

Die Untersuchungen des Jacobi und Gauss-Seidel Glätters durch Ruge und Stüben ergab:

Theorem 10.2

Sei $A \in \mathbb{R}^{I \times I}$ s.p.d. Dann gilt für den Glättungsoperator des symmetrischen Jacobi u. Gauss-Seidel Verfahrens:

$$\|We\|_1^2 \leq \|e\|_1^2 - \alpha \|e\|_2^2 \qquad (10.1)$$

$$\|We\|_1^2 \leq \|e\|_1^2 - \alpha \|We\|_2^2 \qquad (10.2)$$

für beliebiges $e \in \mathbb{R}^I$ mit einer Konstanten $\alpha > 0$

Somit wird der Fehler e nur gut durch die Verfahren reduziert, solange $\|e\|_2$ vergleichbar mit $\|e\|_1$ ist. Die Verfahren sind sehr schlecht für $\|e\|_2 \ll \|e\|_1$

Definition 10.3

Wir nennen den Fehler e „algebraisch glatt", falls er nicht mehr durch den Glättungsoperator W reduziert wird, d.h

$$We \approx e \qquad (10.3)$$

Formen wir das Energie-Skalarprodukt des Fehlers etwas um und benutzen Cauchy-Schwarz:

$$\langle Ae, e\rangle = \langle D^{\frac{1}{2}} D^{-\frac{1}{2}} Ae, e\rangle = \langle D^{-\frac{1}{2}} Ae, D^{\frac{1}{2}} e\rangle$$
$$\leq \|D^{-\frac{1}{2}} Ae\| \; \|D^{\frac{1}{2}} e\| = \|e\|_2 \, \|e\|_0 \qquad (10.4)$$

Somit impliziert $\|e\|_2 \ll \|e\|_1$, dass $\|e\|_1 \ll \|e\|_0$ gilt, bzw. explizit für M-Matrizen

$$\langle Ae, e\rangle = \frac{1}{2} \sum_{i,j} -a_{ij}(e_i - e_j)^2 + \frac{1}{2} \sum_{i,j} a_{ij} e_i^2 + \frac{1}{2} \sum_{i,j} a_{ij} e_j^2 \qquad (10.5)$$
$$= \frac{1}{2} \sum_{i,j} -a_{ij}(e_i - e_j)^2 + \sum_i \Big(\sum_j a_{ij} \Big) e_i^2 \ll \sum_i a_{ii} e_i^2$$

(negativ; $a_{ij} = a_{ji}$ Symmetrie!; A-Matrix)

Für den wichtigen Fall $\sum_{i \neq j} |a_{ij}| \approx a_{ii}$ bedeutet das

$$\frac{1}{2} \sum_{j \neq i} -a_{ij}(e_i - e_j) \ll a_{ii} e_i^2$$
$$\sum_{j \neq i} \frac{|a_{ij}|}{a_{ii}} \frac{(e_i - e_j)^2}{e_i^2} \ll 1 \qquad (10.6)$$

2D - Problem: $-\nabla K \cdot \nabla u = f$ auf Ω

$K = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix}$: $\varepsilon = 0{,}001$

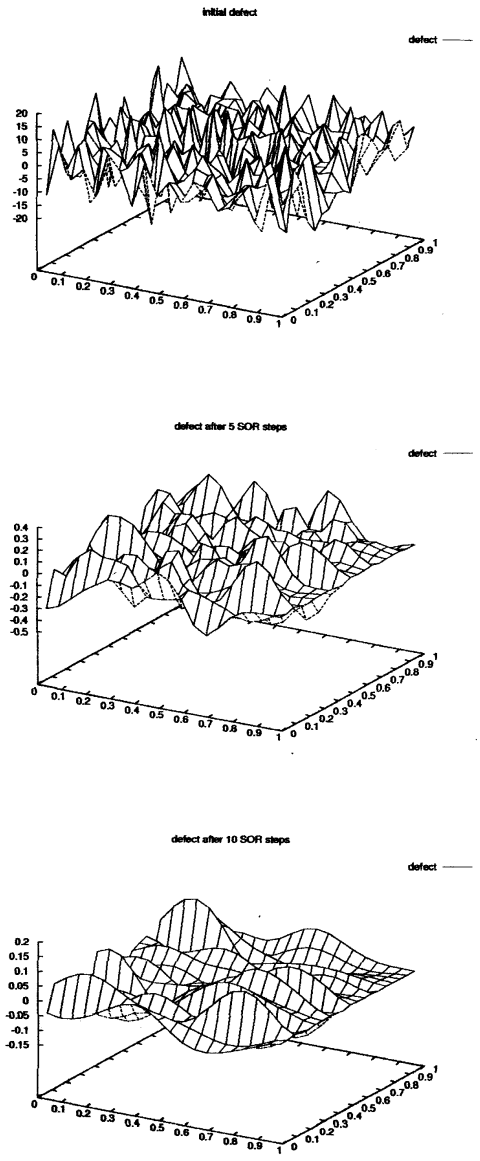


Figure 3.1: Algebraic Smoothing via SOR

Also ändert sich ein algebraisch glatter Fehler wenig von e_i zu e_j, falls $\frac{|a_{ij}|}{a_{ii}}$ relativ groß ist.

Definition 10.4

Sei $A \in \mathbb{R}^{N \times N}$ ein Matrix. Dann nennen wir $G(A) = (V, A)$ den Matrixgraph der Matrix A. Hierbei ist $V = \{v_1, \dots, v_N\}$ die Menge der gerichteten Knoten, die die Unbekannten repräsentieren. $E = \{(i_1, j_1), \dots, (i_m, j_m)\}$ mit $i_k \neq j_k$, die Menge der gerichteten Kanten, so daß (i,j) genau dann existiert, falls $a_{ij} \neq 0$.

Definition 10.5

Knoten v_i ist stark abhängig von Knoten v_j, falls gilt

$$-a_{ij} > \theta \max_{k \neq i} \{-a_{ik}\} \qquad (10.7)$$

für eine Schwelle $0 < \theta \leq 1$.
(Wir sagen auch v_j beeinflußt v_i stark)

Definition 10.6

Die Nachbarschaft des Knotens v_i ist definiert als

$$N_i = \{v_j \in V \mid (i,j) \in E\}.$$

Für einen glatten Fehler gilt wegen $\|e\|_2 \ll \|e\|_1$ insbesondere

$$r_i = a_{ii} e_i + \sum_{v_j \in N_i} a_{ij} e_i \approx 0 \qquad (10.8)$$

(Übung?)

10.2 Interpolatoren

Sei V die Menge der DOFs auf dem feinen Level.

Ziel: Finde $F \subset V$, $C \subset V$, $F \cap C = \emptyset$, $F \cup C = V$ unter Benutzung algebraischer Information

$\Omega_H = C$ ist Menge der DOFs auf dem groben Level

Finde $(R^H)^T : \Omega_H \to \Omega_h$ mit $\left((R^H)^T e\right)_i = \sum_{k \in C} w_{ik} e_k^H$

$w_{ik} = \delta_{ik}$ falls $i \in C$, so dass $(R^h)^T$ lokal arbeitet

Es soll Folgendes erfüllt sein:

- glatte Komponenten sollen genau auf dem groben Level approximiert werden können
- vom groben auf das feine Level soll glatte Funktion gut interpolier werden können
- das grobe Level soll deutlich weniger Unbekannte haben

Interpolationsoperator sei definiert als

$$R^T e^H = \begin{cases} e_i^H, & \forall i \in C \\ \sum_{V_k \in C_i} w_{ik} e_k^H, & \forall i \in F \end{cases} \qquad (10.5)$$

wobei $C_i \in C$ natürlich klein sein soll.

Falls für alle $v_i \in F$ gilt, dass $N_i \subseteq C$ ist, kann man mit (10.8) direkt die Interpolationsgewichte definieren (s. Übung), d.h. $w_{ik} = \frac{-a_{ik}}{a_{ii}}$

Sei $S_i \subset N_i$ die Menge der Knoten, die v_i gemäß (10.7) stark beeinflussen. Ferner $C_i := C \cap S_i$, $D_i := N_i - C_i$, sowie $D_i^s = D_i \cap S_i$ und $D_i^w = D_i - S_i$.
Aus 10.8 haben wir

$$a_{ii} e_i \approx - \sum_{v_j \in C_i} a_{ij} e_j - \sum_{v_j \in D_i^s} a_{ij} e_j - \sum_{v_j \in D_i^w} a_{ij} e_j \qquad (10.10)$$

Für $v_j \in D_i^w$ vertauschen wir e_j mit e_i.
Für $v_j \in D_i^s$ ist dieses vorgehen nicht ausreichend
Da der Wert von e_j von allen stark verbundenen Knoten beeinflusst wird, approximieren wir

$$e_j^h \approx \sum_{k \in C_i} (a_{jk} e_k) / \left(\sum_{k \in C_i} a_{jk}^h \right) \qquad (10.11)$$

Einsetzen von (10.11) in (10.10) und obige Vertauschung liefert nach Auflösen nach e_i die Interpolationsgewichte

$$w_{ij} = - \frac{a_{ij} + \sum_{v_m \in D_i^s} \left(\frac{a_{im} a_{mj}}{\sum_{v_k \in C_i} a_{mk}} \right)}{a_{ii} + \sum_{v_m \in D_i^w} a_{im}} \qquad (10.12)$$

Sobald die Interpolatoren gewählt sind ergibt sich das Grobleuel-Operator als

$$A^H = R A^h R^T ;$$

10.7 "Grobgitter" Problem

Die Unbekannten des Grobgitters (Menge C!) werden heuristisch gefunden, in dem man die Matrix auswertet. Dabei sollen folgende Kriterien berücksichtigt werden:

Kriterium 10.7

Für jeden Knoten $v_i \in F$ soll jeder Knoten $v_j \in N_i^S$ entweder in C sein, oder stark beeinflusst von mindestens einem Knoten aus C_i

(→ liefert gute Interpolation; Muss erfüllt sein)

Kriterium 10.8

C soll eine maximale Untermenge aller Knoten mit der Eigenschaft, dass keine zwei C-Knoten stark miteinander verbunden sind

(→ liefert gute Vergröberungsrate, Richtlinie!)

Algorithmus 10.3

1. Finde Zerlegung $C \cup F = V$ des Matrixgraphen $G(A) = (V, E)$, die Kriterium 10.8 erfüllt:

```
C = ∅; F = ∅; U = V
for (i=1; i≤|V|; i++) z_i = |(N_i^S)^T|
while (U ≠ ∅) {
    hole v_i ∈ U mit z_i maximal
    C = C ∪ {v_i}; U = U \ {v_i};
    for (v_j ∈ (N_i^S)^T ∩ U) {
        F = F ∪ {v_j}; U = U \ {v_j}
        for (v_k ∈ N_j^S ∩ U) z_k = z_k + 1
    }
    for (v_j ∈ N_i^S ∩ U) z_j = z_j - 1
}
```

2. Untersuche alle Knoten aus F, ob Kriterium 10.7 erfüllt ist. ~~Falls das nicht der Fall ist, werden sie zu C-Knoten.~~ Berechne Interpolationsgewichte:

```
T = ∅
while (F\T ≠ ∅){
  wähle v_i ∈ F\T;
  T = T ∪ {v_i}; done = 0;
  C_i = N_i^s ∩ C; D_i^s = N_i^s \ C_i; D_i^w = N_i \ N_i^s; C̄_i = ∅
  while (!done){
    d_i = a_ii + Σ_{v_k ∈ D_i^w} a_ik; d_j = a_ij ∀v_j ∈ C_i; done = 1;
    for (v_k ∈ D_i^s){
      if(N_k^s ∩ C_i ≠ ∅){d_j += a_ik a_kj / Σ_{k∈C_i} a_km ∀v_j ∈ C_i;}
      else {                    // 10.7 nicht erfüllt
        if(C̄_i ≠ ∅){ C = C ∪ {v_i}; F = F \ {v_i}; break;}
        else {
          C̄_i = {v_k}; C_i = C_i ∪ {v_k}; D_i^s = D_i^s \ {v_k};
          done = 0; break;
        }
      }
    }
  }
  if(v_i ∈ F){ C = C ∪ C̄_i; F = F \ C̄_i; w_ij = d_j/d_i ∀v_j ∈ C
}
```

Ist 10.7 für einen Nachbarn nicht erfüllt, wird er versuchsweise C_i und $\bar{C}_i$ Knoten und alle Nachbarn von v_i werden erneut geprüft. Ist 10.7 jetzt immer noch nicht erfüllt, wird v_i zum C-Knoten. Ansonsten werden die $\bar{C}_i$ zu C Knoten

Literatur

O. Axelsson und V. A. Barker (1984). *Finite Element Solution of Boundary Value Problems*. Academic Press.

Mathew Chan (1994). *Acta Numerica*.

P. Deuflhard und A. Hohmann (1993). *Numerische Mathematik I*. de Gruyter.

W. Hackbusch (1991). *Iterative Lösung großer schwachbesetzter Gleichungssysteme*. Teubner.

B. Smith, P. Bjørstad und W. Gropp (1996). *Domain Decomposition*. Cambridge University Press.

Yserentant (1986).